

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

A Simple Nonlinear Predictive Model for Stock Returns

Biqing Cai and Jiti Gao

November 2017

Working Paper 18/17

A Simple Nonlinear Predictive Model for Stock Returns

BIQING CAI

School of Economics

Huazhong University of Science and Technology, China

AND JITI GAO*

Department of Econometrics and Business Statistics

Monash University, Australia

October 13, 2017

Abstract

In this paper, we propose a simple approach to testing and modelling nonlinear predictability of stock returns using Hermite Functions. The proposed test suggests that there exists a kind of nonlinear predictability for the dividend yield. Furthermore, the out-of-sample evaluation results suggest the dividend yield has nonlinear predictive power for stock returns while the book-to-market ratio and earning-price ratio have little predictive power.

Key words: Hermite Functions, Out-of-Sample Forecast, Return Predictability, Series Estimator, Unit Root

JEL Classification Numbers: C14; C22; G17.

*Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: jiti.gao@monash.edu.

1 Introduction

Whether the stock returns are predictable or not is one of the most important questions in finance. It is not only of interest to the practitioner but also has implications for financial models of risk and returns. Meanwhile, the issue of return predictability is also very controversial because people usually reach different conclusions based on different data sample periods as well as different statistical methods.

In the last few decades, based on in-sample evidence, a consensus has emerged among economists that the equity risk premium seems to vary considerably over the business cycle and perhaps at lower frequencies as well. Campbell (2000) concludes that the stock returns contain significant predictable components. At the same time, several authors expressed concern that the apparent predictability of stock returns might be spurious because many of the predictor variables in the literature are highly persistent. Nelson and Kim (1993), and Stambaugh (1999) point out that persistence leads to biased coefficients in predictive regressions if innovations in the predictor variable are correlated with the returns.

Moreover, the out-of-sample evidence of stock return predictability is much weaker. Goyal and Welch (2003; 2008) find that the predictive ability of a variety of popular financial and economic variables in the literature does not hold up in out-of-sample forecasting exercises. They conclude that “the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power.” Under the widely held view that predictive models require out-of-sample validation (e.g., Campbell (2008) argues: “The ultimate tests of any predictive model is its out of sample performance.”), the findings of Goyal and Welch (2003; 2008) cast doubts about the reliability of return predictability.

In the study of predictive regression, linear models are now pervasive in the literature, to name only a few, Stambaugh (1999), Lewellen (2004), Campbell and Yogo (2006). However, the linear model has some pitfalls. First of all, the linear model is usually derived from a simplified present value model, e.g. Campbell and Shiller (1988), which may not be a good approximation in reality. Engstrom (2003) shows that under a reasonable theoretical setting which simply imposes the no arbitrage condition, the linear predictive relationship may not hold and structural treatment of risk may lead to nonlinear relationship between return and dividend yield. Furthermore, it is well known that the predictive financial ratios are highly persistent that we can not reject that they contain unit roots. However, the stock return is far less persistent, and is often regarded as a stationary process (or even a white noise process). A linear model will lead to an unbalanced relationship between the left-hand-side and the right-hand-side variables in predictive regression which is called model inconsistency by Granger (1995).

Recently, there have been more and more interests on nonlinear predicting relationships. Gonzalo and Pitarakis (2012) propose a regime switch predicting model and argue that the strength of predictability to vary across economic episodes, such as expansions and recessions. Kasparis et al. (2015) develop a nonparametric test for predictability. Campbell and Thompson (2008) point out that imposing sign restriction on the regression coefficients may help predicting stock return. Gonzalo and Pitarakis (2012) also show that the return predictability has different characteristics in different economic regimes. Chen and Hong (2010) show that the nonparametric model can outperform the linear model. Lee et al. (2015) further show that a combination of nonparametric model and sign restriction can make the model outperform historic mean model. However, the proposed nonlinear models are not shown to catch the co-existence of highly persistent predictors and far less persistent stock returns.

In this paper, we propose that the predictive function is integrable (see Park and Phillips (1999; 2001) for more discussion) and use the Hermite functions to detect and estimate the predictive function. A nonlinear integrable function of a unit root nonstationary time series significantly reduces the nonstationarity of the original unit root time series and thus partially addresses the model imbalance issue in the linear predictive regression. Cai (2012) shows that the commonly used diagnostic tools will fail to detect such kind of nonstationarity which explains why the stock return is far less persistent. A figure illustration of this point can be found in Phillips (2015). We apply the proposed estimation method to estimate the stock return predictive function and use the estimated function to predict the next-period stock return. Our empirical results support an existence of nonlinear predictive power of the dividend yield, while the predictive power of book-to-market ratio and earning-price ratio is very weak. A detailed analysis is given in Section 3.

The main contribution of this paper is as follows. First of all, we propose an integrable model for the predictive regression and provide a general methodology for testing and estimating the nonlinear function. Thus, it partially addresses the model imbalance issue inherent in the linear predictive regression and gives a closed-form description of the nonlinear function. To be specific, we propose using a simple nonlinear predictive model of the form:

$$r_{t+1} = \alpha + (\beta_0 + \beta_1 x_t) e^{-\frac{x_t^2}{2}} + e_{t+1},$$

where the nonlinear function is expanded using the first two terms of Hermite series. Then it is obvious that a precise functional form is given and furthermore, because of the role of the exponential term $e^{-\frac{x_t^2}{2}}$, the model becomes much less unbalanced than a simple linear model of the form: $r_{t+1} = \theta_0 + \theta_1 x_t + e_{t+1}$ in the case where x_t behaves like a unit root process. Secondly, our empirical results provide some strong and robust evidence of existence

of nonlinear predicting power of the dividend yield.

The organisation of this paper is as follows. In Section 2, we propose the nonlinear predictive model and discuss the testing and estimation based on Hermite functions. Section 3 applies the approach to test for predictability and estimate the stock return predictive function. We then compare the out-of-sample performance of the integrable model with the historic mean and linear models. Section 4 concludes the paper. The technical assumptions and proofs are collected in an appendix.

2 The Integrable Predictive Model: Testing and Estimation

2.1 Integrable Predictive Model

In this section, we discuss the integrable model and its estimation. In the conventional predictive regression analysis, since Mankiw and Shapiro (1986), Nelson and Kim (1993) and Stambaugh (1999), the following model is widely used:

$$r_t = \alpha + \beta x_{t-1} + \varepsilon_t, \quad (2.1a)$$

$$x_t = \phi + \rho x_{t-1} + u_t, \quad (2.1b)$$

where x_t is a predicting variable and r_t is the stock market return. Popular predicting variables include: the dividend yield (e.g., Campbell and Shiller (1988a; 1988b)); the earning-price ratio (e.g., Lamont (1998)) and book-to-market ratio (e.g., Kothari and Shanken (1997)). We refer to Welch and Goyal (2008) for a comprehensive discussion about including more predicting variables. However, it is well known that the predictive financial ratios are nearly unit root or unit root processes, i.e., ρ in (2.1b) is equal to or nearly 1. Since Nelson and Kim (1993) and Stambaugh (1999), researchers have noticed that most of the predicting financial ratios are unit root or nearly unit root processes. The persistence leads to biased coefficients in predictive regressions if the innovations in the predictor variable are correlated with returns and the standard t-test for predictability therefore has incorrect size. Thus, many researchers have proposed more efficient tests for stock return predictability under (2.1a), for example, Lewellen (2004) and Campbell and Yogo (2006). However, as we have mentioned in the introduction, in the linear model when $\beta \neq 0$, there may exist a model imbalance because of the variables with different persistence involved in the left-hand-side and then the right-hand-side variables. This motivates us to propose some nonlinear predicting functions to catch the data characteristics. Specifically, we consider the

following model:

$$r_t = \alpha + f(x_{t-1}) + \varepsilon_t, \quad (2.2)$$

where f is an integrable function, and this paper proposes using an integrable function of the form: $f(x) = (\beta_0 + \beta_1 x) e^{-\frac{x^2}{2}}$. The integrable function contains all the probability density functions as special cases. We refer to Park and Phillips (1999, 2001) for more discussion on this class of nonlinear functions. As shown in Cai (2012), the integrable function can generate the time series pattern of the left-hand-side and right-hand-side variables in predictive regression. Thus it should be a suitable model for predictive regression. A similar discussion can be found in Phillips (2015). From a theoretical point of view, the linear model should be too aggressive considering the regressor contains a unit root because when $\beta \neq 0$ (i.e., there exists predictability), and the return should also be a unit root process. The integrable function serves the role of compressing the strong signal of the unit root process involved. And it is also consistent with the intuition that when there is predictability, it should be weak, otherwise, there will exist arbitrage opportunities.

In practice, the functional form of the the integrable function is unknown, so that we use Hermite series to approximate the integrable function. When the data is stationary, series estimation as a nonparametric alternative to the conventional kernel-based method has been examined in the literature, e.g., Newey (1997), Andrews (1991). When the regressor is a unit root process, Cai and Gao (2013) have established an asymptotic theory for the Hermite series based estimation method for a class of integrable functions. For illustration purpose, in the next subsection, we provide some background knowledge of the Hermite functions.

2.2 Preliminaries of Hermite Functions

The Hermite functions form a complete basis in $L^2(\mathbb{R})$, where \mathbb{R} is the real line and $L^2(\mathbb{R})$ denotes the space of square-integrable functions. They have advantages in the nonlinear nonstationarity setup because unlike the stationary data case where orthogonality is defined according to the marginal (stationary) distribution, in the nonstationary case, the orthogonality is defined with respect to the Lebesgue measure. Because when the regressor is a unit root process, Lebesgue measure is the invariant measure. For more information about the invariant measure for a null recurrent process (which covers the simplerandom walk process as a special case), see, for example, Karlsen and Tjøstheim (2001). Thus, Hermite series are naturally orthogonal in the nonstationary case without fully specifying the data generating process, i.e., we don't need to impose restrictions on the distribution of the innovations of the unit root process. It is well known that the usage of orthogonal basis functions allows us to extract information of different pieces of the unknown function effectively. In this

subsection, we introduce some basic properties about Hermite functions.

Let $\{H_i(x)\}_{i=0,1,2,\dots}$ be the orthogonal Hermite polynomial system with respect to the weight function $\exp(-x^2)$. It's well known that $\{H_i(x)\}_{i=0,1,2,\dots}$ is a complete orthogonal system in Hilbert Space $L^2(\mathbb{R}, \exp(-x^2))$, in which the conventional inner product is used, i.e., $\langle f, g \rangle = \int f(x)g(x) \exp(-x^2)dx$. In addition, the orthogonality of the system can be expressed as follows:

$$\langle H_i(x), H_j(x) \rangle = \sqrt{\pi}2^i i! \delta_{ij},$$

where δ_{ij} is the Kronecker delta function. Put $\varphi(x) = \exp(-\frac{1}{2}x^2)$ and define

$$h_i(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^i i!}} H_i(x) \quad \text{and} \quad F_i(x) = h_i(x)\varphi(x), i \geq 0.$$

Then, $\{F_i(x)\}_{i=0,1,2,\dots}$ are the so-called Hermite series or Hermite functions in the literature. $\{F_i(x)\}_{i=0,1,2,\dots}$ are complete orthonormal in $L^2(\mathbb{R})$, such that any continuous function $f(x)$ in $L^2(\mathbb{R})$ has an expression of the form

$$f(x) = \sum_{i=0}^{\infty} \theta_i F_i(x),$$

where $\theta_i = \int f(x)F_i(x)dx$.

The Hermite functions can be listed as follows: $F_0(x) = \frac{1}{\sqrt[4]{\pi}} \exp(-\frac{1}{2}x^2)$; $F_1(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2}} \times 2x \times \exp(-\frac{1}{2}x^2)$; $F_2(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{8}} \times (4x^2 - 2) \times \exp(-\frac{1}{2}x^2)$; $F_3(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{48}} (8x^3 - 12x) \times \exp(-\frac{1}{2}x^2)$, and so on.

2.3 Test for Predictability

In this section, we discuss how to test for integrable-type predictability. The null hypothesis is:

$$H_0 : r_t = \alpha + \varepsilon_t, a.s. \text{ for } t = 1, 2, \dots, T, \quad (2.3)$$

against an alternative hypothesis of the form:

$$H_1 : r_t = \alpha + f(x_{t-1}) + \varepsilon_t, a.s. \text{ for } t = 1, 2, \dots, T. \quad (2.4)$$

For this purpose, we first regress the left hand side variable on the constant and the Hermite functions $F_0(x), F_1(x), \dots, F_{p-1}(x)$, the estimator of the respective coefficients are

$$\hat{\theta} = \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^T \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})r_t,$$

where $F(x_{t-1}) = [F_0(x_{t-1}), F_1(x_{t-1}), \dots, F_{p-1}(x_{t-1})]^\tau$ and $\bar{F} = \frac{1}{T} \sum_{t=1}^T F(x_{t-1})$.

This paper then focuses on the case where p is fixed. We then propose the following Wald test statistic:

$$Q = \frac{\hat{\theta}^\tau \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right) \hat{\theta}}{\hat{\sigma}^2}, \quad (2.5)$$

where $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \left(r_t - \hat{\mu} - \hat{\theta}^\tau F(x_{t-1}) \right)^2$. We now state the following results; their proofs are given in Appendix A.

Theorem 1. *Let Assumptions A.1–A.3 listed in Appendix A below hold. Under H_0 , we have $Q \rightarrow_D \chi_p^2$.*

Theorem 2. *Let Assumptions A.1–A.3 hold. If, in addition, $\int_{-\infty}^{\infty} F(x)f(x)dx \neq 0$, then under H_1 we have $Q \rightarrow_P \infty$ with rate $O_P(T^{1/2})$.*

Remark: Theorems 1 and 2 show that the Wald test has the standard χ^2 distribution under the null while under the alternative, the divergence rate is $O_P(T^{1/2})$, which is slower than that in the stationary time series case. In applications, we need to select the truncation parameter p . In an unreported Monte Carlo simulation, we find that the proposed test has reasonable size and good power against a wide range of alternatives in the finite sample case where p is 1, 2 or 3. So in our empirical application below, we choose p to be 1, 2 and 3.

It should be pointed out that there are some corresponding results for the conclusions of Theorems 1 and 2 when p is allowed to vary with T . In the case where $p \rightarrow \infty$, for example, we have that $\frac{Q-p}{\sqrt{2p}} \rightarrow_D N(0, 1)$.

2.4 Estimating Predictive Regression Function

The proposed test in (2.5) is suitable for an in-sample test. As we have mentioned in the introduction, the most controversial issue in the predictive regression is the out-of-sample performance. So that we need to discuss how to model the predictive regression for forecasting purposes.

When f in (2.2) is square-integrable, it can be written as

$$r_t \approx \alpha + \sum_{i=0}^{\infty} \theta_i F_i(x_{t-1}) + \varepsilon_t. \quad (2.6)$$

We then use an approximate model of the form:

$$r_t = \alpha + \sum_{i=0}^{p-1} \theta_i F_i(x_{t-1}) + \varepsilon_t. \quad (2.7)$$

Using OLS we can get the estimated coefficients $\hat{\alpha}$ and $\hat{\theta}_i$, $i = 0, \dots, p-1$. In the case where p is allowed to satisfy $p \rightarrow \infty$ at certain rate as $T \rightarrow \infty$, Cai and Gao (2013) derive the asymptotic property of the series estimator and show that the finite sample performance is reasonable even though p is being treated as a fixed integer in practice. Compared with the kernel-based nonparametric estimation, the series based estimator enjoy several advantages: it is convenient in computation, it is easy to be extended to additive model, and it has a close-form expression.

As is well known in the nonparametric estimation literature, the selection of the truncation parameter p is of critical importance. When p increases, the bias of the estimator is smaller while the variance of the estimator is larger and vice versa. It is shown in Cai and Gao (2013) that to guarantee the asymptotic theory, p can be chosen to be $\lfloor cT^\gamma \rfloor$ with $0 < \gamma < 3/22$, where $\lfloor \cdot \rfloor$ denotes taking integer part.

In this paper, similar to Gao et al. (2002) we adopt a data driven choice of the truncation parameter which minimizes the following criterion: $\frac{\hat{\sigma}^2}{(1-p/T)^2}$, where $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\alpha} - \sum_{i=0}^{p-1} \hat{\theta}_i F_i(x_{t-1}))^2$. Intuitively, this criterion is similar to the adjusted- R^2 in the regression model where a penalty is imposed to the increase of the number of regressors. This penalty is even more important in the forecasting context as is well known that good in-sample performance does not necessarily lead to good out-of-sample performance when the model is complicated.

3 Empirical Application

In this section, we apply the proposed Hermite series based test and estimation method to study the predictive regression. Specifically, we first test whether there exists a type of nonlinear predictability for the dividend yield. If the answer is YES, we further estimate the model via Hermite series estimation and compare the performance of one-step-ahead forecast of our model with both the historic mean model and the linear model. Finally, we also examine whether two other widely used return predicting variables: Book-to-Market ratio and Earning Price ratio have nonlinear predictability.

3.1 Data Description

The predicting variables are dividend yield, book-to-market ratio and earning-price ratio, which are commonly used in the literature, see e.g. Lewellen (2004), Welch and Goyal (2008). Price and dividends come from Center for Research in Security Prices (CRSP) dataset ¹. We

¹We thank Professor Lewellen for providing us with his dataset.

focus on NYSE equal and value-weighted indices to be consistent with the existing studies. The dividend yield is calculated monthly on the value-weighted NYSE index, and it is defined as dividends paid over the prior year divided by the current level of index. Thus the dividend yield is based on the rolling windows of annual dividends.

The returns data are from January, 1946 to December, 2007 with a total number of 744 data points. In this paper, we study the predictability of both the nominal return and the excess return. $Vwny$ denotes value-weighted NYSE stock return (nominal return), $ewny$ denotes equal-weighted stock return, $evwny$ denotes excess value-weighted stock return (real return or excess return) which is defined by value-weighted return minus T-bill rate, and $eewny$ denotes excess equal-weighted excess return. Dy and $\log(dy)$, respectively, denote the dividend yield and the logarithm of the dividend yield, having 744 data points. Bm denotes book-to-market ratio with 744 data points². Ep denotes earning-price ratio with sample size 536³.

The sample mean, standard error, skewness, kurtosis and the first order autocorrelation coefficient are summarised in Table 1.

Table 1: Summary Statistics

variables	mean	std	skewness	kurtosis	auto
$vwny$	0.0099	0.0401	-0.3830	4.8756	0.0414
$ewny$	0.0112	0.0470	-0.1863	6.8156	0.1417
$evwny$	0.0062	0.0403	-0.4119	4.8167	0.0483
$eewny$	0.0075	0.0472	-0.2163	6.7615	0.1463
dy	0.0359	0.0128	0.4047	2.5653	0.9904
$\log(dy)$	-3.3948	0.3757	-0.3398	2.4429	0.9927
bm	0.6060	0.2046	0.0703	1.9156	0.9872
$\log(bm)$	-0.5641	0.3676	-0.4561	2.2697	0.9908
ep	0.0657	0.0268	0.9577	2.9629	0.9877
$\log(ep)$	-2.7977	0.3833	0.3150	2.3562	0.9876

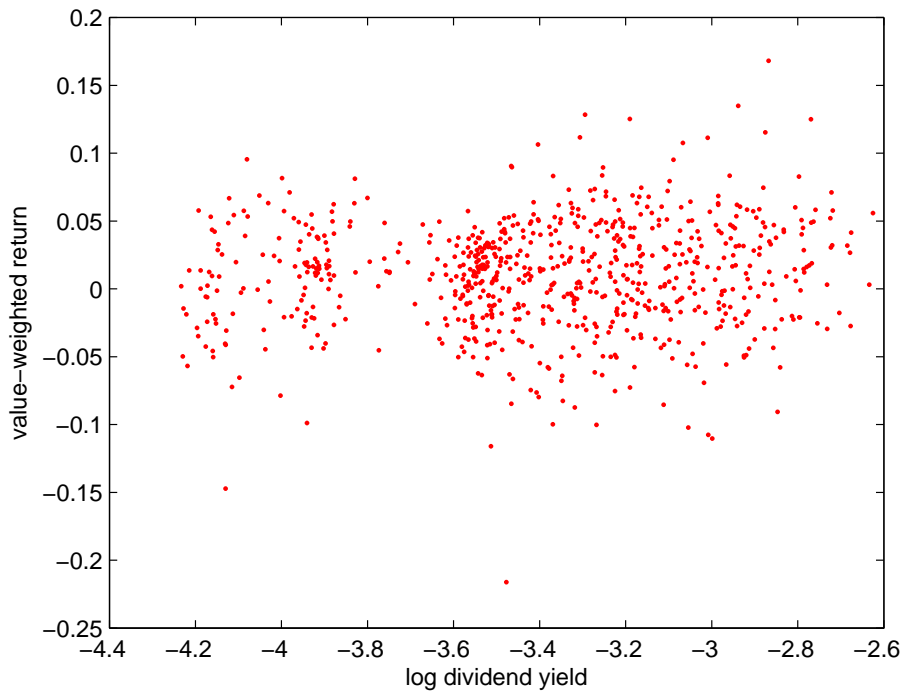
From Table 1, we can find that there are weak serial correlations in stock returns. Especially, the value-weighted return series is nearly white noise process. At the same time, the

²Book equity uses data set (<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>) available at Ken French's website.

³It is from Compustat, thus only starts from 1963.

forecasting dividend yield, book-to-market ratio and earning-price ratio are highly persistent with the first order sample autocorrelation close to 1. This suggests that there is no linear relationship between stock return and its predictors, but can not rule out the possibility that the the predictors can forecast stock return nonlinearly. The sample kurtosis of the return series suggests the stock return has heavy tail as is well known in the literature. The conventional ADF and KPSS tests suggest there are unit roots in dividend yield, book-to-market ratio and earning-price ratio, while the stock returns are stationary. The scatter plots of value-weighted stock return v.s. logarithm of dividend yield, logarithm of book-to-market ratio, and logarithm of earning-price ratios suggest there are no explicit patterns of forecasting relationship of these financial ratios to the stock return (see Figure 1 for the scatter plot of value-weighted stock return v.s. logarithm of dividend yield).

Figure 1: Dividend Yield and Stock Return



These empirical facts suggest that the relationship of the stock return and its predictors should not be linear and the stock return inherently contains a sizable unpredictable component, so the best forecasting models can explain only a relatively small part of stock returns. So that an integrable model rather than a linear model can serve as the benchmark model in stock return predicting study. In the following study, we will examine whether there exists a kind of integrable predictability.

3.2 In-Sample Test of Predictability

In this section, we use the Wald test to test stock return predictability. The explanatory variable will be dividend yield or log dividend yield and the return can be value-weighted NYSE stock return or equal-weighted NYSE stock return. The results of the test are shown in Table 2:

Table 2: Testing Predictability: 1946-2007

Truncation parameter	dy		log(dy)	
	ewny	vwny	ewny	vwny
p=1	6.4043 (0.0114)	7.8697 (0.005)	6.6751(0.0098)	8.1687 (0.0043)
p=2	6.7012 (0.0351)	8.0849 (0.0176)	6.6759 (0.0355)	8.2292 (0.0163)
p=3	6.7224 (0.0813)	8.4108 (0.0382)	6.8223 (0.0778)	8.244 (0.0412)

The numbers in the brackets are p-values.

We can summarize from Table 2: (1) Both dy and log(dy) can predict stock returns (ewny or vwny); (2) After controlling the first component, the addition of the other two components make little contribution to predict stock return; (3) The forecasting power of dividend yield is greater for vwny than ewny. This may be due to the construction of dy is based on value-weighted stock.

As a robustness check, we split the full sample into two subperiods, the first period is return series from 1946.1 to 1994.12 (thus DY is from 1945.12 to 1994.11) with sample size 589; the second period is from 1995.1 to 2007.12, with sample size 155. As noticed by Stambaugh (1999), Lewellen (2004) and Campbell and Yogo (2006), the dividend yield reaches historical low in 1995, and including the samples after that may make different conclusion about predictability of stock return.

The results for the two different periods are collected in the following two tables.

Table 3: Testing Predictability: 1946-1994

Truncation parameter	dy		log(dy)	
	ewny	vwny	ewny	vwny
p=1	8.0681 (0.0045)	10.2474 (0.0014)	7.5954 (0.0059)	9.8162 (0.0017)
p=2	8.7192 (0.0128)	10.6354 (0.0049)	8.3365 (0.0155)	10.2350 (0.0060)
p=3	9.3710 (0.0247)	11.8639 (0.0079)	9.1206 (0.0277)	11.5312 (0.0092)

Table 4: Testing Predictability: 1995-2007

Truncation parameter	dy		log(dy)	
	ewny	vwny	ewny	vwny
p=1	3.8552 (0.0496)	9.7494 (0.0018)	3.1422 (0.0763)	8.1898 (0.0042)
p=2	4.9233 (0.0853)	11.7220 (0.0028)	4.6937 (0.0957)	11.0027 (0.0041)
p=3	4.9968 (0.1720)	12.4531 (0.0060)	5.0874 (0.1655)	12.1073 (0.0070)

From Tables 3 and 4, we can find that: (1) for the first period, the dividend yield can predict stock return as has been documented in previous studies regardless of whether we use equal-weighted, value-weighted stock return, log-dividend yield or raw dividend yield; (2) in the second period, the predictability of equal-weighted stock return is weaker than that for the first period. However, we still find strong evidence of predictability of value-weighted stock return. Considering that the dividend yield dataset is constructed using value-weighted stock, we can still conclude the return predictability still exists for the second period; (3) from the information provided by the p -values, it seems that a simple one component model can well approximate the unknown nonlinear function.

3.3 Out-of-Sample Evaluation Strategy

We estimate the predictive regression relationship using the following three models:

(Model 1) Historic mean model:

$$r_{t+1} = \alpha + \varepsilon_{t+1}. \quad (3.1)$$

(Model 2) Linear model:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}. \quad (3.2)$$

(Model 3) Integrable model:

$$r_{t+1} = \alpha + \sum_{i=0}^{p-1} \theta_i F_i(x_t) + \varepsilon_{t+1}, \quad (3.3)$$

where $p = \lfloor cT^{5/44} \rfloor$ with c being chosen to be 1, 2, 3, or p chosen by the criterion discussed in Section 2 (denoted p^* in Tables 2–13). Model 1 indicates there is no predicting power for x , whereas Model 2 is the linear predicting model. Model 3 is the model proposed in this paper. Let x denote log dividend yield or log book-to-market ratio or log earning-price ratio. The performance of the estimators are evaluated based on the out-of-sample criterion, because the out-of-sample performance provides more convincing evidence as discussed in the introduction.

In the literature of economic forecasting, there are basically two methods widely used. The first estimation method uses recursive (or expanding) window. The initial in-sample period is the first $\lfloor Tr \rfloor$ observations. Then the estimator in each model is used to estimate the following period return using the following period predictors. The performance of different models is evaluated by its prediction accuracy through comparing the estimators with the true returns. The next in-sample period is the first $\lfloor Tr \rfloor + 1$ observations. And then a period ahead forecasting is made. The second estimation method uses a rolling window with the in-sample size always being $\lfloor Tr \rfloor$, i.e., in each recursion, one observation is abandoned and then one new observation is included.

Rolling windows are typically justified by appealing to structure breaks. However, as demonstrated by Pesaran and Timmermann (2007), by the criterion of minimum square forecasting error, the optimal window size is a complicated function of timing and size of breaks. It is not easy to account for the effect of breaks and to choose the window size. The recursive window utilizes all the data when one forecasts next period return, and thus it may increase the precision of in-sample estimation, which in turn, leads to better out-of-sample forecasting. Considering the slow convergence rate of Hermite series estimation, this point is especially important. Although not reported here, the performance of the recursive window method is better than the rolling window for all the choice of r in our experiment for each model. Thus, we focus on the results of the recursive window method in this paper.

The performance of the out-of-sample performance is evaluated by a root mean squared forecasting error (RMSFE) of the form:

$$RMSFE = \sqrt{\frac{1}{T - [Tr]} \sum_{s=[Tr]+1}^T (r_s - \hat{r}_s)^2}, \quad (3.4)$$

where r_s is the true return at period s and \hat{r}_s is the estimated return.

3.4 Predict Returns with Dividend Yield

Among the instruments thought to capture some of the equity premium's variation in the sense of predicting stock returns, the lagged dividend price ratio or dividend yield have emerged as a favourite. Fama and French (1988) find that the dividend-price ratio and book-to-market ratio are useful in predicting stock returns. Similar results are reported in Campbell and Shiller (1988). Campbell (1991) and Cochrane (1992) attribute large fraction of the variance of dividend price ratio to the variation in expected returns. However, as noticed by Mankiw and Shapiro (1986), Nelson and Kim (1993) and Stambaugh (1999), persistence of explaining variables and correlation of explaining variables with error terms lead to over-rejection of the null hypothesis of no predictability in finite sample, thus can be seriously biased towards finding predictability. Stambaugh (1999) finds that after controlling the finite sample bias, there is little evidence of stock return predictability, and the one-sided p-value is 0.15 when the stock return is regressed on dividend yield over the period 1952-1996. Contrary to Stambaugh (1999), Lewellen (2004) finds that the evidence of predictability can be strong if we require the explaining variables to be stationary (i.e., the first order autoregressive parameter is less than 1). Using a robust test based on local to unit root framework, Campbell and Yogo (2006) find that the dividend-price ratio predicts returns at an annual frequency, but they can not reject the null hypothesis of no predictability at quarterly and monthly frequencies.

The out-of-sample evidence of Welch and Goyal (2008) suggests that dividend yield can not outperform the historic mean model. However, Campbell and Thompson (2008) show that by imposing the restriction of the sign of regression coefficient, the dividend yield can predict stock returns. Also, Lettau and Nieuwerburgh (2008) suggest the dividend yield can predict stock returns by controlling the structural break in dividend yield. These papers lead us to conclude that the dividend yield should predict stock returns in a nonlinear fashion.

We apply the evaluation strategy proposed in Section 3.2 to study the predicting power of log dividend yield to stock return (nominal and real returns), and the results are summarised in following four tables.

Table 5: Forecasting VWNYS with Dividend Yield: RMSFE $\times 100$

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.0492	4.0583	4.0534	4.0810	4.0908	4.0511
	1/6	4.0726	4.0689	4.0607	4.0899	4.1052	4.0583
1946.1	1/4	4.1549	4.1470	4.1394	4.1678	4.1724	4.1368
To	1/3	4.2260	4.2212	4.2126	4.2429	4.2462	4.2098
2007.12	1/2	4.0519	4.0623	4.0481	4.0865	4.0841	4.0441
	5/8	3.9586	3.9894	3.9780	4.0178	4.0182	3.9728
	2/3	3.8941	3.9298	3.9100	3.9613	3.9597	3.9036
	[Tr]=240	4.2253	4.2164	4.2084	4.2384	4.2411	4.2055
	[Tr]=360	4.2392	4.2414	4.2276	4.2709	4.2723	4.0092

Table 6: Forecasting EWNYS with Dividend Yield: RMSFE $\times 100$

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.7287	4.7400	4.7352	4.7616	4.7808	4.7323
	1/6	4.7844	4.7834	4.7751	4.8034	4.8303	4.7720
1946.1	1/4	4.9069	4.9001	4.8922	4.9198	4.9356	4.8888
To	1/3	4.9916	4.9845	4.9764	5.0045	5.0225	4.9727
2007.12	1/2	4.4265	4.4325	4.4145	4.4574	4.4720	4.4089
	5/8	4.1610	4.2000	4.1767	4.2182	4.2464	4.1691
	2/3	4.1426	4.1883	4.1588	4.2109	4.2348	4.1481
	[Tr]	4.9932	4.9837	4.9755	5.0044	5.0203	4.9718
	[Tr]	4.4057	4.4109	4.3946	4.4341	4.4496	4.3892

Table 7: Forecasting EVWNY with Dividend Yield: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.0698	4.0805	4.0715	4.1000	4.1071	4.0704
	1/6	4.0937	4.0925	4.0800	4.1102	4.1221	4.0788
1946.1	1/4	4.1754	4.1714	4.1590	4.1882	4.1896	4.1577
To	1/3	4.2456	4.2473	4.2331	4.2648	4.2644	4.2317
2007.12	1/2	4.0566	4.0834	4.0612	4.1013	4.0936	4.0592
	5/8	3.9472	3.9833	3.9683	4.0123	4.0112	3.9656
	2/3	3.8872	3.9290	3.9034	3.9597	3.9574	3.9020
	[Tr]=240	4.2471	4.2438	4.2304	4.2615	4.2606	4.2290
	[Tr]=360	4.0223	4.0475	4.0258	4.0655	4.0590	4.0239

Table 8: Forecasting EEWNY with Dividend Yield: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.7481	4.7608	4.7518	4.7814	4.7973	4.7505
	1/6	4.8041	4.8054	4.7925	4.8242	4.8475	4.7912
1946.1	1/4	4.9262	4.9225	4.9098	4.9407	4.9528	4.9083
To	1/3	5.0112	5.0090	4.9955	5.0272	5.0411	4.9939
2007.12	1/2	4.4367	4.4572	4.4302	4.4790	4.4874	4.4277
	5/8	4.1650	4.2090	4.1800	4.2293	4.2542	4.1771
	2/3	4.1511	4.2025	4.1654	4.2259	4.2471	4.1616
	[Tr]=240	5.0148	5.0094	4.9960	5.0283	5.0401	4.9944
	[Tr]=360	4.4152	4.4344	4.4093	4.4548	4.4643	4.4069

From Table 2, we can see that for the value-weighted return, if the initial estimation window is taken to be 20 years (e.g., Welch and Goyal (2008); Chen and Hong (2010); Rapach and Zhou (2012)), the integrable model outperforms both the historic mean model

and the linear model, especially when the data driven truncation parameter is used. If the initial estimation window is taken to be 30 years (e.g., Lattau and Nieuwerburgh (2008)), the integrable model also outperforms both the historic mean model and the linear model. And the result seems to be robust to a wide range of selections of the initial windows. The result supports the conclusion that dividend yield has predictive power by Lewellen (2004) using in-sample test.

From Tables 3–5, we can see that using different measures of return may lead to somehow different results. However, the conclusion that nonlinear predictive ability of dividend yield seems to be robust to different measures of return.

To be more specific, when $c = 1$, the full-sample choice of p is equal to $\lfloor 744^{5/44} \rfloor = 2$. The estimated regression is:

$$\begin{aligned}\hat{r}_{t+1} &= 0.0065 + 2.2676F_0(x_t) + 0.3379F_1(x_t) \\ &= 0.0065 + (1.7032 + 0.4498 x_t) \cdot \exp(-0.5 x_t^2),\end{aligned}\tag{3.5}$$

where x_t is the log dividend yield, $F_0(x) = \frac{1}{\sqrt{\sqrt{\pi}}} \exp(-\frac{1}{2}x^2)$, $F_1(x) = \frac{\sqrt{2}}{\sqrt{\sqrt{\pi}}} x \cdot \exp(-\frac{1}{2}x^2)$, and r_{t+1} is the value-weighted stock return⁴. Model (3.5) shows that there is no support for neither the historic mean nor the simple linear model.

3.5 Predict Return with Book-to-Market Ratio or Earning-Price Ratio

Book-to-market ratio and earning-price ratio are two other commonly used stock return predictors. As shown previously and elsewhere in the literature, the two ratios are also very persistent. Using the annual data over the period of 1926-1991, Kothari and Shanken (1997) show that book-to-market ratio is helpful in predicting stock returns. Pontiff and Schall (1998) show that the book-to-market ratio of the Dow Jones Industrial Average predicts market returns and small firm excess returns over the period of 1926-1994. Both papers also find that the predictability is not robust to different sub-samples. Using quarterly data over the period of the first quarter of 1947 to the last quarter of 1994, Lamont (1998) shows that earning price ratio can predict stock return. Generally speaking, the evidence of return predictability of these two ratios are weaker than dividend yield. Lewellen (2004) shows that these two ratios have limited predictive power.

The out-of-sample evidence by Welch and Goyal (2008) suggests that these two variables are not helpful in predicting returns. However, Campbell and Thompson (2008) show that

⁴Of course, we can also write the estimated regression function for the case of equal-weighted return or excess returns. To save space, we just report the results for the value-weighted return case.

earning-price ratio can forecast excess return by imposing sign constraints on the regression coefficient using monthly and annual data, while book-to-market can forecast excess return using the annual data.

In this subsection, we will compare out-of-sample performance of the three models with earning-price ratio or book-to-market ratio as predictor. The results are summarized in the Tables 6-13.

Table 9: Forecasting VWNYS with Book-to-Market Ratio: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.0492	4.0641	4.0819	4.0869	4.1240	4.0899
	1/6	4.0726	4.0832	4.1045	4.1069	4.1428	4.1114
1946.1	1/4	4.1549	4.1600	4.1828	4.1835	4.2141	4.1881
To	1/3	4.2260	4.2340	4.2566	4.2518	4.2750	4.2604
2007.12	1/2	4.0519	4.0731	4.1006	4.1037	4.1420	4.1037
	5/8	3.9586	4.0061	4.0365	4.0671	4.2542	4.0671
	2/3	3.8941	3.9398	3.9581	4.0052	4.0423	4.0052
	$\lfloor Tr \rfloor=240$	4.2253	4.2257	4.2491	4.2471	4.2710	4.2559
	$\lfloor Tr \rfloor=360$	4.0174	4.0376	4.0649	4.0677	4.1044	4.0677

Table 10: Forecasting EWN_Y with Book-to-Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.7287	4.7431	4.7646	4.7596	4.8069	4.7538
	1/6	4.7844	4.7960	4.8210	4.8123	4.8588	4.8072
1946.1	1/4	4.9069	4.9103	4.9373	4.9256	4.9650	4.9208
To	1/3	4.9916	4.9906	5.0191	5.0012	5.0233	5.0018
2007.12	1/2	4.4265	4.4337	4.4724	4.4474	4.4825	4.4474
	5/8	4.1610	4.2007	4.2453	4.2424	4.2754	4.2424
	2/3	4.1426	4.1825	4.2162	4.2197	4.2551	4.2197
	[Tr]=240	4.9932	4.9881	5.0187	5.0057	5.0280	5.0062
	[Tr]=360	4.4057	4.4119	4.4495	4.4289	4.4609	4.4289

Table 11: Forecasting EVW_Y with Book-to-Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.0698	4.0848	4.1029	4.1019	4.1388	4.1090
	1/6	4.0937	4.1049	4.1263	4.1223	4.1579	4.1306
1946.1	1/4	4.1754	4.1819	4.2047	4.1985	4.2289	4.2071
To	1/3	4.2456	4.2582	4.2805	4.2685	4.2912	4.2811
2007.12	1/2	4.0566	4.0904	4.1161	4.1106	4.1487	4.1106
	5/8	3.9472	3.9996	4.0317	4.0541	4.0936	4.0541
	2/3	3.8872	3.9388	3.9590	3.9965	4.0334	3.9965
	[Tr]=240	4.2471	4.2502	4.2732	4.2643	4.2878	4.2770
	[Tr]=360	4.0223	4.0549	4.0806	4.0741	4.1110	4.0741

Table 12: Forecasting EEWNY with Book-to-Market Ratio: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.7481	4.7626	4.7845	4.7732	4.8208	4.7670
	1/6	4.8041	4.8162	4.8416	4.8262	4.8728	4.8208
1946.1	1/4	4.9262	4.9308	4.9580	4.9392	4.9789	4.9336
To	1/3	5.0112	5.0133	5.0413	5.0161	5.0382	5.0141
2007.12	1/2	4.4367	4.4546	4.4914	4.4575	4.4923	4.4575
	5/8	4.1650	4.2096	4.2563	4.2422	4.2757	4.2422
	2/3	4.1511	4.1969	4.2338	4.2245	4.2608	4.2245
	[Tr]=240	5.0148	5.0112	5.0413	5.0211	5.0435	5.0192
	[Tr]=360	4.4152	4.4317	4.4676	4.4375	4.4698	4.4375

Table 13: Forecasting VWNY with Earning-Price Ratio: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.2694	4.2995	4.3543	4.8438	8.6921	4.6711
	1/6	4.2179	4.2416	4.2953	4.6907	8.5831	4.6289
1963.5	1/4	4.2512	4.2765	4.3366	4.9392	4.9789	4.7028
To	1/3	4.0969	4.0838	4.0919	4.1446	4.2798	4.1446
2007.12	1/2	3.9929	3.9975	4.0213	4.0222	4.0501	4.0222
	5/8	3.5305	3.5496	3.5677	3.5727	3.6014	3.5727
	2/3	3.5727	3.5828	3.6073	3.5928	3.6423	3.5928
	[Tr]=240	3.9407	3.9418	3.9619	3.9734	4.0008	3.9734
	[Tr]=360	3.5912	3.6005	3.6258	3.6091	3.6563	3.6091

Table 14: Forecasting EWN_Y with Earning-Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	5.0249	5.0571	5.0842	5.4223	9.6600	5.2642
	1/6	4.9193	4.9434	4.9656	5.1813	9.5667	5.1586
1963.5	1/4	4.8053	4.8283	4.8547	5.0752	5.0157	5.0736
To	1/3	4.4650	4.4554	4.4409	4.5015	4.5669	4.5047
2007.12	1/2	4.1776	4.1958	4.1870	4.2122	4.2424	4.2119
	5/8	3.6096	3.6661	3.6356	3.6683	3.6906	3.6571
	2/3	3.6633	3.6993	3.6776	3.6975	3.7365	3.7015
	[Tr]=240	4.1620	4.1775	4.1642	4.1899	4.2296	4.1856
	[Tr]=360	3.6842	3.7162	3.6965	3.7152	3.7513	3.7206

Table 15: Forecasting EVW_Y with Earning-Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	4.2854	4.3289	4.3847	4.8713	8.6528	4.7002
	1/6	4.2309	4.2691	4.3236	4.7164	8.5361	4.6558
1963.5	1/4	4.2583	4.3017	4.3638	4.7242	4.4774	4.6296
To	1/3	4.1021	4.1061	4.1139	4.1630	4.3074	4.2726
2007.12	1/2	3.9898	4.0015	4.0206	4.0248	4.0580	4.0377
	5/8	3.5298	3.5551	3.5716	3.5756	3.6114	3.5866
	2/3	3.5696	3.5875	3.6087	3.5941	3.6512	3.6204
	[Tr]=240	3.9377	3.9466	3.9621	3.9754	4.0074	3.9885
	[Tr]=360	3.5880	3.6050	3.6268	3.6101	3.6654	3.6371

Table 16: Forecasting EEWNY with Earning-Price Ratio: RMSFE \times 100

Sample	r	Model 1	Model 2	Model 3			
				c=1	c=2	c=3	p*
	1/8	5.0460	5.0939	5.1227	5.4579	9.6246	5.3704
	1/6	4.9368	4.9783	5.0021	5.2156	9.5253	5.2676
1963.5	1/4	4.8187	4.8628	4.8923	5.1052	5.0483	5.1929
To	1/3	4.4780	4.4879	4.4749	4.5321	4.6044	4.5664
2007.12	1/2	4.1817	4.2061	4.1952	4.2229	4.2572	4.2377
	5/8	3.6172	3.6767	3.6476	3.6795	3.7072	3.6740
	2/3	3.6687	3.7097	3.6868	3.7081	3.7533	3.7161
	[Tr]=240	4.1676	4.1895	4.1745	4.2010	4.2453	4.2102
	[Tr]=360	3.6891	3.7267	3.7050	3.7258	3.7685	3.7348

From Tables 6–13, we can see that the predictive power of these two ratios are much weaker than that of dividend yield, regardless of whether linear or nonlinear models are used. The forecasting ability of these two ratios are restricted to only some special cases (with colour different from black) for each return series. Thus, the predictability of these two series, if exists, should be very weak. The conclusion is the same as that by Lewellen (2004) where an in-sample test is used.

4 Conclusions and Discussion

In this paper, we introduce a new and simple nonlinear return predicting model and propose testing and estimating the model using Hermite functions. The out-of-sample evaluation results suggest that while dividend yield can predict the stock return nonlinearly, the book-to-market ratio or earning-price ratio can not. As Welch and Goyal (2008) conclude that “the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power.”, our paper suggests that finding *the way* that the variables predicting the return should be of similar importance.

The methodology of this paper can be applied to data sharing similar characteristics to return and dividend yield. For example, we may study the relationship of return of exchange rate and forward rate premium, where the exchange rate return is far less persistent than

the forward rate premium. The current method can be also extended to study predictability of multivariate variables where the following additive model is used:

$$r_t = \alpha + \sum_{i=1}^q g_i(x_{it-1}) + \varepsilon_t,$$

where x_{it} , $i = 1, \dots, q$ are predicting variables. These issues are left for future research.

5 Acknowledgements

Earlier versions of this paper were presented at several workshops and seminars. Comments and suggestions by the workshop and/or seminar participants were much appreciated. The first author acknowledges financial support by the Chinese National Science Foundation under Grant Number: NSFC71703045. Thanks also go to the Australian Research Council Discovery Grants Program for its financial support under Grant Numbers: DP150101012 & DP170104421.

Appendix A

To derive an asymptotic theory for our test, we need to introduce the following technical assumptions.

A.1 Let $x_t = w_t + w_{t-1} + \dots + x_0$, where $x_0 = O_P(1)$ and w_t is generated by $w_t = C(L)\eta_t$, in which L is the lag operator, $C(L) = \sum_{k=0}^{\infty} c_k L^k$, $C(1) \neq 0$ and $\sum_{k=0}^{\infty} k|c_k| < \infty$, $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with an absolutely continuous distribution with respect to the Lebesgue measure, the characteristic function $\phi(u)$ satisfies $\int_{-\infty}^{\infty} \phi(u)^2 du < \infty$, $E[\eta_1] = 0$, $E[\eta_1^2] = \sigma_\eta^2$ and $E[|\eta_1|^q] < \infty$ for some $q > 4$.

A.2 Let f be integrable and squared integrable.

A.3 (i) Suppose that $\{\varepsilon_t, \mathbb{F}_t\}$ is a stationary ergodic martingale difference sequence with respect to \mathbb{F}_t , i.e. $E(\varepsilon_t | \mathbb{F}_{t-1}) = 0$ a.s., where $\mathbb{F}_t \equiv \{x_{t+1}, x_t, \dots, x_1, \varepsilon_t, \dots, \varepsilon_1\}$.

(ii) Let $E(\varepsilon_t^2 | \mathbb{F}_{t-1}) \rightarrow_{a.s.} \sigma^2$ and $\sup_t E[|\varepsilon_t|^m | \mathbb{F}_{t-1}] < \infty$ with $m > 4$.

Remark: It's easy to see that the Hermite series are integrable, and from Park and Phillips (2001), the integrable property is closed under the conventional operators, such as addition, subtraction and production. While we impose the martingale difference assumption for the error term, which is not contrary to the conventional assumption that the stock return and the financial ratios are contemptuously correlated (notice here x_t is lag variable) and the predictor is predetermined, this condition can be relaxed to allow that x_t is to be contemptuously correlated with ε_t using the theory developed in Chang and Park (2011) without any change of the conclusions.

Lemma 1: Under A.1–A.3, $\frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_t) - \bar{F})^\tau \rightarrow_D \int_{-\infty}^{\infty} F(x)F(x)^\tau dx L(1,0)$, where $L(1,0)$ is the local-time variable associated with the standard Brownian motion.

Proof: Since p is assumed to be fixed, we have $\frac{1}{\sqrt{T}} \sum_{t=1}^T F(x_{t-1})F(x_{t-1})^\tau \rightarrow_D \int_{-\infty}^{\infty} F(x)F(x)^\tau dx \cdot L(1,0)$ due to Theorem 3.2 in Park and Phillips (2001) and use the Cramer-Wold device. $\frac{1}{\sqrt{T}} \sum_{t=1}^T F(x_{t-1})\bar{F}^\tau = o_P(1)$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{F}\bar{F}^\tau = o_P(1)$. Because of the orthogonality of Hermite series, $\int_{-\infty}^{\infty} F_i(x)F_j(x)dx = 0$, if $i \neq j$, the matrix is diagonal.

The proof then follows.

Lemma 2 Let A.1–A.3 hold. If, in addition, $F(x)f(x)$ is integrable with $\int_{-\infty}^{\infty} F(x)f(x) \neq 0$, we have $\frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})f(x_{t-1}) \rightarrow_D \int_{-\infty}^{\infty} F(x)f(x)dx L(1,0)$.

Proof: $\frac{1}{\sqrt{T}} \sum_{t=1}^T F(x_{t-1})f(x_{t-1}) \rightarrow_d \int_{-\infty}^{\infty} F(x)f(x)dx L(1,0)$ due to Theorem 3.2 in Park and Phillips (2001) and the Cramer-Wold device. And $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{F}f(x_{t-1}) = o_P(1)$, so the result follows by the Slutsky theorem.

Lemma 3 Under A.1–A.3, $\frac{1}{\sqrt{T}} \sum_{t=1}^T F(x_{t-1})\varepsilon_t \rightarrow_D \sigma(L(1,0) \int_{-\infty}^{\infty} F(x)F'(x)dx)^{1/2} N_1$, where N_1 is a standard normal variable independent of B_1 , thus independent of $L(1,0)$.

Proof: The proof follows from Theorem 3.2 of Park and Phillips (2001) by using the Cramer-Wold device.

Proof of Theorem 1. From Lemmas 1 and 3, we have

$$\begin{aligned}
T^{\frac{1}{4}}\hat{\theta} &= T^{\frac{1}{4}} \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})r_t \\
&= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})r_t \\
&= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})(\alpha + \varepsilon_t) \\
&\rightarrow_D \left(\int_{-\infty}^{\infty} F(x)F(x)^\tau dx L(1,0) \right)^{-1} \times \sigma(L(1,0) \int_{-\infty}^{\infty} F(x)F'(x)dx)^{1/2} N_1 \\
&= \left(\int_{-\infty}^{\infty} F(x)F(x)^\tau dx L(1,0) \right)^{-1/2} \times \sigma N_1,
\end{aligned}$$

which is a mixed normal distribution. In order to prove this theorem, it suffices for us to show

$$\hat{\sigma}^2 \rightarrow_P \sigma^2$$

and then apply the Slutsky theorem.

We have

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 = \frac{1}{T} \sum_{t=1}^T [r_t - \hat{\alpha} - \hat{\theta}^\tau F(x_{t-1})]^2 \\
&= \frac{1}{T} \sum_{t=1}^T [r_t - \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\theta}^\tau F(x_{t-1})) - \hat{\theta}^\tau F(x_{t-1})]^2 \\
&= \frac{1}{T} \sum_{t=1}^T [\alpha + \varepsilon_t - \alpha - \frac{1}{T} \sum_{t=1}^T \varepsilon_t - \frac{1}{T} \hat{\theta}^\tau \sum_{t=1}^T F(x_{t-1}) - \hat{\theta}^\tau F(x_{t-1})]^2 \\
&= \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 + O_P(1) \rightarrow_P \sigma^2.
\end{aligned}$$

Thus, the limiting distribution of the proposed Wald test statistic is $\chi^2(p)$ where p is a fixed truncation parameter of the Hermite series.

Proof of Theorem 2

Observe that

$$\begin{aligned}
\hat{\theta} &= \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})' \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F}) r_t \\
&= \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F})(\alpha + f(x_t) + \varepsilon_t) \\
&= \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F}) f(x_{t-1}) \\
&+ \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F}) \varepsilon_t
\end{aligned}$$

Because

$$\begin{aligned}
&\left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right)^{-1} \sum_{t=1}^T (F(x_{t-1}) - \bar{F}) f(x_{t-1}) \rightarrow_d \\
&\left(\int_{-\infty}^{\infty} F(x) F(x)^\tau dx L(1, 0) \right)^{-1} \int_{-\infty}^{\infty} F(x) f(x) dx L(1, 0) \equiv \theta^*,
\end{aligned}$$

where θ^* can be regarded as a *pseudo true value* under the misspecified model, see, e.g. Kasparis (2011). We conclude that $\hat{\theta}^\tau \left(\sum_{t=1}^T (F(x_{t-1}) - \bar{F})(F(x_{t-1}) - \bar{F})^\tau \right) \hat{\theta}$ is of stochastic order $O_P(\sqrt{T})$.

For the denominator, we have

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 \\
&= \frac{1}{T} \sum_{t=1}^T \left(r_t - \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\theta}^\tau F(x_{t-1})) - \hat{\theta}^\tau F(x_{t-1}) \right)^2
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{t=1}^T \left(\alpha + f(x_{t-1}) + \varepsilon_t - \frac{1}{T} \sum_{t=1}^T (\alpha + f(x_{t-1}) + \varepsilon_t - \hat{\theta}^T F(x_{t-1})) - \hat{\theta}^T F(x_{t-1}) \right)^2 \\
&= \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 + o_P(1) \rightarrow_P \sigma^2.
\end{aligned}$$

Thus, Q diverges to infinity with rate $O_P(T^{1/2})$.

References

- Andrews, D.W.K., 1991. Asymptotic normality of series estimator for nonparametric and semiparametric models. *Econometrica*, 59, 307-345.
- Cai, B., 2012. Integrable Cointegration: Theory and Applications. Ph.D. Thesis, Xiamen University.
- Cai, B. and J., Gao, 2013. Hermite Series Estimation in Nonlinear Cointegration Models. Monash University working paper at <https://ideas.repec.org/p/msh/ebswps/2013-17.html>.
- Campbell, J.Y., 2000. Asset pricing at the millennium. *Journal of Finance*, 55, 1515-1567.
- Campbell, J.Y., 2008. Viewpoint: estimating the equity premium. *Canadian Journal of Economics*, 41, 1-21.
- Campbell, J.Y. and R.J., Shiller, 1988. The dividend price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1, 195-228.
- Campbell, J.Y. and S.B., Thompson, 2008. Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies*, 21, 1509-1531.
- Campbell, J.Y., and M., Yogo, 2006. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81, 27-60.
- Chen, Q. and Y., Hong, 2010. Predictability of equity returns over different time horizons: a nonparametric approach. Working paper at <http://old.ccer.edu.cn/download/11861-1.pdf>
- Cochrane, J.H., 1992. Explaining the variance of price-dividend ratios. *Review of Financial Studies*, 5, 243-280.
- Engstrom, E., 2003. The conditional relationship between equity risk and dividend price ratio. Working paper at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355340.
- Fama, E.F. and K.R., French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22, 3-25.
- Gao, J., H., Tong and R., Ralff, 2002. Adaptive orthogonal series estimation in additive stochastic regression models. *Statistica Sinica*, 12, 409-428.
- Granger, C.W.J., 1995. Modelling nonlinear relationships between extended-memory variables. *Econometrica*, 63, 265-279.
- Gonzalo, J. and J.Y., Pitarakis, 2012. Regime-specific predictability in predictive regression. *Journal of Business and Economic Statistics*, 20, 229-241.
- Goyal, A. and I., Welch, 2003. Predicting the equity premium with dividend ratios. *Management Science*, 49, 639-654.
- Hansen, P.R. and A., Timmermann, 2012. Choice of sample split in out-of-sample forecast evaluation. Manuscript, European University Institute and University of California at San Diego.
- Karlsen, H.A. and D., Tjøstheim, 2001. Nonparametric estimation in null recurrent time series. *Annals of Statistics*, 29, 372-416.

- Kasparis, I.E., Andreou H.A. and P.C.B., Phillips, 2015. Nonparametric predictive regression. *Journal of Econometrics*, 185, 468-494.
- Kothari, S.P. and J., Shanken, 1997. Book-to-market, dividend yield, and expected market returns: a time series analysis. *Journal of Financial Economics*, 44, 169-203.
- Lamont, O., 1998. Earnings and expected returns. *Journal of Finance*, 5, 1563-1587.
- Lettau, M. and S.V., Niewerburgh, 2008. Reconciling the return predictability evidence. *Review of Financial Studies*, 21, 1607-1652.
- Lee, T., Y., Tu and A., Ullah, 2015. Forecasting equity premium: global historical average versus local historical average and constraint. *Journal of Business and Economic Statistics*, 33, 393-402.
- Lewellen, L., 2004. Predicting returns with financial ratios. *Journal of Financial Economics*, 74, 209-235.
- Mankiw, N.G. and M., Shapiro, 1986. Do we reject too often ? small sample properties of tests of rational expectation models. *Economic Letters*, 20, 139-145.
- Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79, 147-168.
- Nelson, C. and M., Kim, 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance*, 48, 641-661.
- Pesaran, M.H. and A., Timmermann, 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137, 134-161.
- Phillips, P.C.B., 2015. Halbert White Jr. Memorial JFEC Lecture: pitfalls and possibilities in predictive regression. *Journal of Financial Economics*, 49, 141-160.
- Pontiff, J. and L.D., Schall, 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics*, 49, 141-160.
- Rapach, D.E. and G.F., Zhou, 2012. Forecasting stock returns. In *Handbook of Economic Forecasting*, Vol 2, edited by Elliott, G. and A. Timmermann.
- Stambaugh, R., 1999. Predictive regression. *Journal of Financial Economics*, 54, 375-421.
- Teräsvirta, T., D., Tjøstheim, and C.W.J., Granger, 2010. *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.
- Welch, I. and A., Goyal, 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21, 1455-1508.