A NOTE ON "PREDICTING RETURNS WITH FINANCIAL RATIOS"

Amit Goyal
Emory University

Ivo Welch
Yale School of Management

A Note on
"Predicting Returns with Financial Ratios"*

Amit Goyal
Goizueta Business School
Emory University[†]

Ivo Welch
Yale School of Management
Yale Economics Department
NBER[‡]

December 16, 2003

Abstract

This note reinterprets methods that seek to use the aggregate dividend price ratio to predict aggregate stock market returns; specifically, methods which use information about time-varying changes in the dividend-price ratio process to improve the prediction equation. It argues that the empirical evidence is still too weak to suggest practical usefulness of these estimators.

In *Predicting returns with financial ratios*, Lewellen (2003) introduces a promising new test to improve on the ability of financial ratios to predict stock returns. This test relies on the fact that dividend yields are close to non-stationary, and have become more so since 1973. The paper has received much attention, even before publication. (For example, Campbell and Yogo (2003) generalize the methodology.) A reader not too familiar with the data would likely conclude that there is no question that dividend yields can help investors predict stock returns, at least prior to 1995.

Yet, in Goyal and Welch (2003), we had also documented that dividend yields have become more non-stationary over time, even indistinguishable from a random walk as of December 2002. We then implemented a test statistic that directly uses Campbell and Shiller's (1988) identities to instrument not only the time-varying properties of the dividend yield, but also the time-varying changes in the dividend growth process. In contrast to Lewellen, we had concluded that neither the dividend yield, nor our instrumented prediction could help investors predict the equity premium.

Predicting the equity premium may well be the most important issue in finance, so it is important to reconcile the two perspectives. Both perspectives have evaluated the same data through similar lenses (time-varying changes in the dividend process) and still have come to opposite conclusions. This note explains why, and gives an alternative perspective on the performance of Lewellen's improved test.

Stambaugh (1999) introduces an underlying process of

$$r_t = \alpha + \beta \cdot dp_{t-1} + \epsilon_{r,t} \tag{1}$$

$$dp_t = \mu + \rho \cdot dp_{t-1} + \epsilon_{dp,t} \quad, \tag{2}$$

where r here is the simple[1] stock return and dp is the log dividend price ratio. The goal is to estimate the slope coefficient β in the return equation (1). The correlation between ϵ_r and ϵ_{dp} violates the OLS assumption that the independent variable dp_{t-1} be uncorrelated with the errors $\epsilon_{r,t}$. Therefore, the simple OLS estimate of β is upwardly biased. Denoting the estimator of β by $\hat{\beta}_T$ and the estimator of ρ by $\hat{\rho}_T$, where

---

[1]Using log-stock returns instead of simple returns does not matter at monthly frequency for our results. We use the simple stock return to remain directly comparable with Lewellen (2003).

T is the number of observations, Stambaugh shows that the estimated beta should be adjusted using the empirical estimate of ρ as follows:

$$\hat{\beta}_T^{\text{adjusted}} = \hat{\beta}_T^{\text{ols}} - \frac{\text{cov}(\epsilon_r, \epsilon_{dp})}{\text{var}(\epsilon_{dp})} \times \text{Estimated Bias in } \hat{\rho}_T \tag{3}$$

In Lewellen's (and therefore our) full sample period, the OLS beta is about 0.009, the covariance-variance ratio term is about –0.905, and the estimated rho is around 0.997. Stambaugh derives a frequentist correction for bias in ρ as $-(1+3\hat{\rho}_T)/T$. This implies a bias in the prediction beta to be

$$\hat{\beta}_T^{\text{Stambaugh}} = \hat{\beta}_T^{\text{ols}} + \frac{\text{cov}(\epsilon_r, \epsilon_{dp})}{\text{var}(\epsilon_{dp})} \times \left[\frac{1 + 3\hat{\rho}_T}{T} \right] \quad . \tag{4}$$

Therefore, Stambaugh reduces the OLS beta by roughly $3.6/T$. With 660 observations, this is about 0.0055.

Like the Stambaugh correction, the Lewellen estimator is also essentially an intelligent shrinkage estimator, but it uses information about dividend process autocorrelations in a different fashion. Lewellen rules out explosive bubbles, and therefore estimates the bias in ρ as $\hat{\rho}_T - 0.9999$. This implies that he can change the forecasting beta to:

$$\hat{\beta}_T^{\text{Lewellen}} = \hat{\beta}_T^{\text{ols}} + \frac{\text{cov}(\epsilon_r, \epsilon_{dp})}{\text{var}(\epsilon_{dp})} \times [0.9999 - \hat{\rho}_T] \quad . \tag{5}$$

Given the empirical estimates, Lewellen reduces the OLS beta by around 0.0025. Because the contemporaneous correlation between innovations is negative, and the estimated $\hat{\rho}_T$ is lower than 1, $\hat{\rho}_T$ enters negatively in Stambaugh correction, but positively in Lewellen correction. The higher Lewellen estimates ρ (closer to a random walk in the dividend process), the more he shrinks the prediction beta towards the OLS beta. Intuitively, with $\hat{\rho}_T$ increasing in the sample period (1946–2000), and especially after 1995, Lewellen's test can find evidence that the dividend yield can predict where earlier papers had seemed to find only lack of significance. An interesting difference between Stambaugh and Lewellen is that as T goes to infinity, Stambaugh suggests zero shrinkage to the OLS beta, while Lewellen suggests a constant shrinkage.

Figure 1 plots both the betas and the dividend-price process autocorrelation. Lewellen's beta has the appealing feature that in the face of drastically changing dividend process autocorrelation, the recursive estimated out-of-sample beta remains remarkably stable, when compared to the Stambaugh or the OLS beta. But the figure also

2

shows that Lewellen's betas are lower than other beta estimates for most of the sample period. Thus, when compared to these other betas, his evidence of stronger predictability is based on statistical grounds (lower estimation standard errors of his beta), and not on economic grounds (higher beta estimates).

Cochrane (2001, p. 406) suggests that an intuitive way of looking at the betas from an annual regression of returns on (simple) dividend price ratio. Because the mean dividend price ratio for our sample period is 3.80%, this can roughly be accomplished by multiplying all our beta numbers by $12/3.80\% \approx 316$. This gives the OLS, Stambaugh, and Lewellen betas as 2.89, 1.16, and 2.09, respectively.[2] Cochrane argues that the benchmark beta for no predictability (when dividend price ratios are random walks) is 1.0, and that for complete predictability (when dividend price ratios are not persistent at all) is 25.0. Therefore, Stambaugh betas imply least predictability and OLS betas imply the most predictability.

In Goyal and Welch, we relied directly on Campbell and Shiller's (1988) identity to derive

$$\hat{\beta}_T^{\text{GW}} = 1 - \kappa \cdot \hat{\rho}_T + \hat{\beta}_{\Delta\text{Div}}^{\text{ols}} \quad , \tag{6}$$

where κ is $1/(1 + \overline{e^{dp}})$, which can be calibrated to about 0.9968 with U.S. monthly data, and $\hat{\beta}_{\Delta\text{Div}}^{\text{ols}}$ is the slope coefficient in an OLS regression of (log) dividend growth on dividend price ratio.[3] The higher autocorrelation $\hat{\rho}_T$ can also reduce the estimated beta coefficient. This specification is consistent with Cochrane (2001, p. 402), who states that "To believe in lower predictability of returns, you must either believe that dividend growth really is predictable, or that the d/p ratio is really much more persistent than it appears to be." On theoretical grounds, our instrumentation is most appealing, because it takes both sources into account.

Table 1 compares the empirical performances of these methods, keeping Lewellen's sample period (1946–2000 and 1946–1995), data frequency (monthly), and specific data (value-weighted NYSE stock returns). The left columns of the table de-

---

[2]Actual annual regressions of returns on simple dividend price ratios give OLS, Stambaugh, and Lewellen betas as 1.12, 0.02, and 0.32 respectively.

[3]Strictly speaking, identity (6) is valid only for log returns and dividend price ratio computed using monthly dividends. As explained later, we use simple returns to be consistent with Lewellen (2003). We also follow the standard practice to compute the dividend price ratio using the last 12-month dividends but divide the mean dividend price ratio by 12 in parameter κ to make it consistent with monthly frequency ($1/(1 + \overline{d/p}/12) \approx 0.9968$, where d/p is the twelve-month moving average un-log-ed dividend price ratio.). Goyal-Welch (2003) use annual data where both of these concerns are not an issue.

scribe in-sample residuals. The historical mean error is the first data row—naturally zero for all methods used here. More interestinglly, if used for one-month ahead prediction, the prevailing mean would have yielded an RMSE of 4.07% and an absolute forecast error of 3.14%. The naïve OLS technique does a tiny bit better than the historical mean.[4] All modification (Stambaugh, Lewellen, and Goyal-Welch) techniques have about the same performance as the OLS forecast. If we end the sample in 1995, OLS, Goyal-Welch, Stambaugh, and Lewellen can all significantly outperform the historical mean *in-sample*. If we end the sample in 2000, however, OLS, Goyal-Welch, and Lewellen perform a little better than Stambaugh's estimator.

The right columns of Table 1 describe out-of-sample performance. When we end the sample in December 2000, Lewellen's technique performs best on the RMSE metric, where it can outperform the historical mean's RMSE by 0.008% per month. This is neither economically significant, even if aggregated to one year, nor statistically significant in a Diebold and Mariano (1995) t-test on the RMSE difference ($t = 0.65$).[5] OLS and Stambaugh perform 0.013% per month worse than Lewellen's method. On the MAE metric, the prevailing mean outperforms all dividend ratio techniques (and sometimes in a statistically significant fashion). Our final metric is the frequency of months in which a method beats the historical mean. Both Stambaugh and Lewellen beat the historical mean 48.0% of the time, while OLS and GW beat it 47.4%. On all metrics, despite its theoretical appeal, the instrumented Goyal-Welch procedure performs no better than the naïve OLS beta. This is a reflection of the fact that there is almost no difference between $\hat{\beta}_T^{\text{ols}}$ and $\hat{\beta}_T^{\text{GW}}$. Furthermore, it is not just the monthly returns: in Goyal-Welch (2003), we explored annual forecasts, and found similarly poor out-of-sample predictive ability.

If we end the sample in 1995 instead of 2000, it is the simple OLS predictor which does best on the RMSE metric, followed closely by Lewellen's and GW's statistic. Because the Lewellen net performance is smoother than OLS's (see Figure 1), on a RMSE difference t-test, the Lewellen out-of-sample performance t-statistic reaches 1.81, which corresponds to a two-tailed p-value of 0.071 and a one-tailed p-value of 0.036. Although OLS and GW perform better than Lewellen's test, their relative performance

---

[4]This is a reflection of the low $\overline{R}^2$. A low $\overline{R}^2$ should not be overinterpreted, because $\overline{R}^2$ of course must be low in monthly regressions. However, in our earlier paper, we had used annual forecasts, where in-sample $\overline{R}^2$ was more respectable (but not out-of-sample performance).

[5]The Diebold and Mariano statistics' properties are well known and well behaved. We had also experimented with jackkniving these errors, and found virtually identical statistical significance.

advantage is not smooth enough to outperform the historical mean in a statistically significant manner. Stambaugh's method cannot significantly outperform, either. On the MAE metric, the historical mean again performs best. On our final metric, Stambaugh and Lewellen can beat the historical mean 49.0% of the time, while OLS could beat it only 48.3% of the time.[6]

Figure 2 repeats our favorite out-of-sample diagnostic from Goyal-Welch (2003). We plot the cumulative out-of-sample (absolute or squared) forecast error of a method minus the forecast error of the historical mean. When a line increases, the dividend-yield method outpredicts the historical mean. When a line decreases, the historical mean outpredicts the dividend-yield method. The figure shows that from about 1975 to 1994, by-and-large, the OLS/GW methods (virtually indistinguishable) performed reasonably well on the RMSE metric, better than Lewellen, Stambaugh, and the historical mean. Ending the sample anywhere around 1990–1995 maximizes the relative predictive ability of all dividend ratio methods. In 1995, Lewellen both beats OLS, and has statistically superior performance relative to the historical mean. This is because Lewellen's technique is steadier, which gives it lower standard errors and thus the aforementioned advantage in statistical significance. On the MAE metric, no method seems to reliability outpredict the historical mean. OLS/GW significantly underperform the mean until 1970. Beginning around 1994, all dividend techniques underperform.

The most important metric may be the economic significance of these techniques. Fortunately, both RMSE and MAE have intuitive magnitudes. A typical relative performance advantage of a magnitude of around 0.01 percent per month is very modest. If we use the trading strategy in Breen, Glosten, and Jagannathan (1989), we can detect *no* gains to using any of these techniques. In our original paper, we used annual returns and similarly failed to find economic relevance. For an investor, this level of predictive ability, even if it were statistically better, is unlikely to be practically useful.

In conclusion, the tests in Goyal-Welch (2003), Lewellen (2003), and Campbell and Yogo (2003) incorporate changes in the dividend yield process in different ways into

---

[6]Our Goyal-Welch estimator, while recognizing the economic sources of predictability, ignores the statistical bias in computing the autocorrelation coefficient $\hat{\rho}_T$. If we use $\hat{\rho}_T$ as 1.0 in equation (6) instead of estimating it, Goyal-Welch estimator's RMSE increases to 4.138% (from 4.134%) in sample period ending 1995. Interestingly, however, on a RMSE difference t-test, the t-statistic reaches 1.95, which corresponds to a two-tailed p-value of 0.051 and a one-tailed p-value of 0.025.

the prediction equation, but with only moderate success. Lewellen's paper is a step in the right direction, and can outperform the prevailing historical stock return mean *out-of-sample* on at least one out-of-sample performance metric in a subperiod—perhaps a first. But a reader of this literature should not be left with the impression that it is unambiguously clear that predicting stock returns with these particular dividend ratio methods would have yielded superior or even statistically significant investment results. The data appear so ambiguous and perhaps uninformative that a Bayesian investor might end up with a posterior that is very close to her priors—at least on monthly and annual forecasting horizons. One reason to believe in dividend price ratios to forecast returns or dividend growth is Cochrane's well-known identity, equation (7), which states that

$$dp_t \quad = \quad \sum_{i=0}^{\infty} \kappa^i (r_{t+1+i} - \Delta\text{Div}_{t+i+1}) + \text{constant} \tag{7}$$

$$= \quad \kappa \cdot dp_{t+1} \quad + \quad (r_{t+1} - \Delta\text{Div}_{t+1}). \tag{8}$$

Because the dividend price ratio has not predicted dividend growth, one might believe that it should predict stock returns. But, equation (8) points out why optimism may be premature: the dividend price ratio has failed in the common empirical practice which forecast stock returns on monthly or annual horizons, because the dividend yield has not so much forecast either dividend growth or stock returns, but primarily because it has forecast itself.

In sum, Lewellen's careful statistical analysis of biases in estimating the autocorrelation of dividend price ratio results in a *lower* predictive in-sample beta on returns, which argues for lower *economic* predictability. The out-of-sample performance does not convince us, either. Given the sum total of the evidence, the conclusion—as to whether these dividend yield forecasting techniques for stock returns have succeeded empirically—should, in our opinion, remain a matter of the reader's evaluation of and philosophy about empirical tests. Caveat Emptor.

References

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance* 44(5), 1177–1189.

Campbell, John Y, and Motohiro Yogo, September 2003, "Efficient Tests of Stock Return Predictability," Working Paper, Harvard University and NBER.

Campbell, John Y. and Robert Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies* 1(3), 195–228.

Cochrane, John H., 2001, *Asset Pricing*, Princeton University Press, Princeton.

Diebold, Francis X., and Roberto S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economics Statistics* 13(3), 253--263.

Goyal, Amit, and Ivo Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49(5), 639–654.

Lewellen, Jonathan, August 2003, "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, forthcoming.

Stambaugh, Robert, 1999, "Predictive Regressions," *Journal of Financial Economics* 54(3), 375–421.

Table 1: Predictive Performance: Statistics on Forecast Errors

Forecasts Ending in 2000

Statistic	In-Sample				Out-of-Sample				Beats
	Mean	RMSE	MAE	$\overline{R}^2$	Mean	SDV	RMSE	MAE	
Historical Mean	0.000%	4.072%	3.139%	0.000%	-0.018%	4.164%	4.160%	**3.170%**	Mean
OLS	0.000%	**4.061%**	**3.134%**	0.409%	-0.348%	4.154%	4.165%	3.225%	256/540
Stambaugh	0.000%	4.065%	3.135%	0.208%	-0.115%	4.168%	4.165%	3.197%	259/540
Lewellen	0.000%	4.062%	3.134%	0.366%	-0.183%	4.152%	**4.152%**	3.186%	259/540
Goyal-Welch	0.000%	4.061%	3.134%	0.409%	-0.334%	4.154%	4.164%	3.221%	256/540

Forecasts Ending in 1995

Statistic	In-Sample				Out-of-Sample				Beats
	Mean	RMSE	MAE	$\overline{R}^2$	Mean	SDV	RMSE	MAE	
Historical Mean	0.000%	4.057%	3.119%	0.000%	0.026%	4.157%	4.152%	**3.146%**	Mean
OLS	0.000%	**4.027%**	3.115%	1.312%	-0.230%	4.131%	**4.133%**	3.182%	232/480
Stambaugh	0.000%	4.030%	3.110%	1.173%	-0.022%	4.149%	4.145%	3.159%	235/480
Lewellen	0.000%	4.036%	**3.110%**	0.902%	-0.090%	4.138%	4.134%	3.153%	235/480
Goyal-Welch	0.000%	4.027%	3.114%	1.310%	-0.222%	4.132%	4.134%	3.180%	232/480

All models predict the value–weighted market rate of return, beginning in 1946. The first forecast is predicted in 1956 to allow for 10 years to elapse before the first forecast is made. "Beats Mean" is the number of months in which the absolute forecast error of a method is less than absolute historical prevailing mean forecast error. Boldface means best performer.

Figure 1: Recursive Beta Coefficients and Time-Varying Dividend Price Ratio Process

This figure plots the recursively computed betas from equation

$$r_t = \alpha + \beta \cdot dp_{t-1} + \epsilon_{r,t},$$

where r is the value-weighted NYSE return, and dp is the log dividend price ratio. Beta adjustments are given by

$$\hat{\beta}_T^{\text{Stambaugh}} = \hat{\beta}_T^{\text{ols}} + \frac{\text{cov}(\epsilon_r, \epsilon_{dp})}{\text{var}(\epsilon_{dp})} \cdot \left[\frac{1 + 3\hat{\rho}_T}{T} \right]$$

$$\hat{\beta}_T^{\text{Lewellen}} = \hat{\beta}_T^{\text{ols}} + \frac{\text{cov}(\epsilon_r, \epsilon_{dp})}{\text{var}(\epsilon_{dp})} \cdot [0.9999 - \hat{\rho}_T]$$

$$\hat{\beta}_T^{\text{GW}} = 1 - \kappa \cdot \hat{\rho}_T + \hat{\beta}_{\Delta\text{Div}}^{\text{ols}}.$$

where the autocorrelation is estimated from the equation

$$dp_t = \mu + \rho \cdot dp_{t-1} + \epsilon_{dp,t}$$

Al estimates are computed recursively using the beginning date till all the data upto current period. The overall sample period is 1945 to 2000. Goyal-Welch beta is virtually indinguishable from the OLS beta and is not plotted separately.
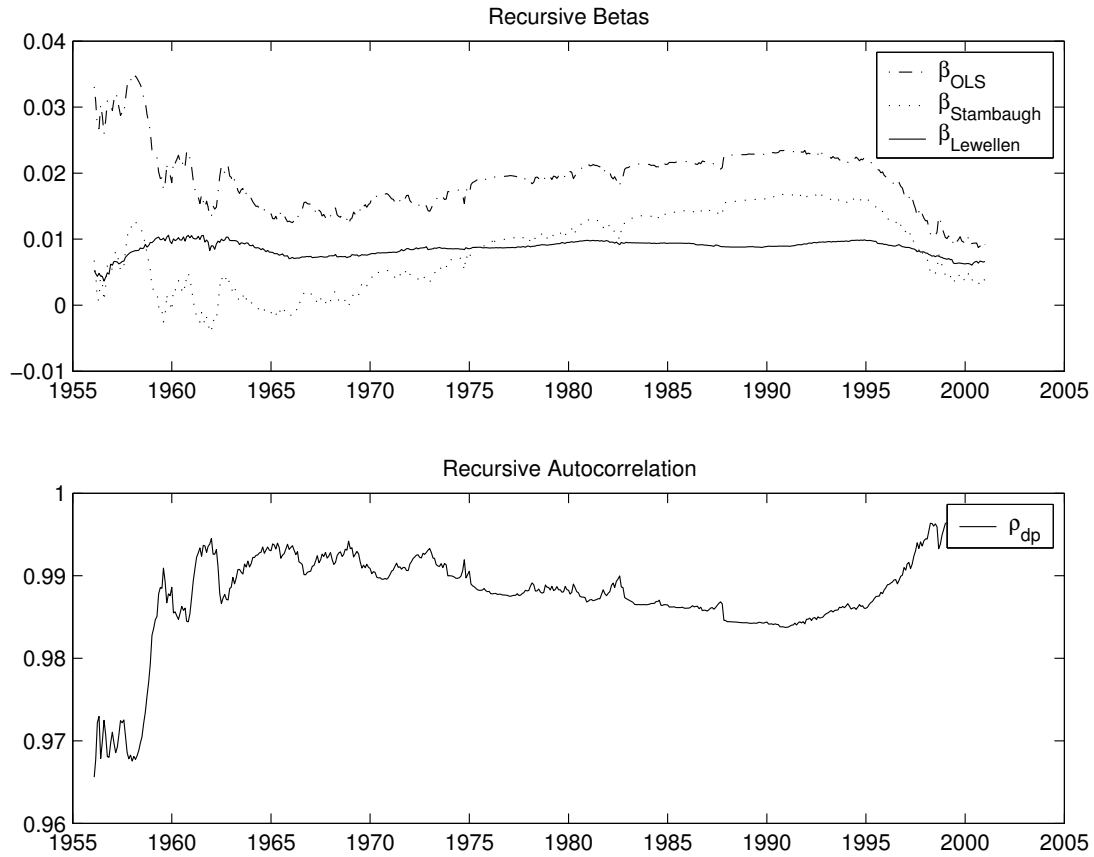

9

Figure 2: Out-of-Sample Cumulative Performance Relative to the Historical Mean

This figure plots

$$\text{Net-SSE}(T) = \sum_{t=1956}^{T} \text{SE}(t)^{\text{Prevailing mean}} - \text{SE}(t)^{\text{Dividend Model}}$$

where $SE(t)$ is either the squared or the absolute out-of-sample prediction error in period t. For a month in which the slope is positive, the dividend ratio regression model predicted better than the unconditional historical average *out-of-sample*. The vertical line is the end of 1995. The OLS and GW estimates are virtually identical.

10