DIVIDEND PERSISTENCE AND RETURN PREDICTABILITY

John G. Powell
Jing Shi
Tom Smith

This Version: August 28, 2004
The Ferson, Sarkissian and Simin paper (2003a, JF) asks whether there is

“Spurious Regressions in Financial Economics?”

This paper provides a definitive answer;

Yes!

Return predictability of the dividend yield is a spurious result!
Statistical Background to the Paper:
A Spurious Correlation Illustrative Example!
(Neyman, 1952)

<table>
<thead>
<tr>
<th>Country</th>
<th>Women (*10k)</th>
<th># of Storks</th>
<th># of Babies</th>
<th>Stork-rate per 10k</th>
<th>Birth rate per 10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>2.00</td>
<td>10.00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>15</td>
<td>2.00</td>
<td>15.00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>2.00</td>
<td>20.00</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>3.00</td>
<td>10.00</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>3.00</td>
<td>15.00</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>3.00</td>
<td>20.00</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>4.00</td>
<td>10.00</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>15</td>
<td>4.00</td>
<td>15.00</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>4</td>
<td>20</td>
<td>4.00</td>
<td>20.00</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>4</td>
<td>15</td>
<td>2.00</td>
<td>7.50</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>9</td>
<td>40</td>
<td>1.80</td>
<td>8.00</td>
</tr>
<tr>
<td>46</td>
<td>6</td>
<td>8</td>
<td>35</td>
<td>1.33</td>
<td>5.83</td>
</tr>
<tr>
<td>47</td>
<td>6</td>
<td>8</td>
<td>40</td>
<td>1.33</td>
<td>6.67</td>
</tr>
<tr>
<td>48</td>
<td>6</td>
<td>8</td>
<td>45</td>
<td>1.33</td>
<td>7.50</td>
</tr>
<tr>
<td>49</td>
<td>6</td>
<td>9</td>
<td>35</td>
<td>1.50</td>
<td>5.83</td>
</tr>
<tr>
<td>50</td>
<td>6</td>
<td>9</td>
<td>40</td>
<td>1.50</td>
<td>6.67</td>
</tr>
<tr>
<td>51</td>
<td>6</td>
<td>9</td>
<td>45</td>
<td>1.50</td>
<td>7.50</td>
</tr>
<tr>
<td>52</td>
<td>6</td>
<td>10</td>
<td>35</td>
<td>1.67</td>
<td>5.83</td>
</tr>
<tr>
<td>53</td>
<td>6</td>
<td>10</td>
<td>40</td>
<td>1.67</td>
<td>6.67</td>
</tr>
<tr>
<td>54</td>
<td>6</td>
<td>10</td>
<td>45</td>
<td>1.67</td>
<td>7.50</td>
</tr>
</tbody>
</table>
1) The following is the correct model:

- \# of babies = \beta_0 + \beta_1 (# of storks) + \beta_2 (women)

An OLS regression is run

\# of babies = 10 + 0 (# of storks) + 5 (women)

**Conclusion:**
- Independence of the number of storks and babies!
2) Researchers, however, are often interested in the birth rate and therefore divide regression variables on both sides of the regression equation by the common “confounding” variable (“women”):

\[
\text{Birth rate} = 2.4 + 3.3 \text{ (stork rate)} \\
(5.76)
\]

The correlation between the birth rate and the stork rate is 0.63!

**Conclusions:**
- The presence of storks influences the birth rate
- Further, the removal of storks would be an effective method of birth control

The investigator therefore concluded that

“... although there is no evidence of storks actually bringing babies, there is overwhelming evidence that, by some mysterious process, they influence the birth rate”!
3) Some researchers pointed out that the birth rate regression equation is incorrect, and that the number of storks (not the stork rate) is the correct independent variable:

\[
\text{Birth rate} = 16.4 - 1.15 \times \text{(number of storks)} \\
(-7.15)
\]

The correlation between the birth rate and number of storks is -0.70!

**Conclusions:**
- The number of storks is actually inversely related to the birth rate!
- The best way to cut the birth rate is to increase the number of storks
4) To avoid the spurious correlation problem, the correct way to reformulate the first regression is to divide everything in the regression by the number of women, e.g.

- \(# \text{ of babies}=\beta_0 + \beta_1(\text{# of storks}) + \beta_2(\text{women})\)

Divide both sides by Women

- \(\frac{\text{# of babies}}{\text{women}}=\beta_2 + \beta_0(\frac{1}{\text{women}}) + \beta_1(\frac{\text{# of storks}}{\text{women}})\)

which leads to

- \(\frac{\text{# of babies}}{\text{women}}=5 + 10(\frac{1}{\text{women}}) + 0(\frac{\text{# of storks}}{\text{women}})\)
What is learned from this spurious correlation example?

- Opposite and incorrect inferences can be made from the same data set when ratios are used.

- Use of ratios can lead to spurious correlation even when all component variables that make up the numerators and the common denominator in a regression are independent.
Objectives of the paper

This paper conclusively answers the question as to whether there is “Spurious Regressions in Financial Economics?” by showing that return predictability of the dividend yield is a spurious result:

- spurious regression occurs when highly persistent variables are used in a time series regression model to predict/explain dependent variables that are at least partially persistent

The paper also demonstrates that the spurious regression problem is strongly reinforced by a spurious correlation when the dependent variable (returns) and independent variable (dividend yield) are both constructed from the same underlying highly persistent variables (the share index level and the dividend level)

A data mining simulation procedure is utilized to take account of this problem

The paper shows that standard dividend behaviour models are also affected by the spurious regression problem

A correctly reformulated Lintner (1956) dividend first-difference model is derived to indicate how this econometric problem can be avoided
Dividend Yield Return Predictability

Many studies find that dividend yields predict a substantial amount of the cross-sectional and time series variation of stock return (see Table 1)

Dividend yields have also been used to predict long horizon returns (e.g. Fama and French, 1988; Campbell, Lo and MacKinlay, 1997)

Only a few studies indicate that there is not a strong statistical relationship between dividend yields and stock returns

  e.g.  
Bossaerts and Hillion (1999)  
Goyal and Welch (2003)  
Stambaugh (1999)
How is dividend yield return predictability tested?

A standard approach to test for dividend yield return predictability is to use the regression equation

\[
    r_{t+1} = \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left( \frac{D_t}{P_t} \right) + \varepsilon_{t+1},
\]

where

\( D_t \) is the level of real annual dividends during the twelve months preceding time \( t \), and
\( P_t \) is the real stock index level at time \( t \)

(see Panel A, Table 2)

Ferson, Sarkissian and Simin (2003a) indicate that the dividend yield’s predictive power is questionable when account is taken of the spurious regression problem combined with data mining;

- but do not conclusively prove it
How does this study prove it?

Starting point:

Step 1. Theoretical justification of dividend yield return predictability relies upon a high current dividend level relative to the share index level predicting higher future returns:

\[ r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[ \frac{D_t}{P_t} \right] + \varepsilon_{t+1} \]

If so, substituting a constant level of dividends \( (c) \) for the dividend term \( (D_t) \) in the numerator of the dividend yield variable \( (D_t / P_t) \) in the regression equation should reduce or eliminate return predictability

\[ r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_2 \left[ \frac{c}{P_t} \right] + \varepsilon_{t+1} \]

(see Panel B, Table 2)

Surprizingly, the Adjusted R\(^2\) actually increases instead of falling:
- this is a first indication that dividend yield return predictability might be spurious, since the constant level of dividends \( (c) \) is even more persistent
Step 2:

To check the robustness of the result, a pseudo dividend yield series is created to replace the actual dividend yield series
  - The numerator is again non-stochastic

(see Panel C, Table 2)

Once again, the adjusted $R^2$ rises, not falls
  - this again hints that something spurious is occurring, since the pseudo-dividend series is extremely persistent

Step 3:

Therefore, check the persistence properties of the variables

*What do we get?*

(See Table 3)

- dividend series are highly persistent, as are $D_t/P_t$ and $c/P_t$
- contemporaneous correlation between dividend yield and return is actually negative
Step 4:

A data-mining simulation procedure is used to provide a cut-off $R^2$ (see Foster, Smith and Whaley, 1997, JF; Ferson, Sarkissian and Simin, 2003a, JF)

- The moments and the serial correlation properties of the regression variables are estimated for each data series

- Uncorrelated dependent and independent variables with the same serial correlation properties and sample moments are then simulated for a time period equal to the sample length (1927 to 1996)

- A regression is run on the simulated series

- The process is repeated 1,000 times

- The adjusted $R^2$s are recorded for each regression and ranked from lowest to highest

- The 95$^{th}$ percent $R^2$ is then reported as the cut-off $R^2$ and compared to the actual adjusted $R^2$ - the 95$^{th}$ percentile is chosen to represent the tendency of researchers to data mine by reporting only regression results that provide the highest adjusted $R^2$s
What do we get?

For instance, for the EW NYSE Index, an $R^2$ of 4.53% would be expected to be obtained by regressing subsequent returns against uncorrelated independent variables that are equally as persistent as the dividend yield variable ($D_t/P_t$)

- actual adjusted $R^2 < \text{cut-off } R^2 \rightarrow \text{spurious result}$

- p-value for each regression is insignificant (proportion of the simulated adjusted $R^2$s that exceed the actual adjusted $R^2$)

Dividend yield return predictability regression results are due to spurious regression!

So the paper is now finished (dividend yield return predictability has been proven to be spurious)?
Hold on!!! *Something else is also going on here!*

The data-mining simulation procedure assumes the variables in the regression equation are independent:

\[
r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[ \frac{D_t}{P_t} \right] + \epsilon_{t+1}
\]

But, the assumption is not appropriate here!

Why?

Return and dividend yield variables both come from the same underlying variables (P and D).
What can be done?

- The simulation procedure is modified to recognize the dependency of both the return and dividend yield variables on the underlying share index and dividend level series.

- So, rather than simulating the dividend yield series $D_t/P_t$ using the properties of the dividend yield series, the dividend ($D_t$) and share index ($P_t$) series are instead simulated using their own estimated properties.

- Dividend yield ($D_t/P_t$) observations are then calculated using the simulated dividend and share index values.
What do we get?

- Much higher simulated modified cut-off $R^2$s in Table 2

- The increase in the modified cut-off $R^2$s is caused primarily by the presence of the highly persistent price index common denominator on both sides of the regression equation

Recall:

$$r_{t+1} = \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[ \frac{D_t}{P_t} \right] + \epsilon_{t+1}$$

- implies a spurious correlation problem interacts with the spurious regression problem, leading to extremely spurious results
The results can be compared to the return predictability literature recommendations regarding spurious regression and regression coefficient bias:

Ferson, Sarkissian and Simin, 2003b, say current return predictability studies are subject to spurious regression, but return predictability can be saved by detrending predictor variables:

- this will create a less persistent independent variable that is less correlated with the regression error

Campbell and Yogo, 2003 also recommend detrending the dividend yield to create a less noisy predictor variable, and Lewellen, 2004 tests (in effect) a detrended dividend yield variable;

So, should the dividend yield be detrended to test for dividend yield return predictability?

Subsequent returns are therefore regressed against a dividend yield that is stochastically detrended using the Ferson, Sarkissian and Simin 2003b recommended procedure (see Table 4)
What do we get?

• An increase in the dividend yield actually predicts lower, not higher subsequent returns!

• The change in sign indicates a spurious correlation effect, as in the Kronmal (1993) incorrectly specified ratio regressions (recall storks and the birth rate)!

• “Temporary mispricing” is also unlikely to be the underlying reason for return predictability, since higher mispricing predicts higher, not lower, subsequent returns!

• Dividend yield return predictability is spurious!
Dividend Behaviour Models

Persistence properties of dividends are responsible for dividend yield return predictability, so does dividend persistence also cause spurious regression in dividend behaviour models?

The two most important dividend behaviour models are examined:

Marsh and Merton (1987)

$$\log\left(\frac{D_{t+1}}{D_t}\right) + \frac{D_t}{P_{t-1}} = \psi_0 + \psi_1 \log\left(\frac{P_t + D_t}{P_{t-1}}\right) + \psi_2 \log\left(\frac{D_t}{P_{t-1}}\right) + \epsilon_{t+1}$$

Lintner (1956)

$$D_t = \theta_0 + \theta_1 E_t + \theta_2 D_{t-1} + \epsilon_t \rightarrow$$

$$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \epsilon_t$$

(the time $t$ dividend choice is modelled only in relation to information that is observable at time $t$)
What do we get?

The results indicate that spurious regression plays an extremely important role in these standard dividend behaviour regression models which regress highly persistent dividends against lagged dividends and other terms (Table 5)
A reformulated Lintner (1956) first-difference dividend behaviour model

Reformulate entirely in terms of first differences to avoid spurious regression:

\[ D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t \]

\[ D_{t-1} = \theta_0 + \theta_1 E_{t-2} + \theta_2 D_{t-2} + \varepsilon_{t-1} \]

\[ D_t - D_{t-1} = \theta_1 (E_{t-1} - E_{t-2}) + \theta_2 (D_{t-1} - D_{t-2}) + (\varepsilon_t - \varepsilon_{t-1}) \]

Incorporate the Marsh and Merton (1987) permanent earnings explanation of dividends by substituting in the price index level for earnings (since price equals the present value of future earnings):

\[ P_t = a + b E_t \]

\[ D_t - D_{t-1} = \theta_0 + \theta_1 (P_{t-1} - P_{t-2}) + \theta_2 (D_{t-1} - D_{t-2}) + (\varepsilon_t - \varepsilon_{t-1}) \]
What do we get?

- The adjusted $R^2$ greatly exceeds the cut-off $R^2$

- The model implies that changes in the aggregate level of dividends are explained by lagged share price innovations as well as lagged dividend innovations, as predicted by Marsh and Merton (1987)
Conclusions

Both the spurious regression and spurious correlation problems are present in dividend yield return predictability and dividend behaviour regression models

Dividend yield return predictability is spurious!

A reformulation of the Lintner (1956) dividend behaviour model entirely in terms of first differences provides a model that is not subject to spurious regression, and also directly incorporates the Marsh and Merton (1987) permanent earnings explanation of dividend behaviour