The Dog and the Straw Man: Response to "Dividend Growth Does Not Help Predict Returns Compared To Likelihood-Based Tests: An Anatomy of the Dog."

John H. Cochrane[*]

April 22, 2020

_____

[*]Hoover Institution, Stanford University and NBER.

It is distressing to be accused of a mistake one has not made. It is more distressing when the contrary statements lie only a few pages deeper in one's paper, and more still when one has pointed this out twice already in writing. Such is the case with Hjalmarsson and Kiss (forthcoming).

Cochrane (2007) also started as a critique. The statistical evidence that returns are predictable is iffy. Correct small sample test statistics hover in the 90% to 95% range. The evidence that dividend growth is predictable is even weaker. So, examining return and dividend growth forecasts one at a time, it is easy to conclude that both dividend growth and returns are not predictable. But this is nonsense. If both dividend growth and returns are unpredictable, then the price-dividend ratio must be a constant, which it obviously is not. One needs to evaluate coherent views of the world, not one statistic at a time when an identity links those statistics. This is of course the sort of observation that no good statistician would make – we don't jump from failing-to-reject $\beta_r = 0$ in a univariate test, failing-to-reject $\beta_d = 0$ in a univariate test, to accepting that $\beta_r = 0$ and $\beta_d = 0$. If an urn has only red and black balls, one fails to reject that the urn has 1/4 red balls, and one also fails to reject that the urn has 1/4 black balls, that does not mean we accept that the urn has 1/4 red and 1/4 black balls. One recognizes they must add up. But this is only clear in retrospect.

Inspired by this observation to consider both return and dividend growth forecastability together, Cochrane (2007) starts with a simple observation. I set up a complete null hypothesis involving all the terms of a dividend growth, return, and dividend yield vector autoregression

$$r_t = a_r + \beta_r(d_{t-1} - p_{t-1}) + \varepsilon_t^r$$
$$\Delta d_t = a_d + \beta_d(d_{t-1} - p_{t-1}) + \varepsilon_t^d$$
$$d_t - p_t = a_{dp} + \phi(d_{t-1} - p_{t-1}) + \varepsilon_t^{dp},$$

in a way that respects the return identity,

$$r_t = -\rho(d_t - p_t) + \Delta d_t + (d_{t-1} - p_{t-1}).$$

This identity links coefficients,

$$\beta_r = \beta_d + (1 - \rho\phi), \tag{1}$$

errors,

$$\varepsilon_t^r = \varepsilon_t^d - \rho\varepsilon_t^{dp},$$

and data series themselves. One of each of set is redundant given the other two of that set.

The null is complete in that it specifies consistent values for both return and dividend growth predictability and all other parameters.

From (1), so long as $\rho\phi < 1$, then if returns are not predictable $\beta_r = 0$, dividend growth must be predictable $\beta_d > 0$. If a high price does not signal lower future returns, it must signal higher future dividends.

Cochrane (2007) Table 2 lists two such null hypotheses

$$\text{Null 1: } \beta_r = 0, \ \phi = 0.941, \beta_d = 1 - \rho\phi = -0.0931$$
$$\text{Null 2: } \beta_r = 0, \ \phi = 0.99, \beta_d = 1 - \rho\phi = -0.046.$$

The displayed nulls also include numerical values for the error covariance matrix $cov(\varepsilon\varepsilon')$. I then compute finite-sample distributions for the regression coefficients $\hat{\beta}_d$ and $\hat{\beta}_r$ under these nulls, displayed in Figure 1. Under these nulls, many simulations exceed the sample estimate of the return-forecast coefficient β_r. Returns this forecastable occur often. Very few simulations exceed the sample estimate of the dividend-growth forecast coefficient β_d. Dividends this *un-*forecastable (or forecastable in the wrong direction) are rare.

Why is there the difference between return and dividend growth forecasts? I trace it down to the fact that the return shock ε^r is highly negatively correlated with the dividend yield shock ε^{dp}, while the dividend growth shock ε^d is not well correlated with the dividend yield shock. Therefore, return forecasts inherit the small-sample distribution problems of the nearly-unit autocorrelation coefficient ϕ, while dividend growth forecasts have essentially standard OLS distributions. More generally, since $\beta_r = \beta_d + (1 - \rho\phi)$, one can phrase observations about the joint estimation of return and dividend forecasts equivalently as an observation about joint estimation of return predictability and autocorrelation ϕ of the dividend yield. The latter phrasing offers less economic intuition, but clearer statistical interpretation since ϕ estimation issues are well known.

How can one object to this? Hjalmarsson and Kiss (forthcoming) criticize "treating the autoregressive parameter of the dividend-price ratio as known" (abstract); "the choice to treat the OLS estimate of the AR parameter as the "true" value;" "the p-values... ignore the uncertainty coming from the fact that the value of the AR parameter in the original data is in fact unknown;" "the AR parameter in the simulation design is set equal to the OLS estimate from the data;" and so forth.

But that's what a null hypothesis *is.* That's what a distribution under the null *means.* One specifies "known" numerical values for the parameters, β_r, β_d and ϕ, along with the error covari-

ance matrix. There is nothing special about zero. In a null hypothesis we treat $\beta_r = 0$ as "known" just as we treat $\beta_d = -0.09$ as "known." One finds the distribution of estimates given those null values of parameters. One calculates whether sample estimates are unusual given this distribution generated under the null. The sample value of β_d is rarer *under this null* than the sample value of β_r.

So what is this paper's complaint? It sets up a straw man, which it calls "Cochrane's test," though you will find no advocacy of such a test in the paper, nor do they quote such. Suppose one repeated this procedure: In every sample, compute the estimate $\hat{\phi}$ in that sample, generate a distribution taking this $\hat{\phi}$ as the null, and test whether β_d is greater than the "data-dependent null" $1 - \rho\hat{\phi}$ (in quotes because it's a misnomer – nulls are numbers not data-dependent statitics). Well, that would be dumb. As the paper demonstrates, the sampling properties of this procedure are not desirable.

But one does not have to look far in Cochrane (2007) to see that's not " Cochrane's test." I considered many values of ϕ. Look at "Null 2" right next to "Null 1" in Table 2 and reproduced above. Look at the *second* row of Figure 1, that plots the sampling distribution of $\hat{\beta}_r$ and $\hat{\beta}_d$ under the null $\phi = 0.99$, rather than the "data-dependent" $\phi = 0.94$. Continue reading for a whole page to section 2.3 "Long-run estimates and tests" based on $\beta_r^{lr} \equiv \beta_r / (1 - \rho\phi)$ and $\beta_d^{lr} \equiv \beta_d/(1 - \rho\phi)$, of which I write "One great advantage of using long-horizon regression coefficients is that we do not need to choose between return and dividend-growth tests, as they give precisely the same results." Read section 3, a discussion of the joint distribution of β_r, β_d, and ϕ estimates which certainly do not treat ϕ as "known." Plow all the way to section 4.2 "results for different ϕ values," and in particular Table 5 which compiles probability values for a range of ϕ from $\phi = 0.90$ through $\phi = 1.01$. Examine section 4.3 which samples ϕ from a continuous prior and then computes distributions. How can one possibly actually read the paper and claim that it advocates a "test" in which one uses one and only one value of ϕ, the sample estimate, and treats that as "known?" And what would Hjalmarsson and Kiss have me do? When I reported distributions under every null from 0.90 to 1.01 in Table 5, should I have *omitted* the sample mean ϕ? Wouldn't that look like a curious omission?

All the paper has to say about any of this is a footnote 2 acknowledging the existence of long-run tests, though mis-stating their nature (they do use dividend growth regressions), and footnote 10 acknowledging that "Cochrane (2008) is clearly aware that the choice of ϕ is important and he considers several alternative values," though the body of the paper continues on the sins of using only the sample estimate and treating it as "known." There is also a short section 2.4 "altering the value of ϕ in the simulations" which writes "if one sets the AR parameter

in the simulations equal to some maximum feasible value for ϕ, say ϕ^{Max}, one ends up with a test that is similar to Lewellen's (2004) test, interpreted by Campbell and Yogo (2006) as a sup-bound test." But a point ϕ null is not the same thing as an upper bound on ϕ. If anything, my section 4.3 sampling of ϕ from a prior distribution with an upper bound is similar to Lewellen and Campbell and Yogo, both of which I cite. Most of all, Hjalmarsson and Kiss just throw in this as a minor referee-pacification acknowledgement and then plow on as if nothing had happened.

We do not "know" that ϕ equals its sample value (of course), but the whole point of Cochrane (2007) is that we do

> "...know something about the dividend-yield autocorrelation ϕ. The Wald test on b_r [β_r here] uses no information about other parameters. It is the appropriate test of the null $\{b_r = 0, \phi = \text{anything}\}$. But we know ϕ cannot be too big. If $\phi > 1/\rho \approx 1.04$, the present value relation explodes and the price-dividend ratio is infinite, which it also is obviously not. If $\phi \geq 1.0$, the dividend yield has a unit or larger root, meaning that its variance explodes with horizon. Economics, statistics, and common sense mean that if our null is to describe a coherent world, it should contain *some* upper bound on ϕ as well as $b_r = 0$, something like $\{b_r = 0, \|\phi\| < \bar{\phi}\}$. A good test uses information on *both* $\hat{b}_r$ and ϕ to evaluate such a null, drawing regions in $\{b_r, \phi\}$ space around the null $\{b_r = 0, \|\phi\| < \bar{\phi}\}$, and exploiting the fact that under the null $\hat{b}_r$ should not be too big *and* $\hat{\phi}$ should not be too big..."

The distribution of test statistics given a prior in which ϕ must be less than one, as implemented in Cochrane (2007), their relation to Lewellen and Campbell and Yogo's tests, and the optimal thing to do in this circumstance remain interesting avenues for research, that I wish this paper had pursued.

The paper also does not leave us with any guidance about what to do. As far as I can tell, Hjalmarsson and Kiss want us to go back to simple tests of the one-period OLS coefficient β_r. But now we are back to the beginning: If you just test OLS coefficients, you do not reject $\beta_d = 0$ and you do not reject $\beta_r = 0$. Yet the resulting view of the world is nonsense. How do we avoid this nonsense, and test a reasonable joint view of the world, using all the data at hand together? How do we better impose the prior view that $\phi < 1$? How do we better avoid the small-sample near-unit root estimation problems with ϕ, and how they bleed into β_r? Cochrane (2007) has some advice on these questions, which could be analyzed, critiqued and extended: Use long-run forecast estimates, implement a prior with $\phi < 1$, and so forth. But just go back to OLS and testing β_r in isolation is not fruitful advice at this stage.

As a critique of Cochrane (2007), then, the paper fails because it doesn't address the actual critiqued paper. As an alternative way to handle the issues raised by the paper, it doesn't even try. A null is a null and always specifies "known" parameters. My paper considered many values for the crucial parameter ϕ and never advocated a test based on one value, the OLS sample estimate. All this is clear if one reads past the first result.

Perhaps the paper's beef is not with the original but with "implementations" (abstract) of the ideas. Perhaps some of the implementers cited by name but without comment in footnote 1 slipped in to the mistake this paper alleges. But a critique paper must read and carefully document that people actually have made the mistake one alleges, not just cite "implementations" and a long list of footnotes.

The paper makes one correct criticism in footnote 6. Cochrane (2007) consistently misuses the word "power" to mean the probability of rejecting the null hypothesis. Given my persnicketyness about language and my criticisms of others who misuse, say "market efficiency" or write "hold" rather than "issue" bank capital, it is an embarrassing slip. However, the body of the paper continues with my incorrect use of the word, starting in the second paragraph. Let us all stop misusing language and not pass this canard on.

One may object to the sharp tone of this rejoinder. I was asked to review this paper for two other journals before it came to the *Critical Finance Review*. I provided essentially the above response. (For the record, I advised revision and publication. I do not believe a critiqued author should try to quash publication of critiques. The decision whether a critique has merit, and whether the issue is important enough to merit publication, must rest with editors and other reviewers.) The result was really only the pro-forma acknowledgements of footnotes 2 and 10. So, I believe the basic point of this paper is wrong, and the authors are reluctant to accept that fact. But perhaps I am wrong. Fortunately, the *Critical Finance Review* allows you to decide.

References

Cochrane, John H. 2007. "The Dog That Did Not Bark: A Defense of Return Predictability." *The Review of Financial Studies* 21 (4):1533–1575. URL https://dx.doi.org/10.1093/rfs/hhm046.

Hjalmarsson, Erik and Tamás Kiss. forthcoming. "Dividend Growth Does Not Help Predict Returns Compared To Likelihood-Based Tests: An Anatomy of the Dog." *Critical finance review* .