Stambaugh Correlations, Monkey Econometricians and Redundant Predictors

August 8, 2011

Abstract

We consider inference in a widely used predictive model in empirical finance. "Stambaugh Bias" arises when innovations to the predictor variable are correlated with those in the predictive regression. We show that high values of the "Stambaugh Correlation" will arise naturally if the predictor is actually predictively redundant, but emerged from a randomised search by data mining econometricians. For such predictors even bias-corrected conventional tests will be severely distorted. We propose tests that distinguish well between redundant predictors and the true (or "perfect") predictor. An application of our tests does not reject the null that a range of predictors of stock returns are redundant.

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

$$\begin{aligned} r_{t+1} &= \alpha + \beta z_t + e_{t+1} \\ z_{t+1} &= \phi z_t + w_{t+1} \end{aligned}$$

where the first equation captures the degree of predictability of some variable, r_t (usually some measure of returns or excess returns) in terms of some predictor variable z_t, (frequently some indicator of value, such as the price-dividend ratio); while the second models the predictor as an AR(1) process. Usually we are interested in testing $H_0 : \beta = 0$, i.e., that z_t is predictively redundant for r_{t+1}. Since Stambaugh (1999) it has been well known that if the predictor variable is persistent and there is a high absolute "Stambaugh Correlation" between the two innovations, e_t and w_t, then conventional OLS will lead to estimates of β 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, Sarkissian and Simian, 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).

While our analysis will largely focus on this simple system, the framework generalises fairly straightforwardly to allow for multiple predictors, structural breaks, etc. We can in principle incorporate considerable generality, by letting z_t be a weighting of underlying predictors, eg let $z_t = h'Z_t$, where Z_t is a vector of p predictors, that follows the vector autoregressive process, $Z_{t+1} = \Phi Z_t + W_{t+1}$.This replaces the second equation in the system, but leaves the first unchanged. The Stambaugh Correlation then naturally generalises to $\rho = corr\left(H'W_t, e_t\right)$. By allowing for non-standard but serially independent innovations we can also incorporate Markov switching models, etc (see Hamilton (1994, pp 678-9). We discuss these generalisations further in the main paper.

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

In most of the existing literature, a high Stambaugh Correlation is typically simply treated as a nuisance that complicates inference. In more recent contributions, Cochrane (2008a) and Pastor & Stambaugh (2009) have both put forward strong arguments that, under the natural alternative hypothesis that z_t is (in Pastor & Stambaugh's terminology) a "perfect" predictor (in the sense that a model with $\beta \neq 0$ and e_t white noise is the true data generating process), then on a priori grounds we would *expect* the Stambaugh Correlation to be high.

In this paper we show that high Stambaugh Correlations may also occur for a very different reason: if the predictor variable is redundant (in the sense that it tells us nothing about the future of returns that is not already in the history of returns), but simply proxies univariate predictability. This is all the more likely to occur if the predictor has emerged from a process of data mining. By analogy with the infinite number of monkeys ultimately typing the works of Shakespeare, we can think of this as an experiment in which an infinite number of monkey econometricians randomly sift through available data and run regressions, with the experimenter then simply choosing the predictor that has the highest R^2 for future returns. For such predictors, even the most stringent test procedures used in past research may not be enough to detect redundancy.

The standard null $\beta = 0$ implies $r_t = \alpha + e_t$, where in most of the existing literature, e_t is assumed to be white noise. We propose a generalised null, in which we allow for a degree of predictability of returns from some information set that may not be observable to the econometrician. Even if the true predictable component is unobservable, the history of returns itself will in general have some (but usually quite weak) predictive power, that can be captured by an ARMA representation.[1]

We show that, in this more general null model, if a predictive regression is estimated by OLS, conditioning only on z_t, then even if z_t is predictively redundant, it may be sufficiently correlated with the predictions from the ARMA representation to reject the standard null in its usual form, even after correcting for biases in the usual way, and even if the ARMA

[1] This necessary link goes back at least as far as Fama & French (1988) and underpins, whether explicitly or implicitly, the literature on "mean reversion" of asset returns.

representation has very weak predictive power. Furthermore, if z_t is a redundant predictor that has emerged from a data mining by monkey econometricians, we show that it will in some key respects strongly resemble Pastor & Stambaugh's (2009) "perfect predictor": it will have a high Stambaugh Correlation, and similar persistence to the true predictable component of returns; and the innovations in the predictive regression will be close to white noise.

Despite these superficial similarities, we show that there is a simple testing procedure for the null that z_t is a redundant predictor. Our tests exploit the property that if z_t is redundant, its *marginal* predictive power must disappear, once we condition correctly upon the univariate prediction.[2] Our tests have good power against the obvious alternative, that z_t is a "perfect predictor" (Pastor & Stambaugh, 2009*)*. We also note a simple diagnostic: redundant predictors will tend to resemble closely a "univariate predictor" that is simply an exponentially weighted average of past returns, and will tend to predict better, the closer is this resemblance. In contrast, a "perfect" predictor (or even an imperfect, but non-redundant proxy thereof) will tend to look *less* like the univariate predictor, the better it predicts.

To illustrate our proposed testing procedure, we examine four candidate predictors that have been used in past research, several of which appear to reject the standard 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. This conclusion is reinforced by the observation that these predictors closely resemble the univariate predictor, that simply summarises the history of returns. Indeed we note that two widely used indicators, the price-dividend and price-earnings ratios, actually predict *less* well than the univariate predictor.

The analysis of this paper can be linked to two important recent contributions to the literature.

Pastor & Stambaugh (2009) analyse a system very similar to our own, in which predictors are in general "imperfect" in that they are imperfectly correlated with the true predictable

[2] The null of redundance as used in this paper is identical to the null of No Granger Causality as originally stated in Granger (1969). But our test procedures differ from the usual Granger Causality tests in that we allow r_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.

component of returns. They also propose conditioning on the joint histories of returns and observable predictors. They do not however consider the possibility that some imperfect predictors may be redundant; nor, in their Bayesian framework, do they consider tests of the null of redundance. They also propose that low Stambaugh Correlations and non-white noise residuals in the predictive regression can provide a useful diagnostic of predictor imperfection. Our analysis runs directly counter to their proposed strategy, since it suggests that neither diagnostic will screen out redundant predictors, particularly if these emerge from horse races.

Ferson, Sarkissian and Simin (2003) consider systems that can be viewed as special cases of our own system, but in which a redundant predictor is always assumed to have a Stambaugh Correlation of zero. In our framework such predictors would, in sufficiently long samples, never win horse races; but in short samples they show that even such predictors may appear significant due to a spurious regression problem, if the predictable component of returns has high persistence. Whiel Ferson et al make an important contribution to our awareness of the risks of data mining, we argue that, by restricting their definition of redundant predictors to the subset with zero Stambaugh Correlations, they neglect a crucial feature of observable predictors: whether redundant or otherwise. Indeed we argue elsewhere (Robertson & Wright, 2009) that observable univariate properties of returns *require* that a "perfect predictor" must have a high Stambaugh Correlation.

The paper is structured as follows. In Section 1 we derive the relationships between the true process for returns and a redundant predictor. In Section 2 we describe our proposed test procedures. In Section 3 we describe our empirical applications, and Section 4 concludes. Appendices provide technical details.

1 Univariate Properties of Returns and Redundant Predictor Variables

1.1 The predictive system

We assume that the joint process for r_t and some predictor z_t can be represented in what Pastor & Stambaugh (2009) refer to as a "predictive system".[3]

The process for returns is given by

$$r_{t+1} = \mu_t + u_{t+1} \tag{1}$$

$$\mu_t = \overline{r} + a(L)v_t \tag{2}$$

where $\overline{r} = E(r_t)$; $[\begin{array}{cc} u_t & v_t \end{array}]'$ is a vector of serially uncorrelated shocks, with “true” Stambaugh Correlation $\rho =$corr(u_t, v_t); and the persistent process μ_t (which may not be observable) captures the predictable components of the true process for returns, with $a(L) = a_0 + a_1 L + a_2 L^2 + ...$some (possibly infinite order) polynomial in the lag operator (defined by $Lx_t = x_{t-1}$).[4] We define $R^2_\mu = 1 - \sigma^2_u/\sigma^2_r$ as the proportion of the total variance of returns explained by this predictable component.

The representation of returns is supplemented by an autoregressive representation of an observable predictor variable, z_t (which we assume for simplicity has a zero mean)

$$z_{t+1} = \phi z_t + w_{t+1} \tag{3}$$

where z_t may in principle be, in Pastor & Stambaugh’s (2009) terminology, either an "imperfect" or "perfect" predictor, depending on the value of ϕ, and the correlation between w_t and v_t.[5]

[3] Very similar systems are found in, for example, Campbell Lo and Mackinlay (1997) Chapter 7; Ferson, Sarkissian and Simin (2003); Cochrane (2008b).

[4] μ_t is often referred to as the expected return, since clearly, for some information set Ω_t, $\mu_t = E(r_{t+1}|\Omega_t)$. However, except in the context of a model of perfectly functioning complete markets with a rational representative investor who knows the full model, it is not clear *whose* conditional expectation this will be. But viewed as a state space model in which μ_t is the unobserved state variable for r_{t+1} (see, for example, Cochrane, 2008b) the model is considerably more general.

[5] The predictor is “perfect” if $\mu_t = \beta z_t$, implying $|corr(w_t, v_t)| = 1$; $\phi = \lambda$, but more generally “imperfect” if $|corr(w_t, v_t)| < 1$; $\phi \neq \lambda$

We shall show that it may also be predictively redundant, in a sense to be defined below, even when this correlation is non-zero.

We follow most of the literature in assuming an AR(1) representation of μ_t in (2), ie we set $a(L) = 1/(1-\lambda L)$, so that (2) becomes

$$\mu_{t+1} = (1-\lambda)\, r + \lambda \mu_t + v_{t+1} \tag{4}$$

It is well known (see, for example, Campbell, Lo and Mackinlay, 1995, Chapter 7; Cochrane, 2008b; Pastor & Stambaugh, 2009) that if the predictable component of returns takes this AR(1) form then the reduced form for r_t will be an ARMA(1,1). We could in principle make the dynamics of both expected returns and the predictor variable more complex. Both could in principle be linear combinations of elements of autoregressive vector processes . Much of our analysis extends to this more complex setup (we comment below on such extensions). However, we follow Pastor & Stambaugh's example in focussing in the bulk of the paper on the simple representations in (3) and (4), both on grounds of transparency, and because univariate predictability in returns is sufficiently weak that it provides no basis for assuming a higher order ARMA process.[6]

1.2 Conditioning on the history of returns

Substituting from (4) into (1) we can derive the reduced form ARMA(1,1) process for r_t:

$$r_{t+1} = (1-\lambda)\, \overline{r} + \lambda r_t + \varepsilon_{t+1} - \theta \varepsilon_t = \overline{r} + \left(\frac{1-\theta L}{1-\lambda L}\right) \varepsilon_{t+1} \tag{5}$$

where ε_t is a white noise innovation. If $\theta = \lambda$ the AR and MA components cancel and returns are univariate white noise. Note that this special case can arise even when $R^2_\mu > 0$.[7]

Using the ARMA(1,1) representation in (5) we can reverse-engineer a useful limiting repre-

[6] See Robertson & Wright, 2009.

[7] This feature of the reduced form is well-known: see for example Campbell, Lo and Mackinlay (1997), Pastor & Stambaugh (2009). For a full derivation of the ARMA representation, and a discussion of how the moving average parameter, θ, relates to the properties of the underlying system, see Robertson & Wright, 2009.

sentation of the same form as (1) and (2). Let

$$\mu_t^R \equiv E(\mu_t|R_t) \equiv E[r_{t+1}|R_t] \tag{6}$$

where $R_t = \left[\{r_i, \varepsilon_i\}_{i=0}^t, \lambda, \theta, \overline{r}\right]$ contains the history of returns and the components of the ARMA representation. Thus μ_t^R would be the expected return, if we conditioned upon an infinite history of returns.[8] We can then write down an alternative representation of (5), that takes the same form as (1) and (2), ie,

$$r_{t+1} = \mu_t^R + \varepsilon_{t+1} \tag{7}$$

$$\mu_{t+1}^R = (1-\lambda)\overline{r} + \lambda\mu_t^R + (\lambda - \theta)\varepsilon_{t+1} \tag{8}$$

where in this special case the Stambaugh Correlation ρ_R is equal to unity in absolute value. In the case of univariate white noise returns ($\theta = \lambda$) this collapses to a model of constant expected returns, conditional upon R_t. More generally the proportion of the variance of total returns explained by μ_t^R, which we denote $R_R^2 \equiv 1 - \sigma_\varepsilon^2/\sigma_r^2$, must lie between zero (if $\theta = \lambda$) and R_μ^2.

While R_t is effectively an infinite history, and hence in finite samples the univariate prediction μ_t^R will not be observable, its finite sample equivalent $E\left[\mu_t | \{r_i\}_{i=0}^t\right]$ will be observable, and will converge on μ_t^R as the sample size increases. In contrast the true state variable for returns, μ_t, may be permanently unobservable.[9]

[8] Our finite sample definition of the information set would make it possible to generate forecasts for any period $t > 0$ using the true ARMA representation in (5). This information could also be derived from an infinite sample of returns $\{r_i\}_{i=-\infty}^t$. We can also think of μ_t^R as the limiting conditional forecast of μ_t by applying the Kalman Filter to a finite sample of returns, given knowledge of the structural model in (1) and (2), once the Kalman Filter has converged to its steady state.

[9] More precisely, for $|\rho| \neq 1$, if we condition only on R_t,

$$\lim_{t\to\infty} var\left[\mu_t^R - E\left(\mu_t | \{r_i\}_{i=0}^t\right)\right] = 0 \text{ but } \lim_{t\to\infty} var\left[\mu_t - \mu_t^R\right] > 0$$

1.3 The Univariate Predictor

In setting up our tests, below, it will be helpful to define a "univariate predictor" x_t^R, which drives the time variation in the univariate prediction, μ_t^R, such that, using (8)

$$\mu_t^R = \overline{r} + (\lambda - \theta)\, x_t^R \tag{9}$$

where we note that the univariate predictor x_t^R satisfies

$$x_t^R \equiv \frac{\varepsilon_t}{1-\lambda L} \equiv \frac{r_t - \overline{r}}{1-\theta L} \equiv \sum_{i=0}^{\infty} \theta^i \left(r_{t-i} - \overline{r} \right) \tag{10}$$

Thus the univariate predictor is simply an exponentially weighted moving average of lags of r_t. Introducing this predictor variable has the advantage that in the special case of white noise returns, while the variance of μ_t^R collapses to zero, x_t^R is still well defined (despite having, from (9), no predictive power for returns).

1.4 A Redundant Predictor

We wish to consider the null hypothesis that the observable predictor z_t is predictively redundant, in the following general sense:

Definition 1 *A redundant predictor for returns,* r_{t+1} *contains no information of predictive value that is not already in the history of returns,* R_t

Note that this is simply a statement that a predictor does not Granger-Cause r_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 r_t (and possibly a set of other variables).[10] When the reduced form for r_t contains a moving average component, as in (5) (which it will always do if there are unobservable state variables) this distinction can be very important, as we shall show below.

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

Supppose that z_t is a redundant predictor in the sense of Definition 1, and we run the predictive regression

$$r_{t+1} = \alpha + \beta z_t + e_{t+1} \tag{11}$$

which is simply a least squares linear projection of r_{t+1} on z_t and a constant term. The resulting error process e_{t+1} will, by construction, be orthogonal to z_t but will not typically be white noise. We can then derive the following properties of the predictor and the predictive regression:

Proposition 1 *Assume that z_t follows the AR(1) process in (3).*

a) A sufficient condition for z_t to be redundant for r_{t+1} by Definition 1 is that w_t, the innovation to z_t can be written as

$$w_t = \psi \varepsilon_t + \varpi_t \tag{12}$$

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

b) Let $R_z^2 \equiv 1 - var(e_t)/var(r_t)$ be the predictive R-squared from a linear projection of r_t on z_{t-1}, of the form in (11); and let $\rho_z = corr(e_t, w_t)$ be the associated Stambaugh Correlation. If z_t is a redundant predictor by Definition 1, with innovations as in (12), then

$$\frac{R_z^2}{1 - R_z^2} \leq \rho_z^2 \frac{R_R^2}{1 - R_R^2} \tag{13}$$

where $R_R^2 \equiv 1 - \sigma_\varepsilon^2/\sigma_r^2$ is the predictive R-squared for the fundamental ARMA representation of r_t in (5). The expression for R_z^2 in (13) holds with equality if $\phi = \lambda$, and the implied upper bound $R_z^2 = R_R^2$ is attainable for some z_t since $\rho_z \in [-1, 1]$.

Proof. See Appendix 1. ■

The characteristics in Proposition 1 define a set of redundant AR(1) predictors with relatively simple and empirically relevant properties.[11] The specification of the innovation in Part a) of the proposition means that z_t is clearly redundant by Definition 1 (since any representation

[11] 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_R^2, or will have innovations containing a process ω_t that must satisfy very tight restrictions such that w_t remains white noise (which we require for z_t to be AR(1)).

of the history R_t has an equivalent representation in terms of $\{\varepsilon_i\}_{i=-\infty}^{t}$). For any specification that allowed the innovation w_t to be independently correlated with either current or lagged values of the true innovations u_t and v_t, z_t would not be redundant.

Part b) of the Proposition shows that a redundant predictor can appear to have significant predictive power in the predictive return regression (11). The apparent degree of predictive power will depend, first, on how well the ARMA representation of returns itself predicts (ie, on how high R_R^2 is) and second, on how good a proxy the observable predictor is for the univariate prediction defined in Section 1.3. This is in turn determined by two factors. The first is the predictor's Stambaugh correlation, ρ_z. In the Appendix we show that this is very closely related to the correlation between the innovations to the redundant predictor, in (12), and those to the ARMA representation in (5) (indeed we show that for r_t reasonably close to white noise, $\rho_z \approx$ corr(w_t, ε_t)). The second is the extent to which the the persistence of the redundant predictor matches the AR(1) parameter of μ_t, the true state variable for returns; we show that the closer ϕ is to λ, the better the prediction.

Proposition 1 provides an important insight into the characteristics of observable predictors. We frequently observe high Stambaugh correlations in predictive return regressions. In the existing empirical finance literature this characteristic has mainly been treated simply as a nuisance that complicates inference. In more recent contributions, Pastor & Stambaugh (2009) and Cochrane (2008a) have put forward *a priori* arguments why the "true" Stambaugh Correlation, $\rho =$corr(u_t, v_t) is likely to be strongly negative. But Proposition 1 suggests that high (absolute) Stambaugh Correlations may also arise for quite different reasons.

Consider the case where a given predictor variable has been the result of a wider search of candidate predictors (cf Sullivan, Timmerman and White, 1999; Ferson, Sarkissian and Simin, 2003) by way of some form of repeated data-mining horse race procedure. For a redundant predictor to win out in this procedure, a high absolute Stambaugh correlation is a *necessary* characteristic, since, from (13), the higher is the absolute Stambaugh correlation, the higher is R_z^2. Furthermore, 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, the proposition shows that this will also push up

R_z^2.[12] 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 of the predictive regression (11). If there is no constraint on the number of such potential indicators or on the manner in which they are constructed, then, as the sample size increases,*

a) The redundant predictor with the best track record will yield one-period-ahead predictions of r_{t+1} arbitrarily close to μ_t^R, as defined in (6);

b) Its Stambaugh correlation ρ_z will be arbitrarily close to unity in absolute value;

c) The innovations e_t in the predictive regression will be arbitrarily close to white noise.

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:[13] 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 returns.[14]

Proposition 1 and Corollary 1 point up a distinct contrast between our framework and that of Ferson et al (*op cit*). Their model can be viewed as a special case of ours, in which the Stambaugh Correlation ρ_z is zero. Proposition 1 implies that in this special case, and in sufficiently long samples, estimates of R_z^2 would converge to their population value of zero. However, Ferson et al show that in small samples, even such predictors may appear predictively significant, given a spurious regression problem, when both ϕ and λ are close to unity. In our

[12] See equations (23) and (25) in the proof of the proposition in Appendix 1.

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

[14] The data-mining monkey econometricians 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 empirical examples show, some predictors of stock returns get very close.

framework, since we allow data mining econometricians to pick predictors with $\rho_z > 0$, R_z^2 is in general positive in population, hence the apparent predictive power of redundant predictors would not disappear even in an infinite sample. We would also argue strongly that the case with $\rho_z > 0$ is more empirically relevant. Virtually all observable predictors of returns that have been proposed in the literature have high Stambaugh Correlations. They are also typically quite strongly correlated with the univariate prediction and thus with each other (see Section 2). That being the case, the data mining process modelled by Ferson et al, in which econometricians search through a sequence of predictors that both have $\rho_z = 0$ and are mutually uncorrelated, would in practice yield a very restricted set of predictors to choose from.

1.5 Predictor Characteristics, and a Predictive Hierarchy

As noted above, *a priori* arguments put forward by Pastor & Stambaugh (2009) and Cochrane (2008a) argue for the true Stambaugh correlation, ρ, being strongly negative. Robertson & Wright (2009) also show that observable univariate properties of returns must imply the same characteristic.[15] It is also fairly straightforwardly the case that the innovations, u_t in the "true" predictive regression must be white noise. Somewhat disconcertingly, the Proposition and Corollary imply that a redundant predictor *also* has both characteristics.[16]

This implies a caveat to the suggestion made by Pastor & Stambaugh (2009), drawing on their analysis of predictive systems, that low absolute Stambaugh Correlations and serially correlated residuals in the predictive regression should be viewed as indicating that predictors are "imperfect". While undeniably correct within their framework (in which, by construction, predictors in general are *not* redundant) Proposition 1 and Corollary 1 imply that a purely redundant predictor will appear to do very *well* on both criteria - and particularly so if it has arisen from a process of data mining. Thus use of these diagnostic tools in isolation may well lead to misleading conclusions.

[15] It arises by necessity if there is "mean reversion", in the sense that the variance ratio of long-horizon returns slopes downwards, and the true state variable has positive persistence.

[16] Since the limiting case of the Corollary has $\phi = \lambda$, $\rho_z = 1$, hence $e_t = \varepsilon_t$, and hence is white noise. (For a discussion of the more general case, see Appendix). Note that, while, for generality, the Corollary relates to the absolute Stambaugh Correlation, the same univariate characteristics that Robertson & Wright (2009) show must pin down the sign of the true Stambaugh correlation ρ will also imply the same sign for ρ_z if we normalise z_t to have the same sign as the predictions in (11).

However, there *is* a characteristic that does clearly distinguish between redundant predictors and either imperfect or "perfect" predictors: namely, the degree of predictive power, both absolute and marginal. It follows fairly straightforwardly from our analysis that, if z_t is redundant, we have the following ranking in terms of one-period-ahead goodness of fit:

$$0 \leq R_z^2 \leq R_R^2 \leq R_\mu^2 \tag{14}$$

where all the inequalities hold in strict form if $|\rho| \neq 1$, $|\rho_z| \neq 1$, and $\theta \neq \lambda$. A redundant predictor must predict less well than the ARMA, which in turn must predict less well than the true state variable, μ_t. There is also an equivalently ordered predictive hierarchy in terms of *marginal* predictive power: ie, z_t is redundant if we observe μ_t^R (or equivalently x_t^R) while μ_t^R in turn is redundant if we observe μ_t.[17] This is the basis for our tests.

Note also that our framework suggest a further useful diagnostic. We have:

$$\begin{aligned}
\frac{R_z^2}{R_R^2} &= \operatorname{corr}\left(z_t, \mu_t^R\right)^2 \\
\frac{R_\mu^2}{R_R^2} &= \frac{1}{\operatorname{corr}\left(\mu_t, \mu_t^R\right)^2}
\end{aligned}$$

If z_t is redundant, it will predict better, the higher is its correlation with μ_t^R (and hence x_t^R, as defined in Section 1.3) whereas, for the true state variable μ_t, the better it predicts, the *lower* is its correlation with μ_t^R. So if a predictor is strongly correlated with the univariate prediction, this should be a warning signal of potential redundance.

Note that, while we have exploited the properties of the predictive hierarchy in (14) for the case of an ARMA(1,1) and a single redundant predictor, Robertson & Wright (2011) show that in principle it applies equally well to the case where r_t is an ARMA(p,p) , and z_t is a some combination of multiple predictors.

[17] Note that this does not follow directly from the inequalities in (14). In fact, the inequalities stem from the marginal predictive hierarchy.

2 Testing for predictive redundance

2.1 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 r_t is given by the predictive system (1) and (2), or equivalently, the reduced form (5). We consider tests of this null against the alternative, $H_1 : \mu_t = \alpha + \beta z_t$, ie, that the observable predictor is a "perfect predictor" in Pastor & Stambaugh's (2009) terminology.[18]

Under our null, if we condition only on z_t, it may appear to have statistically significant predictive power for r_{t+1}; however it is redundant once we condition on the history of returns. Both null and alternative hypotheses can be represented straightforwardly in the following multivariate predictive regression:

$$r_{t+1} = \gamma_0 + \gamma_1 x_t^R + \gamma_2 z_t + \xi_{t+1} \tag{15}$$

where x_t^R, as defined in (10), captures the time variation in the univariate prediction, μ_t^R. The null of predictive redundance is $H_0 : \gamma_2 = 0$ (and $\xi_t = \varepsilon_t$), whereas under the alternative H_1:$\gamma_1 = 0$ (and $\xi_t = u_t$).

A key feature of this testing framework is that, in general, r_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 (11) and the predictor autoregression (3) are usually analysed on the assumption that e_t is white noise, hence under the null $H_0 : \beta = 0$, r_t is white noise plus a constant. But the alternative hypothesis in the standard framework is the same as in ours, namely $H_1 : \mu_t = \alpha + \beta z_t$, hence r_t is ARMA(1,1) in reduced form. Thus for tests of the standard null the order of the ARMA representation differs between the null and the alternative.

[18] We could in principle consider intermediate cases, for example where z_t is an imperfect predictor, but with some predictive power, even conditioning upon the ARMA prediction, because it provides a noisy signal of μ_t. However in this case, as noted by Pastor & Stambaugh (2009) 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 that z_t is "perfect".

Note that our null nests the standard null as a special case if (and only if) $\theta = \lambda$, and hence r_t is univariate white noise. In this case both γ_1 and γ_2 are zero under the null. This does not rule out the existence of returns being predictable conditional upon the "true" information set Ω_t (or even some subset thereof) but under the null z_t is as useless in predicting r_{t+1} as the univariate predictor, x_t^R, hence the standard null applies.

The practical obstacle to the implementation of a test procedure based on (15) is that it requires an estimate of the univariate predictor, x_t^R, or equivalently of the one-step-ahead prediction from the ARMA representation, μ_t^R. As noted above, in finite samples this will not be observable.

We consider three tests that use alternative approaches to this problem. We first describe the tests, and then, in Section 2.4, discuss their sampling properties.

2.2 Three Tests of Predictive Redundance

2.2.1 A two stage ARMA-based approach (RP_1)

By definition, μ_t^R the one-step-ahead prediction of r_t from the ARMA representation is, from (9) simply a scaling of the univariate predictor, x_t^R. The reverse is also, straightforwardly, the case, and our first test exploits this equivalence. In the first stage of the test procedure we estimate an ARMA(1,1) representation of r_t. We then take the one-step-ahead predictions from the estimated ARMA model and include them as a proxy for the univariate predictor in a multivariate predictive regression of the same form as (15). We then conduct an F-test for the marginal significance of z_t in this regression, and denote the resulting test statistic RP_1 ((**R**edundant **P**redictor$_1$)).[19]

2.2.2 A one stage ARMAX-based approach (RP_2)

A more direct approach is to estimate an equation of the same form as the standard predictive regression (11), but allowing the error term e_t to follow an ARMA(1,1). Under the null that z_t

[19] Note that here, as with the other two tests, since our objective is hypothesis testing, rather than prediction per se, we use full sample parameter estimates (for a discussion of the contrast between the two approaches, see Cochrane, 2008a). For predictive testing it would clearly be possible in principle to derive recursive ARMA predictions and run recursive F Tests.

is a redundant predictor, once we allow this specification of the error term then we again have $H_0 : \beta = 0$, since under the null the process for e_t can capture the reduced form for r_t itself as in (5). The test statistic RP_2 is the likelihood ratio statistic for this restriction against the unrestricted ARMAX(1,1) representation of r_t.

Note that in principle the methodology of both of our ARMA based tests, RP_1 and RP_2 can be extended straightforwardly to the more general ARMA processes that arise when true expected returns and the redundant predictor arise from a multiple predictor framework.

2.2.3 The indirect univariate predictor approach (RP_3)

A practical problem with the first two tests discussed above is that they rely on ARMA estimation, which is well-known to be problematic, particularly when (as will typically be the case) returns are close to white noise. We thus also consider an alternative approach, which, while more convoluted, has the advantage that it does not rely on ARMA estimation, but instead exploits information from the estimated predictive regression (11) and the properties of the observable predictor z_t.

A useful feature of the univariate predictor x_t^R defined in (10) is that it can be expressed solely in terms of the history of returns and the MA parameter, θ. Our third test procedure circumvents the problems of ARMA estimation 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 regression (11) and the predictor autoregression (3). We shall show below that, under this modified null, the resulting indirect estimate has distinctly better sampling properties than the direct estimate. We give a precise description of the derivation in Appendix 2; here we simply provide a brief sketch.

It is possible to show (see Robertson & Wright, 2009) that the MA parameter θ in (5) can be expressed in terms of three unit-free characteristics of the underlying process for returns:

$$\theta = \theta\left(\lambda, \rho, R_\mu^2\right) \tag{16}$$

where, as defined after (1) and (2), ρ is the "true" Stambaugh Correlation, R_μ^2 is the proportion of the variance of r_{t+1} due to the true state variable, μ_t and λ is its persistence. We noted above that if we replace μ_t with μ_t^R, the univariate prediction, defined in (6) we can derive a special case of the system for returns, as in (7) and (8), with a Stambaugh Correlation of unity in absolute value, which by construction must be consistent with the ARMA representation. Thus θ must also satisfy

$$\theta\left(\lambda, \rho, R_\mu^2\right) = \theta\left(\lambda, \rho_R, R_R^2\right) \tag{17}$$

where, from (8), $\rho_R = sign\left(\lambda - \theta\right)$.

While the arguments of the right-hand side of (17) depend on univariate properties, a natural amendment to our null hypothesis allows us to estimate them, and hence θ, *indirectly* from the predictive system in terms of z_t in (11) and (3).

Consider the joint null that z_t is redundant *and* that $\lambda = \phi$. From Corollary 1 this is a natural restriction on the null model if the predictor may have arisen from a process of data mining. (A similar argument is also used by Ferson et al, 2003).[20] Under this amended null, the estimate of λ is straightforward since we can set $\widehat{\lambda} = \widehat{\phi}$. But we can also exploit the fact that if $\phi = \lambda$, the inequality between R_z^2, ρ_z and R_R^2 in (13) in Proposition 1 holds with equality. Thus if we have estimates of R_z^2 and ρ_z from the system in (3) and (11) then we can derive an implied estimate of R_R^2 from (13), which coupled with an estimate of ϕ, is sufficient to derive an estimate of θ. We refer to this estimate as $\widehat{\theta}_z$.[21]

Using this indirect estimate of θ, an estimate of the univariate predictor $\widehat{x}_t^R$ consistent with the null can then be derived from (10), conditional upon some initial estimate $\widehat{x}_0^R$ 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 (15), with x_t^R replaced by $\widehat{x}_t^R$. We refer to this test statistic as RP_3.

The convoluted nature of this indirect approach can be readily justified by sampling properties. We show in Appendix 2 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 (partic-

[20] This does however have implications for the choice of critical values (see discussion in Section 3 below).

[21] We give a precise description of how $\widehat{\theta}_z$ is calculated in Appendix 2, where we also consider the impact of small sample biases, and discuss sampling properties.

ularly when the true value of λ and θ are very close), and hence we have a better estimate of the univariate predictor in deriving the test statistic. A corresponding disadvantage is the usual joint null problem, that we may reject the null, even when z_t is truly redundant, because ϕ differs from λ. However, as noted above we are, given Corollary 1, likely to be particularly interested in this joint null. While our other two tests do not suffer from a joint null problem, they have the offsetting disadvantage that they rely on ARMA estimation.

2.3 A link with the standard Stambaugh Bias problem

Consider the multivariate predictive regression (15). As already noted, we can consider our tests as alternative ways of producing an estimate of the univariate prediction μ_t^R, and hence of the true univariate predictor, x_t^R. Suppose, for the sake of argument, that we actually had data for x_t^R, and knew the true values of λ and θ. Then, using (7) and (9), it follows that, by setting $\gamma_1 = \lambda - \theta$ in (15), we would have $r_{t+1} - \left[\overline{r} + \gamma_1 x_t^R\right] = \varepsilon_{t+1}$, the true univariate innovation. If this were the case then (15) could be re-written as

$$\varepsilon_{t+1} = \gamma_2 z_t + \xi_{t+1}$$

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 (11) but with a dependent variable that is univariate white noise by construction. Hence in this case the only form of small sample biases we would have to worry about would be those identified by Stambaugh (1999). Given that, as we have already noted, a redundant predictor 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.

However, in practice we must form an estimate of the univariate predictor x_t^R (or, equivalently, of μ_t^R, the univariate prediction). 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_1 and RP_2 rely) are more problematic, particularly when r_t is close to being white noise. For this reason we also examine the distribution of the test statistic RP_3 under the more general null, $\phi \neq \lambda$, despite the fact that in this more general case it is mis-specified.[22]

2.4 Monte Carlo Evidence

While all our simulations are carried out for a wide range of values of λ and θ, it should be borne in mind that we are likely to be particularly interested in combinations where the two parameters are close to being, or are actually equal (ie, near or on the diagonal in the table) since in all such cases returns are near-white noise. Additionally, Robertson & Wright (2009) argue that if returns are both near-white noise and have a declining horizon variance ratio (commonly termed "mean reversion"), then there is a strong argument for focussing on combinations with relatively high values of λ, and with θ lying strictly between λ and unity.

Table I 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, as we assume, an ARMA(1,1).

[Table 1 about here]

Panel A of Table I shows the implied population R-squared, R_z^2, from a least squares linear projection of the form (11) 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 $\phi = \lambda$, 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, (13) in Proposition 1 holds with equality, and R_z^2 is to a reasonable approximation just a scaling-down of R_R^2, the ARMA R-squared. The table shows that the key determinant of R_z^2 is therefore how far the

[22] While the first two tests generalise readily and straighforwardly to multiple predictor models, our third test becomes even more convoluted in its construction, but in principle the approach remains valid.

two key ARMA parameters, λ and θ, are from the diagonal. Along the diagonal the predicted process, r_t, is white noise, so the true values of both R_R^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 I shows the simulated size[23] of a 1-sided t-test on the OLS estimate of β in (11), 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 ϕ increases, the bias increases.[24] 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 ARMA predictions. 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 II to IV 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,[25] for a range of values of λ and θ. As in Table I we set ρ_z, the Stambaugh Correlation of the redundant predictor, to 0.9. For the AR parameter ϕ of the redundant predictor, we consider three different variants. In Table II we assume $\phi = \lambda$; in Table III we assume that it is systematically lower ($\phi = \frac{\lambda}{2}$), while in Table

[23] For details of simulation methodology see Appendix 3.

[24] The size distortion along the diagonal increases for lower values of T: For example, for $\phi = \lambda = \theta = 0.95$ the simulated size of a nominal 5% test increases to 10% and 14% for $T = 100$ and $T = 50$ respectively.

[25] 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 and third tests but the results are very similar.

IV we assume it is systematically higher $\left(\phi = \lambda + \frac{1-\lambda}{2}\right)$.[26]

[Tables II to IV about here]

The most notable feature of these tables is that, while all three tests clearly display size distortions, these are typically much more modest, and, crucially, 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 I. The comparison with the standard testing framework is also, for all three tests, most favourable in the cells next to the diagonal, where r_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:

- The two-stage ARMA-based test RP_1 (as described in Section 2.2.1) fairly systematically under-rejects the null in the neighbourhood of the diagonal, due to a generated regressor problem more than offsetting Stambaugh bias. In contrast the ARMAX-based test, RP_2 (as described in Section 2.2.2) 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).

- Under the restricted null that $\phi = \lambda$, size distortions for our third test, RP_3 (as described in Section 2.2.3) are very much more 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 2.3) , and thus can in principle be allowed for fairly easily. But it is also striking that even when ϕ is not equal to λ, as shown in Tables III and IV, and thus the test statistic is mis-specified, the size distortions remain fairly modest, and are again largely explicable by Stambaugh bias.[27] Well away from the diagonal the size distortions

[26] It is reasonable to assume some relationship between ϕ 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.

[27] 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 IV (since ϕ is systematically higher), and systematically less in the top panel of Table III (since ϕ is systematically lower).

become more significant, particularly in Tables III and IV, but univariate predictability in such cases is likely to be so readily detectable that these are much less likely to be relevant.

Table V provides an indicator of the power of the three tests. It shows rejection rates at a nominal 5% level under the alternative hypothesis that $\mu_t = \alpha + \beta z_t$, ie, that z_t is a "perfect" predictor. To quantify the alternative hypothesis we need to make some assumption about the nature of the true process for returns. We have noted already, in Section 1.2, that the univariate representation puts a lower bound on the true R-squared in (1); ie we have $\min(R^2_\mu) = R^2_R$ which is a function solely of the ARMA parameters, λ and θ. Robertson & Wright (2009) show that it is also possible to derive an *upper* bound on R^2_μ that also depends on just λ and θ.[28] In the simulations summarised in Table V, we set $R^2_\mu = 0.75 \min\left(R^2_\mu\right) + 0.25 \max\left(R^2_\mu\right)$, such that μ_t, the true state variable for returns offers only a relatively modest improvement over the univariate representation. For reference, the table also shows, in the bottom two panels, the implied value of R^2_μ and of the true Stambaugh Correlation, ρ. The lowest panel shows that for a wide range of values of λ and θ the true Stambaugh Correlation will be quite close to unity in absolute value: hence Stambaugh bias would affect small sample estimation even if we had data on the true state variable for returns.[29]

For most values of θ and λ all three tests correctly reject the null of redundance with high probability.[30] Only for values of $\theta > \lambda$ and both close to unity do rejection rates fall off.[31] Thus power is generally good even for a modest degree of true predictability of returns.

[Table V about here]

[28] See Appendix 2.1 for a brief summary.

[29] The Table also illustrates a result proved in Robertson & Wright (2009) that if $\theta > \lambda$ (which implies that the variance ratio of long-horizon returns slopes downwards) the Stambaugh correlation must be negative (consistent with the a priori arguments of Cochrane, 2008a and Pastor & Stambaugh, 2009) and tends to -1 as θ tends to unity.

[30] Note that the comparison between the tests is complicated by the differences in size distortions - the relevant values of which are shown in Table II. 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.

[31] In such cases, Robertson & Wright (2009) show that the upper and lower bounds for R^2_μ become very close, and, as shown in the bottom panel of Table V, ρ, the true Stambaugh Correlation, tends to -1. Thus in these limiting cases μ_t becomes harder to distinguish from μ^R_t, and hence it becomes harder to reject the null that it is redundant. However, in such cases, arguably this reduction in power is of no great importance, because the *marginal* predictive power of μ_t is so modest.

3 Testing for Redundant Predictors of Real Annual US Stock Returns

In Table VI we apply our proposed test procedure to a range of alternative predictors of real stock returns, over a long sample of annual data:[32] 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.[33]

[Table VI about here]

In Panel A we show the two key characteristics of each of the predictors on which our analysis has focussed. All are strongly persistent (ie $\widehat{\phi}$ is high), and most have high absolute Stambaugh Correlations ($\widehat{\rho}_z$).[34] We noted in Section 1.4 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 (11), estimated by OLS, reject the "standard null" $H_0 : \beta = 0$ 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 have quite strongly significant predictive power, and a third is significant at the 10% level.[35]

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

[32] 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.

[33] 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 (2008a); 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.

[34] The positive sign is consistent both with our priors and those of Cochrane (2008a) and Pastor & Stambaugh (2009) because all indicators have price in the numerator, and hence $\beta < 0$, so innovations to the indicators are all negatively correlated with innovations to predicted returns.

[35] Bootstrapping methodology is described in Appendix 3.

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 $\phi = \lambda$, so we can rely simply on our third test, RP_3, that uses an indirect estimate of the univariate predictor, and which, as we noted in Section 2.4, has the most reliable sampling properties under the null. We thus cannot reject the null that, for these three predictors, their apparent predictive power simply reflects their correlation with the univariate 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 (shown in Table VI, Panel B) suggests insignificance.

For the cyclically adjusted P/E, which appears significant at at least the 10% level on standard (Stambaugh-proofed) tests, it is noteworthy that the very high Stambaugh correlation of 0.98 shows a striking similarity with the result we proposed in Corollary 1. This suggests that 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 third test, RP_3. However, there is a joint null problem, arising from the restriction that $\phi = \lambda$. We therefore need to look at the first two tests, RP_1 and RP_2, that do not impose this 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 1.4. To the extent that a predictor variable is chosen on the basis of horse races between predictive regressions, this prior filtering of the data means that there can be significant distortions to p-values (cf Ferson, Sarkissian and Simin, 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.[36]

Figure 1 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 univariate predictor, constructed using the formula in (10) from the history of returns, and the indirect estimates of $\widehat{\theta}_z$ derived from the properties of each of the predictor variables as described in Section 2.2.3. Unsurprisingly the estimated univariate 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 univariate predictor is very strong (around 0.85 for both predictors), while for the dividend yield (plotted as the price-dividend ratio) the correlation is lower (0.68). This is consistent with the analysis of Section 1.4, in which we noted that, for a redundant predictor, there will be a direct correspondence between this correlation and its apparent predictive power. Thus it seems reasonable to conclude that the dividend yield is simply a poorer proxy for the univariate predictor than are the P/E or q.

Cochrane (2008a) argues strongly that the weak predictive power of the dividend yield 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 univariate predictor.

Figure 1. Predictors and Univariate Predictors for US Stock Returns

[36] We are thus undermining the 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 VI. Robertson & Wright showed 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 VI makes these results look distinctly more marginal. However, 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 simply on the basis of its predictive power.

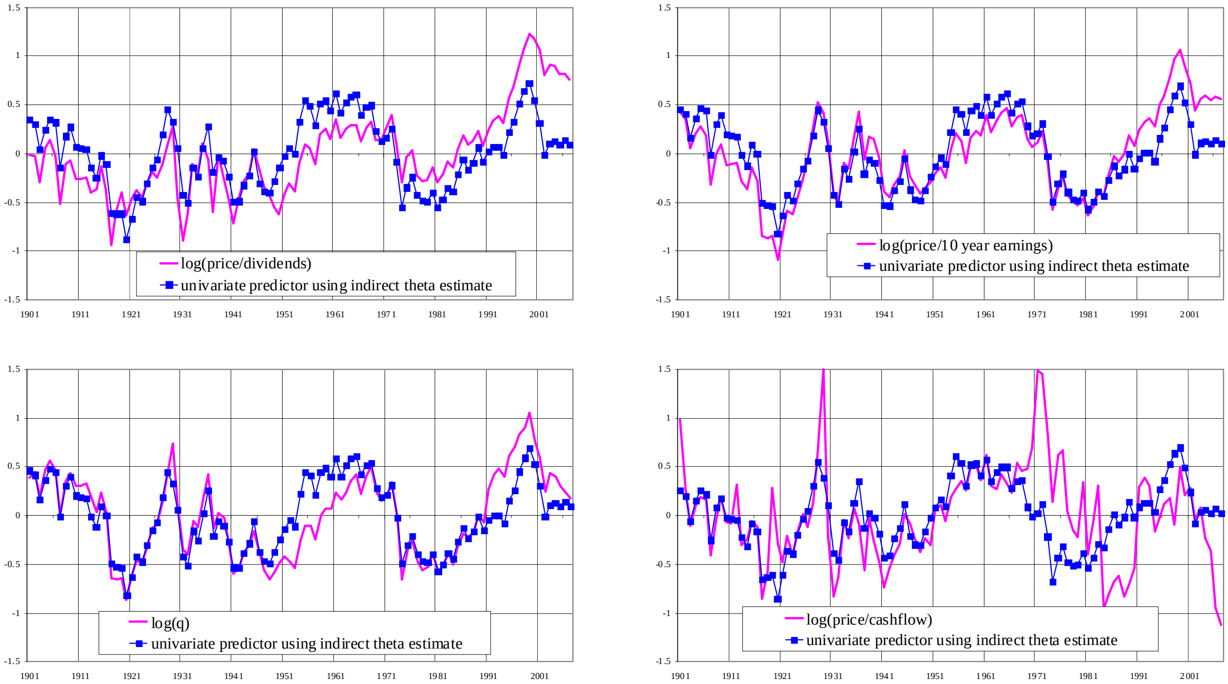

Figure 1 also shows that the correlation between the price-dividend ratio and the univariate predictor is much higher up until the mid 1980s. In this sample the apparent statistical significance of the dividend yield is also higher - consistent with the conclusion of Goyal and Welch (*op cit*) that the evidence of predictability from the dividend yield is an artefact of sample selection. Our results suggest the interpretation that during this shorter sample the dividend yield was simply a better proxy for the univariate predictor. Note however that even if we carry out our three tests over the truncated sample, we still do not reject the null that it is a redundant predictor.

In contrast Figure 1 shows that the cashflow yield has a much lower correlation with the univariate predictor (0.46) despite (as Table VI showed) having stronger predictive power for returns. This is again broadly consistent with our discussion in Section 1.4, in which we noted that the true predictor may have a quite low correlation with the univariate predictor. The marginal nature of the results for this variable in Table VI mean that we certainly cannot claim confidently that the cashflow yield is a "perfect" predictor in Pastor & Stambaugh's (2009) terms, but they do suggest that it is at least a non-redundant imperfect proxy for the true state variable.

It should be stressed that the evidence presented in both Table VI and Figure 1 does not *rely* on the assumption that stock returns have a significant degree of univariate predictability. Simulated p-values are also shown in Table VI, 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 to the simulated p-values for $\widehat{\beta}$, 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 univariate 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. (It is noteworthy that their respective univariate predictors actually predict better, in-sample, than both the P/E and the dividend yield.)

4 Conclusions

We have examined a predictive model that is widely used in empirical finance. In such models the "Stambaugh Correlation" between the innovations in a predictive regression, and to the predictor variable is usually very high in absolute value. It is well-known (since Stambaugh, 1999) that this may lead to small sample bias problems when testing the null that the coefficient on the lagged predictor variable is zero (and hence returns are white noise). We show that, in a more general process for returns, high Stambaugh Correlations will arise naturally if the predictor is redundant, but is correlated with the predictions from an ARMA representation of returns. This tendency will be reinforced if the predictor has emerged from data mining. For such predictors even bias-adjusted tests on the predictive regression may fail to detect redundancy. We propose three new tests that appear on the basis of simulation evidence to discriminate well between such redundant predictors and the true or "perfect" (Pastor & Stambaugh, 2009) predictor. When we apply these tests to four observable predictors used in past research, we cannot reject the null that three out of the four (the dividend yield, the price-earnings multiple and Tobin's q) are redundant, and are simply acting as proxies for the history of returns.

While we have focussed primarily on one-period-ahead prediction of a single return processs, by a single predictor, most of what we propose generalises readily to multiple periods, multiple predictors, and even multiple predicted processes.[37]

Do our theoretical and empirical results represent a counsel of despair for those interested in stock market predictability? Not necessarily. First, the finding that a given predictor is redundant on our definition does not necessarily invalidate the insights that it provides, especially if these are derived from a framework that can be related to an underlying present value model (as argued forcefully by Cochrane, 2008a, for example). Second, the reminder that certain key aspects of predictability may be essentially univariate in nature does not of itself diminish the economic significance of this predictability (*cf* Campbell & Thompson, 2007). Nonetheless the approach we advocate does suggest strongly that a reorientation of predictability testing is required, the essence of which is that any predictive model must outperform the ARMA if if it is to be taken seriously. We believe that this reorientation should also extend beyond predictive return regressions to a much wider class of predictive models.[38]

[37]Boudoukh et al's (2006) analysis suggests strongly that Stambaugh bias contaminates tests of long-horizon predictability at least as much as one-period-ahead tests. A more fundamental problem that we analyse in Robertson & Wright (2009) is that sufficiently strong long-horizon predictability must of itself be essentially a univariate phenomenon.

[38]Two obvious examples of time series that have similar near-white noise properties to those of real returns are real GNP growth and changes in real exchange rates. Predictor variables for such series also tend to have high Stambaugh Correlations, particularly if they are derived as cointegrating combinations; hence these are both obvious areas for extension of our testing approach.

Appendix

1 Redundant Predictors

1.1 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 (3) be given by

$$\begin{aligned} w_t &= \psi\varepsilon_t + \varpi_t \\ &= \rho_{\varepsilon w}\varepsilon_t + \sigma_\varepsilon\sqrt{1-\rho_{\varepsilon w}}q_t \end{aligned} \tag{18}$$

where $E(q_t) = 0, E(q_t^2) = 1$. This gives the useful normalisation

$$\begin{aligned} \sigma_w &= \sigma_\varepsilon; \\ corr(w_t, \varepsilon_t) &= \rho_{\varepsilon w} \end{aligned}$$

The AR(1) property of z_t requires w_t to be white noise, for which a sufficient (but not necessary) condition is that ϖ_t is white noise. We first consider the benchmark case where ϖ_t is white noise, and is also orthogonal to all lags of ε_t, as assumed in Proposition 1, before considering more general cases where w_t is white noise, but ϖ_t is not.

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

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

$$\beta = \frac{cov(r_{t+1}, z_t)}{var(z_t)} = \frac{cov(r_{t+1}, z_t)}{\sigma_\varepsilon^2/\left(1-\phi^2\right)}$$

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

$$\begin{aligned} cov\left(r_{t+1}, z_t\right) &= cov\left(\varepsilon_{t+1} + \left(\lambda - \theta\right)\frac{\varepsilon_t}{1-\lambda L}, \frac{w_t}{1-\phi L}\right) \\ &= \left(\lambda - \theta\right)\rho_{\varepsilon w} cov\left(\frac{\varepsilon_t}{1-\lambda L}, \frac{\varepsilon_t}{1-\phi L}\right) \\ &= \left(\lambda - \theta\right)\rho_{\varepsilon w} cov\left[\left(1+\lambda L + \lambda^2 L^2 + ...\right)\varepsilon_t, \left(1 + \phi L + \phi L^2 + ...\right)\varepsilon_t\right] \\ &= \frac{\left(\lambda - \theta\right)\rho_{\varepsilon w}\sigma_\varepsilon^2}{1-\lambda\phi} \end{aligned}$$

hence we have, for the general case

$$\beta = \left(\lambda - \theta\right)\rho_{\varepsilon w}\left(\frac{1-\phi^2}{1-\lambda\phi}\right) \tag{19}$$

and hence

$$R_z^2 = \frac{\beta^2\sigma_z^2}{\sigma_r^2} = \left(\theta - \lambda\right)^2\rho_{\varepsilon w}^2\left(\frac{1-\phi^2}{1-\lambda\phi}\right)^2\frac{\sigma_\varepsilon^2}{\left(1-\phi^2\right)}\frac{1}{\sigma_r^2} \tag{20}$$

but we have, using the Yule-Walker equations,

$$R_R^2 \equiv 1 - \frac{\sigma_\varepsilon^2}{\sigma_r^2} = \frac{\left(\theta-\lambda\right)^2}{1-\lambda^2+\left(\theta-\lambda\right)^2} \tag{21}$$

hence we have

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

so that substituting into (20) we can write

$$R_z^2 = \rho_{\varepsilon w}^2 R_R^2 g\left(\lambda, \phi\right) \tag{22}$$

where

$$g\left(\lambda,\phi\right) = \frac{\left(1-\lambda^2\right)\left(1-\phi^2\right)}{\left(1-\lambda\phi\right)^2} \tag{23}$$

The expression in (22) is defined in terms of $\rho_{\varepsilon w} = corr(\varepsilon_t, w_t)$. To show the link with the Stambaugh Correlation, note that we have

$$e_t = r_t - \beta z_{t-1} = \left(\frac{1-\theta L}{1-\lambda L}\right)\varepsilon_t - \beta\left(\frac{\rho_{\varepsilon w}\varepsilon_{t-1} + \sigma_\varepsilon\sqrt{1-\rho_{\varepsilon w}^2}q_{t-1}}{1-\phi L}\right)$$

hence

$$\rho_z \equiv corr(e_t, w_t) = \frac{\rho_{\varepsilon w}\sigma_\varepsilon^2}{\sigma_e\sigma_\varepsilon} = \rho_{\varepsilon w}\frac{\sigma_\varepsilon}{\sigma_e} = \rho_{\varepsilon w}\sqrt{\frac{1-R_R^2}{1-R_z^2}} \tag{24}$$

Note that, as discussed in Section 1.4, if returns are sufficiently close to white noise, both R_R^2 and R_z^2 are close to zero, and hence $\rho_z \approx \rho_{\varepsilon w}$. Using (24) to substitute into (22), we have

$$\frac{R_z^2}{1-R_z^2} = \rho_z^2 g(\lambda,\phi)\left(\frac{R_R^2}{1-R_R^2}\right) \tag{25}$$

Equivalently, as in the Proposition

$$\frac{R_z^2}{1-R_z^2} \leq \rho_z^2\left(\frac{R_R^2}{1-R_R^2}\right) \tag{26}$$

since $g(.)$ has a maximum value of unity at $\phi = \lambda$. This 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 (24) ρ_z lies within $[-1, 1]$, hence for $|\rho_z| = 1$ we have $R_z^2 = R_R^2$.■

1.3 Time series properties of e_t, the error term in the predictive regression (11) if z_t is redundant.

We have

$$e_t = r_t - \beta z_{t-1} = \left(\frac{1-\theta L}{1-\lambda L}\right)\varepsilon_t - \beta\left(\frac{\rho_{\varepsilon w}\varepsilon_{t-1} + \sigma_\varepsilon\sqrt{1-\rho_{\varepsilon w}^2}q_{t-1}}{1-\phi L}\right)$$

so in general w_t is an ARMA(2,2). In the limiting case of Corollary 1 we have $\phi = \lambda$, so the above expression simplifies to

$$\begin{aligned} e_t &= \left(\frac{(1-\theta L)\,\varepsilon_t - \beta \rho_{\varepsilon w} \varepsilon_{t-1} + \sigma_\varepsilon \sqrt{1-\rho_{\varepsilon w}^2} q_{t-1}}{1-\lambda L} \right) \\ &= \frac{\left(1 - \left[\theta\left(1-\rho_{\varepsilon w}^2\right) + \lambda \rho_{\varepsilon w}^2\right] L\right) \varepsilon_t + (\theta - \lambda)\, \rho \sigma_\varepsilon \sqrt{1-\rho_{\varepsilon w}^2} q_{t-1}}{1-\lambda L} \end{aligned}$$

since, from (19), setting $\phi = \lambda$, $\beta = (\lambda - \theta)\,\rho_{\varepsilon w}$. Thus in this case e_t is an ARMA(1,1). If we write this as

$$e_t = \left(\frac{1-\psi L}{1-\lambda L} \right) \zeta_t$$

for some white noise process ζ_t, then the moment condition for ψ is

$$\frac{\psi}{1+\psi^2} = \frac{\theta\left(1-\rho_{\varepsilon w}^2\right) + \lambda \rho_{\varepsilon w}^2}{1 + \left[\theta\left(1-\rho_{\varepsilon w}^2\right) + \lambda \rho_{\varepsilon w}^2\right]^2 + (\theta-\lambda)^2 \left(1-\rho_{\varepsilon w}\right)^2}$$

which gives the special cases

$$\begin{aligned} \rho_{\varepsilon w} &= \pm 1 \Rightarrow \psi = \lambda \Rightarrow e_t = \varepsilon_t \\ \rho_{\varepsilon w} &= 0 \Rightarrow \psi = \theta \Rightarrow e_t = r_t \end{aligned}$$

hence the ARMA(1,1) for e_t lives between white noise and the ARMA(1,1) for r_t in (5). But Corollary 1 implies $\rho_z \to 1 \Rightarrow \rho_{\varepsilon w} \to 1$, so that redundant predictors that are most likely to win horse races, will be closest to having white noise residuals.

1.4 More general processes for ϖ_t

Consider the more general processes for w_t and ϖ_t

$$w_t = \gamma_0 \varepsilon_t + \varpi_t; \quad E\left(\varepsilon_t \varpi_t\right) = 0$$

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

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

where s_t is white noise hence we require

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

hence if w_t is white noise, which we require for AR(1)ness of z_t, we require

$$\begin{aligned} E\left(w_t w_{t-k}\right) &= E\left[\left(\sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i} + \sum_{i=0}^{\infty} \pi_i s_{t-i}\right)\left(\sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-k-i} + \sum_{i=0}^{\infty} \pi_i s_{t-k-i}\right)\right] \\ &= \sigma_\varepsilon^2 \left[\gamma_k \gamma_0 + \gamma_{k+1} \gamma_1 + ...\right] + \sigma_s^2 \left[\pi_k \pi_0 + \pi_{k+1} \pi_1 + ...\right] \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \gamma_{k+i} \gamma_i + \sigma_s^2 \sum_{i=0}^{\infty} \pi_{k+i} \pi_i = 0 \ \forall \ k > 0 \end{aligned} \tag{27}$$

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 ϖ_t is white noise uncorrelated with all lags of ε_t.

We next consider two cases that satisfy the autocovariance condition in (27) for more general processes

1.4.1 Special case: ϖ_t is white noise but $E\left(\varpi_t \varepsilon_{t-j}\right) \neq$ for some $j > 0$

For this condition to hold we may allow $\gamma_j \neq 0$ for some $j > 0$. But by inspection of (27) this requires $\gamma_i = 0$ for $i \neq j$ (which in turn implies $\gamma_0 = 0$ - hence a zero contemporaneous correlation of w_t 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

$$w_t = \gamma_j \varepsilon_{t-j} + \varpi_t$$

which is a white noise process, as is $\varpi_t = s_t$. From this specification it follows that z_t will be a scaling of x^R_{t-j}, plus an orthogonal but serially correlated error. Hence it also follows that we must have

$$R_z^2 \leq \lambda^{2j} R_R^2$$

(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.

1.4.2 A more general case: ϖ_t not white noise

In this more general case, while ϖ_t may not be white noise, the autocovariance condition (27) puts a very tight restriction on the nature of the two underlying polynomials $\gamma(L)$ and $\pi(L)$, such that w_t is white noise. Any non-zero γ_i put corresponding restrictions on the π_i, which in turn increases the noise element in w_t, 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).

2 Properties of $\widehat{\theta}_z$, used in the indirect univariate predictor based test, RP_3

2.1 Derivation

Robertson & Wright (2009) show that, by using the moment condition satisfied by the MA component of the univariate representation, it is possible to derive an expression for θ of the form[39]

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

[39] The expression used here modifies the original expression in Robertson & Wright (2009) to allow for the different sign convention for ρ in this paper.

where

$$\kappa(\lambda, \rho, R_\mu^2) = \frac{\lambda - \rho F}{1 + \lambda^2 + F^2 - 2\lambda\rho F}$$

$$F(R_\mu^2, \lambda) = \sqrt{\left(1 - \lambda^2\right) \frac{R_\mu^2}{1 - R_\mu^2}}$$

To derive $\widehat{\theta}_z$ as defined in Section 2.2.3, using this formula, we need estimates of λ and R_R^2. The first is easy given the additional restriction $\phi = \lambda$. It also allows us to write (25), using $g(\lambda, \lambda) = 1$, as

$$f\left(R_z^2\right) = \rho_z^2 f\left(R_R^2\right) \tag{29}$$

where $f(x) = x/(1-x)$, so for a given Stambaugh correlation, inverting (29), we have

$$R_R^2 = \frac{f\left(R_z^2\right)}{\rho_z^2 + f\left(R_z^2\right)} \tag{30}$$

Hence we can write, given $\lambda = \phi$,

$$\theta\left(\lambda, \rho_R, R_R^2\right) = \theta\left(\phi, \rho_R, \frac{f\left(R_z^2\right)}{\rho_z^2 + f\left(R_z^2\right)}\right) = \theta_z\left(\phi, \rho_z, R_z^2\right). \tag{31}$$

where

$$\begin{aligned} \rho_R(\phi, \theta) &= sign(\phi - \theta); \quad \theta \neq \phi \\ &= -1; \quad \theta = \phi \end{aligned}$$

Under the restricted null the final expression in (31) implicitly defines a functional relationship that holds exactly in terms of population parameters. For purposes of estimation, in both Monte Carlo simulations and in our applications, we find the value of θ that satisfies $\widehat{\theta}_z = \theta_z\left(\phi, \widehat{\rho}_z, \widehat{R}_z^2\right)$,given $\rho_R = \rho_R\left(\widehat{\phi}, \widehat{\theta}_z\right)$

2.2 Sampling Properties of $\widehat{\theta}_z$

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 $\phi = \lambda$, as for our proposed test statistic RP_3.

[Table A1 about here]

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 (as 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 $\widehat{\phi}$ and $\widehat{\theta}_z$) to reflect the prior that the true values are both positive. The program then 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 univariate predictor used in deriving the test statistic RP_3 is, under the null, very close to being the true univariate 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 predictor hinges crucially upon the assumption that $\phi = \lambda$. Hence if the joint null underlying RP_3 is rejected this may be because z_t is not redundant, but it could also be because the assumption that $\phi = \lambda$ is incorrect (which, for sufficient differences between ϕ and λ will imply that the indirect estimate $\widehat{\theta}_z$ may be severely biased).

3 Simulation Methodology

3.1 Monte Carlos

The input parameters for each simulation are λ, θ, ρ_z and ϕ and a weight, τ, such that $R_x^2 = (1-\tau)\min(R_\mu) + \tau \max(R_\mu)$, where, $\min\left(R_\mu^2\right) = R_R^2(\lambda, \theta)$, as defined in (21), and, using Robertson & Wright (2009), Proposition 1, we have $\max\left(R_\mu^2\right) = R_R^2\left(\lambda, \theta^{-1}\right)$. By inverting the formula for $\theta\left(\lambda, \rho, R_x^2\right)$ in (28) this yields a value of ρ, the true Stambaugh Correlation. We then simulate underlying joint normal white noise innovations u_t v_t and ϖ_t (in equations (1), (2) and (12) respectively with the appropriate correlations. This in turn generates processes for μ_t, r_t ε_t (using $\varepsilon_t = (1-\theta L)^{-1}\left[(1-\lambda L) u_t + v_{t-1}\right]$) and z_t. Note that for Tables I to IV and A1 we could equally well simulate ε_t and generate r_t from the ARMA$(1,1)$, but for Table V 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.

3.2 Bootstrapped p-Values

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

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

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

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 r_t directly from the ARMA representation (since we do not need to make any assumption on the nature of the true predictor).

References

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*

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

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

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

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

Fama, Eugene F and French, Kenneth R (1988) ."Permanent and Temporary Components of Stock Prices". *Journal of Political Economy*, 96 (2),

Ferson W E , S Sarkissian and T Simin (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 (1994) *Time Series Analysis* 1994 Princeton University Press

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

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

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) The Limits to Stock Return Predictability, *Working Paper*

Robertson D and S Wright (2011) Univariate Bounds for Predictive Systems, *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 I. OLS-Based Tests of a Redundant AR(1) Predictor of an ARMA(1,1) returns 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, 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 I assumes that the returns process is ARMA(1,1): $r_t =(1-\theta L)/(1-\lambda L)\varepsilon_t$; and z_t is a redundant AR(1) predictor with $\phi = \lambda$ in the predictive system (1) to (3) and Stambaugh correlation ρ_z=corr(e_t,w_t)=0.9 in the predictive regression (10). Panel A gives the value of R_z^2 from Proposition 1 (where the inequality holds precisely since $\phi = \lambda$). Panel B shows the simulated size of a *t*-test on $\hat{\beta}$ in equation (9) in 5000 replications.

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

$\phi = \lambda$

		θ					
RP₁	λ	0	0.5	0.7	0.8	0.9	0.95
	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**
RP₂							
	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**
RP₃							
	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**

Table II shows the simulated size, in 5000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) returns process, r_t as described in Section III, Subsections B.1 to B.3. The processes for r_t and z_t are as for Table I. *T=200*

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

$$\phi = \frac{\lambda}{2}$$

		θ					
***RP*₁**	λ	0	0.5	0.7	0.8	0.9	0.95
	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**
***RP*₂**							
	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**
***RP*₃**							
	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**

Table III shows the simulated size, in 5000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) returns process, as described in Section III, Subsections B.1 to B.3. The processes for r_t and z_t are as for Table I, but with $\phi = \lambda / 2$. *T=200*

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

$$\phi = \lambda + \frac{1-\lambda}{2}$$

		θ					
RP₁	λ	0	0.5	0.7	0.8	0.9	0.95
	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**
RP₂							
	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**
RP₃							
	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**

Table III shows the simulated size, in 5000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) returns process, as described in Section III, Subsections B.1 to B.3. The processes for r_t and z_t are as for Table I, but with $\phi = \lambda + (1-\lambda)/2$. *T=200*

Table V Simulated rejection rates of the three tests under H_1: $\mu_t = \beta z_t$.

$$R_\mu^2 = 0.75\min\left(R_\mu^2\right) + 0.25\max\left(R_\mu^2\right)$$

		θ					
RP_1	λ	0	0.5	0.7	0.8	0.9	0.95
	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**
RP_2							
	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**
RP_3							
	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**
Memo: R-Squared of True State Variable for Returns (R_μ^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: True Stambaugh Correlation (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 V shows the rejection rate at a nominal 5% size, in 5000 replications, of the three tests that z_t is a redundant predictor, as described in Section III, Subsections B.1 to B.3, under the alternative hypothesis H_1: $\mu_t=\beta z_t$. The true state variable μ_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: $R_\mu^2 = 0.75\min\left(R_\mu^2\right) + 0.25\max\left(R_\mu^2\right)$, 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 state variable, derived from formulae in Appendix 1A.

Table VI. Testing for Redundant Predictors of Real Annual Stock Returns, 1901 2007 (107 observations)

	Predictor			
	log(price / dividend)	log(price / 10 year earnings)	log(Tobin's q)	log(price / total cash transfers)
Panel A. Predictor Characteristics				
Stambaugh Correlation ($\widehat{\rho_z}$)	0.835	0.983	0.914	0.552
Predictor AR(1) parameter ($\hat{\phi}$)*	0.922	0.928	0.905	0.694
Panel B. Nominal *p*-Values				
RP_1	0.898	0.885	0.567	0.068
RP_2	0.872	0.255	0.132	0.014
RP_3	1.000	0.888	0.440	0.011
$t(\beta)$	*0.141*	*0.034*	*0.009*	*0.002*
Panel C: Bootstrapped *p*-Values, r_t =ARMA(1,1)**				
RP_1	0.9218	0.9456	0.659	0.0994
RP_2	0.9038	0.8368	0.6876	0.0578
RP_3	0.9996	0.8984	0.6882	0.0178
$t(\beta)$	0.9474	0.9184	0.6088	0.0308
Panel D: Bootstrapped *p*-Values, r_t =white noise				
RP_1	0.8898	0.8546	0.5222	0.0706
RP_2	0.8524	0.2516	0.1712	0.0362
RP_3	0.9994	0.9078	0.5286	0.0164
$t(\beta))$	0.1838	0.0634	0.0148	0.0026

Table VI summarises tests that each of the four predictors shown is a redundant predictor. Panel A summarises the two key predictor characteristics. In Panels B, C and D, the first three lines show two-sided *p*-values for our three tests of predictor redundance, as described in Section III, Subsections B.1 to B.3; the fourth line shows 1 sided *p*-values for a t-test on β in the predictive regression (9).

Notes: * AR(1) estimates include bias-correction. ** Bootstrapped p-values for RP_3 set $\phi = \lambda$; $\theta=\theta_z(.)$ as in equation (15); bootstrapped p values for other tests use direct ARMA estimates of λ and θ . See Appendix 3 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, $\hat{\theta}_z = \theta_z\left(\hat{\phi}, \widehat{\rho_z}, \widehat{R}_z^2\right)$ derived from the properties of the predictive regression (9) and the predictor autoregression (3) under the joint null that $\phi = \lambda$ and z_t is redundant (see Section III, Subsection B.3 and Appendix 2). Results are shown for 5000 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 ϕ and θ_z but discard these if zero starting values yield a higher value of the estimated log likelihood.