Journal of Banking & Finance 36 (2012) 2632–2640

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Does the choice of estimator matter when forecasting returns?

Joakim Westerlund *, Paresh Kumar Narayan

Financial Econometrics Group, School of Accounting, Economics and Finance, 70 Elgar Road, Burwood Highway, Deakin University, Melbourne, VIC 3125, Australia

ARTICLE INFO

Article history:
Received 4 October 2011
Accepted 5 June 2012
Available online 13 June 2012

JEL classification:
C22
C23
G1
G12

Keywords:
Predictive regression
Stock return predictability
Heteroskedasticity
Predictor endogeneity

ABSTRACT

While the literature concerned with the predictability of stock returns is huge, surprisingly little is known when it comes to role of the choice of estimator of the predictive regression. Ideally, the choice of estimator should be rooted in the salient features of the data. In case of predictive regressions of returns there are at least three such features; (i) returns are heteroskedastic, (ii) predictors are persistent, and (iii) regression errors are correlated with predictor innovations. In this paper we examine if the accounting of these features in the estimation process has any bearing on our ability to forecast future returns. The results suggest that it does.

1. Introduction

The use of financial ratios to predict returns has attracted much attention, and many studies have shown that ratios such as dividend–price, price–earnings, dividend–payout, and book-to-market are able to predict future stock returns (see for example Campbell and Shiller, 1988a,b, 1998; Fama and French, 1988; Kothari and Shanken, 1997; Lamont, 1998; Chen, 2009).[1] The prevailing tone in the literature is perhaps best summarized by Lettau and Ludvigson (2001, p. 842): "It is now widely accepted that excess returns are predictable by variables such as dividend–price ratios, earnings–price ratios, dividend–earnings ratios, and an assortment of other financial indicators."

However, while there is some evidence of predictability, there is also evidence to the contrary, and in recent years this has become a major source of tension in the literature. In particular, while most evidence in favor of predictability are based on in-sample tests, the little out-of-sample evidence that exists is mainly negative. This disparity makes an overall assessment of return predictability difficult. The following passage, taken from Welch and Goyal (2008, p. 1455), serves as an illustration: "The literature is difficult to absorb. Different articles use different techniques, variables, and time periods. Results from articles that were written years ago may change when more recent data is used. Some articles contradict the findings of others. Still, most readers are left with the impression that 'prediction works' – although it is unclear exactly what works."

In an attempt to sort out "what works", Welch and Goyal (2008) reconsider much of the empirical evidence reported in the literature. They find that most commonly used predictors are unable to deliver consistently superior out-of-sample forecasts of the US equity premium relative to a simple forecast based on the historical average. Similar conclusions are drawn by Bossaerts and Hillion (1999), who did not find any evidence of out-of-sample predictability in a collection of industrialized countries for a number of predictors. Hence, based on these results it would appear as that nothing works.

Amid this debate, in this paper we ask to what extent the out-of-sample forecasting performance is influenced by the choice of estimator of the predictive regression? This is a relevant question, because some of the differences in the literature may well be due to estimation problems. Indeed, as Rapach et al. (2010, p. 288) point out:

> The lack of consistent out-of-sample evidence in Welch and Goyal indicates the need for improved forecasting methods to better establish the empirical reliability of equity premium predictability.

* Corresponding author. Tel.: +61 3 924 46973; fax: +61 3 924 46283.
E-mail addresses: j.westerlund@deakin.edu.au (J. Westerlund), paresh.narayan@deakin.edu.au (P.K. Narayan).

[1] There is a related branch of the literature which considers currency volatility predictability (see Scott and Tucker, 1989) and whether stock markets predict real activity (see Choi et al., 1999).

0378-4266/$ - see front matter
http://dx.doi.org/10.1016/j.jbankfin.2012.06.005

One of the potential problems is that many of the predictors are highly persistent, and their innovations tend to be correlated with return innovations. As Stambaugh (1999) shows, this combination of persistency and endogeneity can be quite devastating in that it induces a small-sample bias in the conventional ordinary least squares (OLSs) estimator. Another potential problem is that returns are highly volatile. Indeed, one of the most well-documented features of financial time series is that returns are highly heteroskedastic. Thus, even if returns are predictable, to the extent that the heteroskedasticity is strong enough to dwarf the signal coming from the predictors, this might be difficult to detect.

Our approach to this issue is as follows. We consider monthly US time series data covering the period January 1871 to December 2008. The variables included are excess returns, the dividend–price ratio (DP), dividend yield (DY), the earnings–price ratio (EP), and the dividend–payout ratio (DE). Three estimators are applied to these data; (1) the OLS estimator, (2) the bias-adjusted OLS (AOLS) estimator of Lewellen (2004), and (3) the feasible generalized least squares (FGLSs) estimator of Westerlund and Narayan (2011). The first estimator is included because it is the workhorse of the industry, the second is included because of its popularity, and the third is included because it is relatively new and therefore represents the state of the art. The difference between the estimators lies in their ability to accommodate possible heteroskedasticity, endogeneity and persistency of the regressor. While the OLS estimator is least general and does not account for any of these features, FGLS estimator is most general and accounts for all three features. The AOLS estimator lies somewhere in between and corrects for the endogeneity and persistency of the regressor.

The results suggest that the choice of estimator does make a difference, and that the FGLS estimator generally performs best. The results are shown to be significant not only from a statistical point of view, but also from an economic point of view. Moreover, while there is some variation coming from the choice of predictor and forecasting horizon, our results seem to be rather robust to the choice of sample period.

2. Econometric discussion

2.1. The predictive regression

As mentioned in Section 1, certain empirical features that characterize predictive regressions can make inference difficult and it is therefore important that these features are acknowledged already at the modeling stage. Our starting point is the following system of equations:

$$r_{t+h} = \alpha + \beta x_t + \epsilon_{t+h}, \tag{1}$$

$$x_{t+1} = \mu(1-\rho) + \rho x_t + \varepsilon_{t+1}, \tag{2}$$

where $|\rho| \leqslant 1$. This is the prototypical predictive regression model that has been widely used in the financial economics literature, in which x_t is a variable believed to be able to predict the h-period-ahead value of excess returns, r_{t+h}. In our case, x_t is either DP, DY, EP or DE. Thus, in this model testing the null hypothesis of no predictability is equivalent to testing the restriction that $\beta = 0$. As in previous studies, it is reasonable to assume that the correlation between ϵ_t and ε_t is negative. For example, if x_t is DY, then an increase in the stock price will lower dividends and raise returns. In order to capture endogenous effects of this sort, the following relationship between the error terms is assumed:

$$\epsilon_t = \gamma\varepsilon_t + \eta_t, \tag{3}$$

where ε_t and η_t are mean zero and uncorrelated with each other. The variances of these errors are typically assumed to be constant over time. However, this does not fit well with the fact that returns are almost always found to be heteroskedastic. The most popular approach by far to model this type of behavior is to assume an autoregressive conditional heteroskedasticity (ARCH) model, which in the case of η_t can be written as

$$\mathrm{var}(\eta_t|I_{t-1}) = \sigma^2_{\eta t} = \lambda_0 + \sum_{j=1}^{q}\lambda_j\eta^2_{t-j}, \tag{4}$$

where I_t is the information available at time t. A similar equation is assumed to apply to $\mathrm{var}(\varepsilon_t|I_{t-1}) = \sigma^2_{\varepsilon t}$. Provided that $\lambda_0 > 0$ and $\sum_{j=1}^{q}\lambda_j < 1$, this implies that the unconditional variance of η_t can be written in terms of the coefficients of (4) as $\mathrm{var}(\eta_t) = \sigma^2_\eta = \frac{\lambda_0}{\left(1-\sum_{j=1}^{q}\lambda_j\right)}$ with $\mathrm{var}(\varepsilon_t) = \sigma^2_\varepsilon$ having a similar representation. The unconditional variance of the composite error term ϵ_t can now be written as $\mathrm{var}(\epsilon_t) = \sigma^2_\epsilon = \gamma^2\sigma^2_\varepsilon + \sigma^2_\eta$.

Having laid out the model of interest, we now turn to the effects of endogeneity, persistency and ARCH:

- The correlation between ϵ_t and ε_t is given by $\rho_{\epsilon\varepsilon} = \gamma\frac{\sigma_\varepsilon}{\sigma_\epsilon}$ and is, as already mentioned, a source of major complication. The reason is that if $\rho_{\epsilon\varepsilon} \neq 0$, then x_t is no longer exogenous, thereby violating one of the most important OLS assumptions; if x_t is endogenous the OLS estimator of β is no longer unbiased.
- The main effect of the persistency of x_t is to aggravate the OLS bias caused by the endogeneity. In fact, as Stambaugh (1999) shows, the OLS bias is given by $-\gamma(1+3\rho)/T$, suggesting that while decreasing in T, the bias is increasing in γ and ρ. Moreover, the persistency is only a problem to the extent that x_t is indeed endogenous such that $\gamma \neq 0$.
- While unattended endogeneity and persistency are matters of bias, unattended ARCH is a matter of efficiency. Indeed, one of the most well-known results from classical regression theory is that OLS is inefficient in the presence of heteroskedasticity. In agreement with this, Westerlund and Narayan (2011) show that there are important efficiency gains to be made by accounting for the ARCH information.

2.2. The estimators

In view of the above discussion, the question of how to estimate the predictive regression in (1) naturally arises. One way is to simply ignore the issues of bias and inefficiency altogether, and run OLS. However, given the potential problems involved, this approach is clearly suboptimal, and Lewellen (2004) has therefore proposed a bias-adjusted estimator that deals with the first issue. The idea is to make (1) conditional on ε_t by substituting from (2) and (3), leading to the following augmented test regression:

$$r_{t+h} = \theta + \beta x_t + \gamma(x_{t+h} - \rho x_{t+h-1}) + \eta_{t+h}, \tag{5}$$

where $\theta = \alpha - \gamma\mu(1-\rho)$. But while this removes the correlation between the regression error and the regressors, since ρ is unknown, (5) is not really feasible, and Lewellen (2004) therefore suggests replacing the true ρ with a "guess". Let ρ_0 denote this guess. The feasible version of (5) can be written as

$$r_{t+h} = \theta + \beta_0 x_t + \gamma(x_{t+h} - \rho_0 x_{t+h-1}) + \eta_{t+h}, \tag{6}$$

where $\beta_0 = \beta - \gamma(\rho - \rho_0)$ has the interpretation of a "bias-adjusted slope coefficient", which can be estimated by simply applying OLS to (6). The FGLS estimator is based on the same regression and is therefore very similar in spirit. One difference between the two estimators is the treatment of the persistency of x_t. In particular, while Lewellen (2004) assumes that $\rho_0 = \rho = 0.9999$, Westerlund and Narayan (2011) assume that $\rho = 1 + \frac{c}{T}$, where $c \leqslant 0$ is a drift