

PREDICTION OF STOCK RETURNS: A NEW WAY TO LOOK AT IT

BY

JENS PERCH NIELSEN¹ AND STEFAN SPERLICH²

ABSTRACT

While the traditional R^2 value is useful to evaluate the quality of a fit, it does not work when it comes to evaluating the predictive power of estimated financial models in finite samples. In this paper we introduce a validated R_V^2 value useful for prediction. Based on data from the Danish stock market, using this measure we find that the dividend-price ratio has predictive power. The best horizon for prediction seems to be four years. On a one year horizon, we find that while inflation and interest rate do not add to the predictive power of the dividend-price ratio then last years excess stock return does.

KEYWORDS

Prediction, Stock returns, Dividend price ratio, Cross Validation.

1. INTRODUCTION

Long term investors have the contradicting aims of minimizing risk and maximizing return over the long run. Much financial literature investigates trading patterns and strategy among long term investors, for example, Barber and Terrance (2000) argue for a buy-and-hold type of strategy that does not eat up returns by trading costs and many professional advisers argue that stocks are better over the long run, see Siegel (1998) and Jagannathan and Kocherlakota (1996) for particular easily read accounts on this. Other professional financial advisers say that expected returns in financial markets vary over time and contain a significant predictable component. Consequently time periods exist where the long term investor might choose to sell stocks and buy

* The research was supported by the Danish Social Science Research fund and the Spanish DGES BEC2001-1270. Comments and suggestions from Thomas Engsted, Carsten Tangg ard and two anonymous referees are gratefully acknowledged.

¹ Codanhus, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark. Email: npj@codan.dk.

² Universidad Carlos III de Madrid, Email: stefan@est-econ.uc3m.es.

bonds, because the return on stocks in these time periods do not match the risk involved. The dividend-price ratio and the earning-price ratio, in particular, has proven to have some predictive power for future stock returns, see Campbell, Lo, and MacKinlay (1997, Chapter 7) for an up-to-date account regarding of the dividend yield based predictability of stock returns, and see Shiller (2000, p.8) for a warning of an overvalued American stock exchange based on the earning-price ratio. Campbell, Lo, and MacKinlay (1997) argued that the predictable component of stock returns is increasing with the time horizon, since the measure of fit, the R^2 , increases rapidly with the time horizon. In the actuarial literature Wilkie (1993) gets to a similar conclusion replicating the linear modeling approach of Fama and French (1988). The following quote is from Wilkie (1993, p.341), who found that predictive power seems to be strongest for a six and a half year time horizon for British data: "A 1% difference in the dividend yield at the time of purchase of the stock makes a difference... equivalent to about 4.1% a year compound for about six and a half years". So, there is a considerable financial and actuarial literature on the predictive power of the dividend yield, see also Richardson and Stock (1989) and Wilkie (1995). However, most literature uses traditional in-sample methodology like goodness of fit, the traditional R^2 value or parametric estimation procedures combined perhaps with some testing. In this paper we consider an adjusted measure of predictive power, the R_V^2 value, that is an out-of-sample measure in the sense that it measures how the model actually predicts into the future, see Eun and Resnick (1988,1994) for financial papers using a similar type of out-of-sample approach to evaluate their times series of stocks, bonds and exchange rates as we use in this paper to analyze our financial time series.

The paper follows its historical development. First we go through an analysis of Danish data from 1922 to 1996 and then we add the analysis of the updated data set from 1922 to 2001.

Regarding the data set 1922-1996

Based on our out-of-sample measure, it seems that the models have strongest predictive power for a time horizon of four years for Danish stock returns, at least with respect to our criterion. Dividend-price does seem to have predictive power whereas knowledge of inflation and short interest rates do not seem to add to this predictive power. However, our study shows that the one year lagged returns do. The best predictive filter on a one year basis turns out to be a two-dimensional fully nonparametric estimator based on the dividend-price ratio and last years lagged excess return. Last years excess return enters with a tendency towards reversal, such that good years tend to follow bad years and vice versa. The dividend-price ratio is, however, still the most indicative parameter while estimating the excess returns of the coming years.

Moreover, based on the current level on the dividend-price ratio in Denmark, around 1%, we concluded (in december 2000) that expected excess returns

on stocks are indeed below zero, for all the considered time horizons with good prediction power, namely one, two, three, four and five years time horizons. Based on this finding we then concluded further that the current market and political situation in Denmark was out of balance, since all institutional investors heavily increased their percentages of stocks in their portfolios right then. On average, an increase from around 20% invested in stocks to around 40% invested in stocks have been seen for long term institutional investors in Denmark over the last seven years. The model of this paper argue that this strategy increases the risk without increasing the average return.³ We believe that the considerations of this paper can be helpful while developing a modern information system for the long term investor.

Regarding the updated data set 1922-2001

It turns out that while there still is predictive power in the updated data set, it is much lower than in the original data set. This remarkable finding can have two explanations. Either the Danish main index, the KFX index, has had an exceptional behavior in some of the last five years or the entire world market has followed exceptional rules during this period. In either case, our findings show that some care has to be taken regarding predictive power of dividend yields and that further studies based on international data sets should be added to the current to get a fuller understanding of the problem. It could for example be relevant to consider a regime shift model such as the one of Harris (1999) to understand this question further. However, while a regime shift model does add to the understanding of historical facts, it does not help much when it comes to predictive power. For a deep insight into the nature of uncertainty in prediction, see Cairns (2000, p 314). In this paper we only consider the first of the three steps considered by Cairns, namely Method 1 that finds the best fit to a model according to a certain criterion. However, since our criterion is a validated measure of error, we implicitly take care of the errors dealt with in Cairns Method 2 and Method 3, that consider uncertainty due to parameter estimation and model estimation. Another relevant extension would be to combine the world wide data of Dimson, March and Staunton (2002) with the predictive methodology of this paper.

We first motivate our choice of regression variables by noting the basic relationship between stock returns and economic factors in Section 2. In Section 3 we describe our data. Our framework for prediction is given in Section 4 followed by the prediction results based on the data set 1922-1996 in Section 5 when dividend yield alone is used for prediction. In Section 6 we consider the use of more regression variables for the data set 1922-1996 and in Section 7 we shortly comment on the updated results for the data set 1922-2001.

³ This remark is from a version of this paper dated february 2001. A talk based on the conclusions of this paper was given to the Danish Actuarial Society in december 2000 under the title "Be careful: the Danish stocks are too expensive".

2. THE BASIC RELATIONSHIP BETWEEN STOCK RETURNS AND ECONOMIC FACTORS

One traditional equation for the value of a stock is

$$P_t = \sum_{j=1}^{\infty} (1 + \gamma)^{-j} (1 + g)^{j-1} D_t,$$

where most of the entering quantities on the right hand side are unknown, γ , discount rate, g , constant growth of dividend yields, i inflation and D_t is the dividend paid out during the period t . This model was introduced to the financial theory by Williams (1938) and Gordon and Shapiro (1956). Campbell and Shiller (1988) referred to the model as the “dividend-ratio” in absence of uncertainty, see also Goetzman and Jorion (1993), Hodrick (1992), and Fama and French (1988). For simplicity the discount rate and the growth rate do not depend on time in this model although this is well known to be incorrect. The point of the above identity is however, that it shows that the price of stocks depend on quantities such that dividend yield, interest rate and inflation. The two latter being highly correlated with almost any relevant discount rate. It is also clear from the above identity that a decrease in discount rate, which is highly correlated with an increase in bond yield, are related to an increase in the stock return and vice versa.

3. THE DATA AND OUR DEFINITION OF PREDICTION

In this paper we use the annual Danish stock market data from Lund and Engsted (1996), respectively the extended sample period 1922-1996 from Engsted and Tangg ard (2000). We have ourselves extended the period to 1922-2001. We consider the time series

$$W_t = (S_t, d_t, I_t, r_t),$$

where S_t is stock return, I_t is inflation and r_t is the short-term interest rate. The stock index is based on a value weighted portfolio of individual stocks chosen to obtain maximum coverage of the marked index of the Copenhagen Stock Exchange (CBS). Notice that CBS was open during the second world war. In constructing the data corrections were made for stock splits and new equity issues below market prices. Further, $d_t = D_t/P_t$ denotes (nominal) dividends D_t paid during year t divided by the (nominal) stock price P_t at the end of year t . The appendix in Lund and Engsted (1996) contains a detailed description of the data from where we have taken the following quote: “A nominal stock index and accompanying dividend series was constructed from the original price quotation sheets from the Copenhagen Stock Exchange. In order to avoid a possible tax-induced distortion due to the well known January effect, the stock index at the end of the year t is defined as the value in (mid) February of year $t + 1$. Similarly, dividends for year t are defined as dividends paid out between February of year t and February of year $t + 1$. However, no Danish companies pay dividends in

January, so the dividend series is effectively the dividends paid during the year t . Corrections are made for stock splits and new equity issues below the market price using techniques similar to those described in Shiller (1981). The stock index is a value-weighted portfolio consisting of approximately 16 individual stocks (companies), which are generally chosen in order to obtain the maximum coverage of the ‘market’ index of the Copenhagen Stock Exchange. The composition of the stock index is changed about every 10 years, and the weights for the individual stocks are only adjusted in connection with changes in the composition of the stock index.”

We have updated the data set of Lund and Engsted (1996) following their original approach. The leading Danish stock index, the KFX, index has been used for this purpose. As a measure of the short-term interest rate, R_t , the discount rate of the Danish Central Bank’s is used up to 1975, spliced together with a short-term zero-coupon yield for the period thereafter. In computing real values, we deflate nominal values by the consumption deflator⁴. The real excess stock return is defined as

$$S_t = \log\left\{\frac{(P_t + D_t)}{P_{t-1}}\right\} - r_{t-1},$$

where

$$r_t = \log(1 + R_t/100).$$

The resulting average of these excess stock returns are 2.5% for the period 1922-2001 (2.1% for 1922-1996) and 3.4% for the after war period 1948-2001 (3.2% for 1948-1996).

4. OUR FRAMEWORK FOR PREDICTION

The problem of prediction is considered as follows: Let $Y_t = \sum_{i=0}^{T-1} S_{t+i}$ be the excess stock return at time t over the next T years. We base our prediction on the assumption that Y_t can be approximated by a model of the form:

$$Y_t = g(W_{t-1}) + \epsilon_t, \quad t \in \{K_1, \dots, K_2\}, \tag{1}$$

where the error variable ϵ_t are mean zero stochastic variables given the past, W_1, \dots, W_{t-1} and S_1, \dots, S_{t-1} ⁵. Ideally we would like to be able to predict Y_t . We do, however, only have information of W_{t-1} and no information with respect to the error term ϵ_t . Therefore, estimating $g(\bullet)$ and using it for our

⁴ The consumption series has been from Hansen (1974) and various publications from the Danish Central Statistical Bureau. The consumption series covers private consumption of durable and non-durable goods. Unfortunately, there is no price deflator for private consumption, so nominal prices, dividends and consumption are deflated using the consumption deflator for total consumption.

⁵ Note that in our implementation W_{t-1} does not contain time lagged information as is the case e.g in the Wilkie model (1995). An investigation of time lagged variables would be an obvious extension of our approach to prediction.

prediction is the best we can do. Due to the definition of Y_t the time period (K_1, K_2) depends on T and is $(T_{first}, T_{last} - T + 1)$ with $T_{first} = 1923$ or 1949 and $T_{last} = 1996$ or 2001 .

Let

$$X = \{(W_{K_1-1}, S_{K_1}), \dots, (W_{K_2-1}, S_{K_2})\}.$$

For t 's where $K_1 \leq (t - T) \leq K_2$ we wish to be able to consider data points which exclude direct information about S_t . We therefore introduce

$$X^{(t)} = \{(W_{K_1-1}, S_{K_1}), \dots, (W_{t-T-1}, S_{t-T}), (W_{t+T-1}, S_{t+T}), \dots, (W_{K_2-1}, S_{K_2})\}.$$

Now let the set H represent different estimation principles and let for $\hat{g}_h, h \in H$, be some estimator based on X and let $\hat{g}_h^{(t)}$ be the (equivalent) estimator based on $X^{(t)}$.

For a given time horizon T , we define the loss of the estimator \hat{g}_h as

$$Q(\hat{g}_h) = \sum_{t=K_1}^{K_2} \{g(W_{t-1}) - \hat{g}_h(W_{t-1})\}^2 \tag{2}$$

which can be estimated by

$$\hat{Q}(\hat{g}_h) = \sum_{t=K_1}^{K_2} \{Y_t - \hat{g}_h^{(t)}(W_{t-1})\}^2, \tag{3}$$

i.e. we predict $g(W_{t-1})$ without the information contained in Y_t , compare also with Appendix 3. Notice that $Q(\hat{g}_h)$ is not estimated well by the goodness of fit measure

$$\bar{Q}(\hat{g}_h) = \sum_{t=K_1}^{K_2} \{Y_t - \hat{g}_h(W_{t-1})\}^2,$$

since this measure always will be in favor of the most complex model. Such complicated models are often in contradiction to the aim of predicting well.

While predicting, we find our optimal prediction scheme by minimizing $\hat{Q}(\hat{g}_h)$ over all principles h . This gives us the best predictor within H .

Now let h_0 correspond to the trivial prediction strategy based on the parametric model

$$Y_t = \mu + \epsilon_t, \tag{4}$$

where μ is estimated by $\hat{\mu} = (K_2 - K_1 + 1)^{-1} \sum_{t=K_1}^{K_2} Y_t$. For a given modeling and estimation principle h , we define our new R^2 value, that we call $R_{V,h}^2$, where V stands for validated, as

$$R_{V,h}^2 = 1 - \frac{\hat{Q}(\hat{g}_h)}{\hat{Q}(\hat{g}_{h_0})}. \tag{5}$$

Notice that $R_{V,h}^2$ measures how well a given model and estimation principle h predicts compared to the simple estimation principle h_0 . If $R_{V,h}^2$ is positive then we say that the modeling and estimation principle h predicts otherwise we say that the principle h does not predict. In the following we suppress h in the notation and rewrite $R_{V,h}^2$ as R_V^2 . No confusion can occur since it will always be clear what h is under consideration. Note that $R_V^2 \in (-\infty, 1]$ and $R_V^2 > 0$. The interpretation of R_V^2 is similar to the one of the classical R^2 that can be defined in a similar way as

$$R^2 = 1 - \frac{\overline{Q}(\hat{g}_h)}{\overline{Q}(\hat{g}_{h_0})}$$

for a strategy h .

We illustrate the difference between our prediction procedure and traditional goodness of fit by considering two different estimators of stock returns. In particular, we consider nonparametric estimators based on the full data

$$W_t = (S_t, d_t, I_t, r_t) \tag{6}$$

and on the simple subset

$$\overline{W}_t = d_t = D_t / P_t. \tag{7}$$

See Appendix 1 for the mathematical definition of the local linear kernel estimators used. The quality of fit of these two models are given in Figure 1, where the estimators of the regression function $g(\bullet)$ are used to fit next years stock return. We talk about fitting rather than predicting, because the graphs are based on an in-sample approach, where $g(\bullet)$ is estimated from the same stock returns as we fit. From the graphs it is quite clear that one can fit our data set pretty well from the full four dimensional time series, whereas the one dimensional time series consisting of the dividend yield alone fits the data much less. Based on a traditional goodness of fit measures as the R^2 value, see for example Kvålseth (1985), one would clearly prefer the four dimensional covariate to predict stock returns to the simpler one dimensional time series based on dividend yield alone. As a matter of fact, a goodness of fit type of procedure will always have a tendency to chose the most complicated model. Kvålseth (1985) is aware that goodness of fit has this problem and suggests a correction using degrees of freedom. In nonparametric regression, it is, however, unclear what degrees of freedom is. Hastie and Tibshirani (1990) give ad hoc suggestions that seemed to work in their simulations for testing using splines but did not work well in other contexts, see e.g. Sperlich, Linton and Härdle (1999) or Müller (1998). There are certainly other selection criteria for (usually particular) nonparametric models as e.g. the improved Akaike criterion of Hurvich, Simonoff and Tsai (1998). Inside their formulae appear also expressions we might interpret as approximations of the degrees of freedom but it is neither clear whether this criterion can be applied

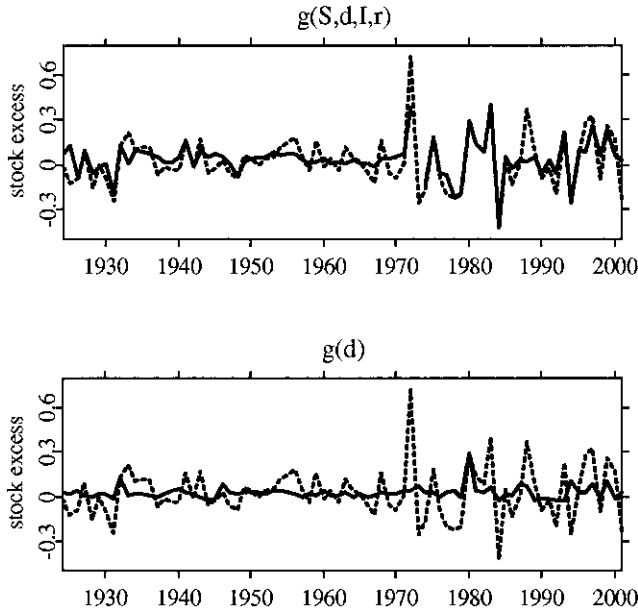


FIGURE 1: Fit of S_t using nonparametric regression (local linear) estimators. Upper curve: based on lagged stock excess, dividend, inflation and short term interest rate, i.e. W_t from (6). Lower curve: based on dividend by price ratio, i.e. \bar{W}_t from (7).

to all the smoothers we use here nor how to interpret the value the criterion takes. We therefore prefer our more straightforward prediction criterion. We will see later that, if we base our conclusions on R_V^2 values, then sometimes we end up with the opposite conclusion of that one we arrived at using the traditional goodness of fit R^2 . Namely, we will see that the dividend yield can indeed help a bit to predict stock returns, while the nonparametric estimator of the full data set is too noisy to be useful for prediction and giving a strongly negative R_V^2 value.

We conclude by pointing out, that even when allowing for any kind of flexible model, it takes a selective choice of the most important explanatory variables to beat even the simple cross validated mean $\hat{Q}(g_{h_0})$ in practical prediction. Indeed, complexity is one of the worst enemies of a good prediction.

5. ESTIMATING AND EVALUATING THE POWER OF PREDICTION

In this section we enter the methodological question of finding a good estimator of prediction power, first we follow Campbell, Lo and Mackinlay (1997, p. 269) and calculate R^2 for different prediction horizons. As mentioned above

we will first concentrate on the period up to 1996. In the first table, Table 1, we consider two versions of the regression

$$Y_t = S_{t+1} + \dots + S_{t+T} = \alpha + \beta\delta_t + \epsilon_{t+T}, \tag{8}$$

where $\delta_t = d_t$ (left-hand columns) or $\delta_t = \ln(d_t)$ (right-hand columns).

TABLE 1
CLASSICAL R^2 VALUES FOR T -YEAR EXCESS STOCK RETURNS ON δ_t , MODEL (8)

<i>horizon</i>	$\delta_t = d_t$		$\delta_t = \ln(d_t)$	
	<i>1923-1996</i>	<i>1949-1996</i>	<i>1923-1996</i>	<i>1949-1996</i>
<i>T</i>				
1	3.8%	7.3%	3.2%	5.9%
2	8.8%	14.9%	6.6%	11.5%
3	13.0%	21.1%	10.5%	17.1%
4	17.5%	25.8%	14.2%	21.0%
5	18.7%	24.2%	15.7%	20.6%

We see that for the linear model the R^2 values are increasing with the time horizon, the same conclusion as Campbell, Lo and Mackinlay (1997, p. 269) arrived at for their American data set. This might imply that prediction over longer horizons is more easy than prediction over short horizons. In the next table, Table 2, we investigate this using our validated criterion for the linear model based on the dividend yield.

TABLE 2
 R_V^2 VALUES FOR T -YEAR EXCESS STOCK RETURNS ON δ_t , MODEL (8)

<i>horizon</i>	$\delta_t = d_t$		$\delta_t = \ln(d_t)$	
	<i>1923-1996</i>	<i>1949-1996</i>	<i>1923-1996</i>	<i>1949-1996</i>
<i>T</i>				
1	-0.2%	1.4%	-1.1%	-0.3%
2	4.9%	8.2%	2.2%	3.0%
3	7.8%	14.2%	4.6%	7.7%
4	10.3%	16.0%	7.4%	9.4%
5	10.3%	9.5%	6.5%	0.5%
6	6.9%	-4.6%	5.2%	-19.5%

We see that, while the quality of prediction is smaller than the R^2 values considered above might suggest, the validated R_V^2 does indeed indicate predictive power of the dividend yield. The period with strongest predictive power seems to be a four year time period with an improved quality of prediction of around 10%. This corresponds to a 10% decrease of the variance of the error term involved in the prediction.

In Table 1 and 2 we see, that a linear regression based on the dividend yield itself instead of the logarithm to the dividend yield gives a better power of prediction. Notice that negative values do not occur in Table 1, but they are present in Table 2. Negative values can not occur with the classical R^2 measure. The classical R^2 measure always favors a more complicated model than the trivial one with a constant mean. The R_V^2 gives negative values in those cases where the prediction model is estimated to perform worse than the trivial model. Since it is indeed very difficult to predict stock markets, the R_V^2 measure will be negative for most attempted prediction models. The surprise here is perhaps that it actually does seem that the dividend yield has predictive power for most of the considered horizons.

Before we get to the nonparametric estimation, we first consider the period 1948-1996 once again, but this time estimation is performed using all the data from 1922 to 1996. However, only the time interval 1948-1996 is used while evaluating the predictive power of the filter. The results are presented in Table 3.

TABLE 3
PREDICTABILITY FOR 1948-1996 OF T -YEAR EXCESS STOCK RETURNS ON d_t ,
RESPECTIVELY ON $\ln(d_t)$, SIMPLE MODEL (8), EVALUATED WITH
THE R_V^2 WHEN USING ALL DATA FROM 1922-1996 FOR PREDICTION

T	d_t	$\ln(d_t)$
1	3.3%	1.8%
2	10.8%	7.3%
3	16.7%	12.4%
4	19.8%	15.6%
5	18.0%	13.2%
6	17.8%	14.9%

The conclusion is that using the entire data set is better while predicting the post war period than just using the post war data itself. It seems that the increased estimation accuracy obtained by using more data outweigh the advantage of only using post war data while estimating the post war period. The argument for the latter methodology is off course that the post war period might be different in nature from the pre war period.

Finally we consider the power of prediction by choosing the functional relationship between the dividend-price ratio and the return by a nonparametric kernel estimator. Specifically, we use local linear kernel estimation what means that in the limit (with bandwidth $h \rightarrow \infty$) the function is linear, and thus, in the limit, coincides with the linear regression, see Appendix 1 and Appendix 2 for details. The bandwidth or smoothing parameter has been chosen such that it maximizes the R_V^2 . Since this functional relationship can be arbitrary, the above discussion on using the raw dividend price ratio or taking the logarithm is irrelevant. We get the results drawn in Table 4.

TABLE 4
 PREDICTABILITY OF T -YEAR EXCESS STOCK RETURNS
 ON USING NONPARAMETRIC MODELS AND MEASURED IN $R^2_{\bar{r}}$.
 EXPLANATORY VARIABLE WAS DIVIDEND YIELD, d_t

T	1923-1996	1949-1996
1	-0.2%	3.3%
2	6.1%	11.5%
3	9.0%	20.7%
4	12.9%	24.5%
5	11.6%	21.7%
6	6.9%	17.8%

Again, when considering the post war period, then data from the entire period is used to fit the nonparametric functional relationship, and the evaluation of the quality of the fit is, however, based on the data in the post war period. While the nonparametric power of prediction for the period 1922-1996 is already slightly better than the strictly linear power of prediction, we see a clear improvement of prediction power for the nonparametric method when considering the period 1948-1996.

For a graphical visualization of the impact of the dividend-price ratio at excess stock returns, see Figure 2 and Figure 3 for respectively the one-year horizon and the four year horizon versions of the prediction of excess stock returns based on the dividend-price ratio. Both the parametric and nonparametric versions are shown. The graphs clearly indicate the impact of the dividend yield on future returns and we also see, that the Danish level of the dividend-price ratio around 1.5% (in 2000) was so low, that according our predictive filter it was a dangerous time to invest in stocks and we did not expect the average excess return on stocks to match this danger. As a matter of fact our model predicted excess returns in the year 2001 to have an average value

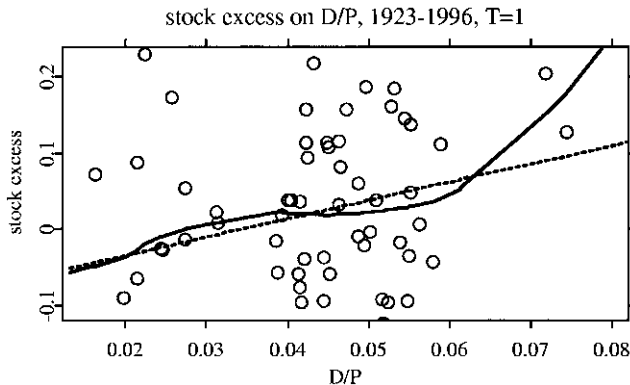


FIGURE 2: Optimal parametric (dashed) and nonparametric (solid) regression fit of stock excess on D/P with real data points.

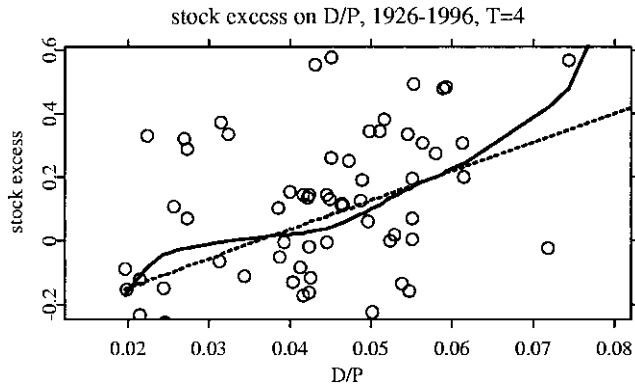


FIGURE 3: Optimal parametric (dashed) and nonparametric (solid) regression fit of stock excess on D/P and real data points.

below zero. So, at that time the extra risk inherent in investments in stocks was not followed by a corresponding extra return on stocks. As a consequence our advice to Danish long term investors was (and is) not to increase their percentage of stocks in their portfolio right now.⁶

6. LOOKING FOR THE BEST PREDICTION MODEL

In this section we investigate the potential advantages that we can obtain by including other variables than just dividend divided by price in our prediction. Due to the complexity of the study of the section, we have chosen to restrict our investigation to a time horizon of one year. Based on the considerations given in Section 2, we have chosen to consider a time series regression problem of the following form:

$$S_t = g(S_{t-1}, d_{t-1}, I_{t-1}, r_{t-1}) + \epsilon_t \quad (9)$$

using the data described in Section 3. The full four-dimensional model corresponds to estimate the function $g(\bullet)$ without any parametric assumptions nor assumptions of structure such as additivity or multiplicativity. This model is most often too complex for both to visualize and/or to predict well. The lack of prediction is due to the error of estimation rather than that the model is insufficient. Therefore we suggest some structure on $g(\bullet)$ to predict well. We have chosen to consider additive models such as

$$g(S, d, I, r) = c + g_1(S) + g_2(d) + g_3(I) + g_4(r), \quad (10)$$

⁶ This remark is from a version of this paper dated february 2001.

compare also Appendix 2, especially for estimation.

Furthermore we consider both the situation where the entering g_i 's are nonparametric and the situation where all the entering g_i 's are parametric and follow a linear model. In our study we consider three types of models with all combinations of subsets of $(S_{t-1}, d_{t-1}, I_{t-1}, r_{t-1})$, namely

- Linear models
- Nonparametric additive models
- Fully nonparametric models

Note that we always applied local linear kernel smoothers applying the bandwidth h that maximizes the R_V^2 , see Appendix 3. The more complex the model is, the bigger the estimation error will be but the smaller the modeling error will be. To be able to choose among the entering models, we use the validated R_V^2 defined in Section 5. All in all, we have 26 models to consider, namely 15 full models (that include automatically the 15 linear models) and 11 nonparametric additive models (leaving out the one-dimensional models that we counted among the full ones). As mentioned and explained in the appendices we always looked for the optimal bandwidths in the nonparametric procedures using Cross Validation, i.e. maximizing our R_V^2 .

Some findings of the estimation respective model structure are the following.

Though the multidimensional nonparametric additive model reaches a positive R_V^2 for some of the considered models, the corresponding full model always did better. This is a clear indicator for having here a more complex structure than additivity. This is not surprising when we consider the complicated relationship between these variables as described in Section 2. From our calculated R_V^2 values we also concluded that the only linear model that does better than the simple constant is the linear model based on the dividend divided by price for the period 1948-2001 as described in the sections before. However, best among all estimators is the fully nonparametric two-dimensional model based on dividend divided by price and lagged excess stock return. This two-dimensional model has a R_V^2 value of 5.5% for the period 1923-1996 and 9.1% for the period 1948-1996. This is much better than the (negative) values of the R_V^2 obtained in Section 5.

Once again have a look on the relation excess returns to dividend by price. In Figure 5 we see the 3-D plot of the two-dimensional predictive filter based on the dividend-price ratio and the lagged excess return of stocks. Further, in Figure 4 we see three slices from this filter plotting the dependency on the dividend yield for three fixed values of excess returns: -25%, 0.7% and 29.5% corresponding to the lower 5% quantile, the median and the upper 95% quantile. We see a clear tendency of the excess stock return to be increasing with the dividend-price ratio. For small dividend yields (below the historical mean of about 4%) the stock return is decreasing with last years excess return. For higher dividend yields (above the historical mean of about 4%) the stock return is increasing with last years excess return. While the intuition of the dependency on the dividend yield is straightforward, it is less straightforward to understand the relationship between last years stock return and current

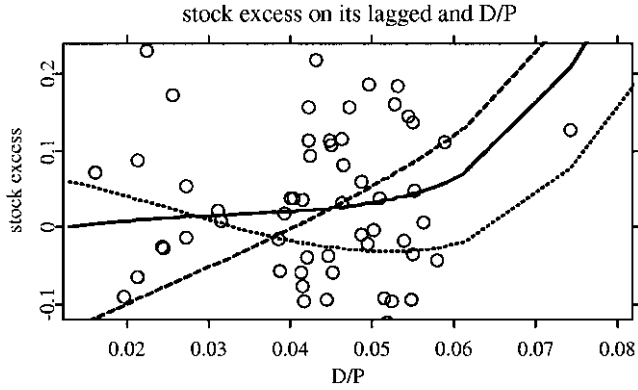


FIGURE 4: Nonparametric regression fits of stock excess on D/P and stock excess lagged fixed at -25% (dotted, starting above zero), at 1% (solid), and at 30% (dashed) for the period 1923-1996.

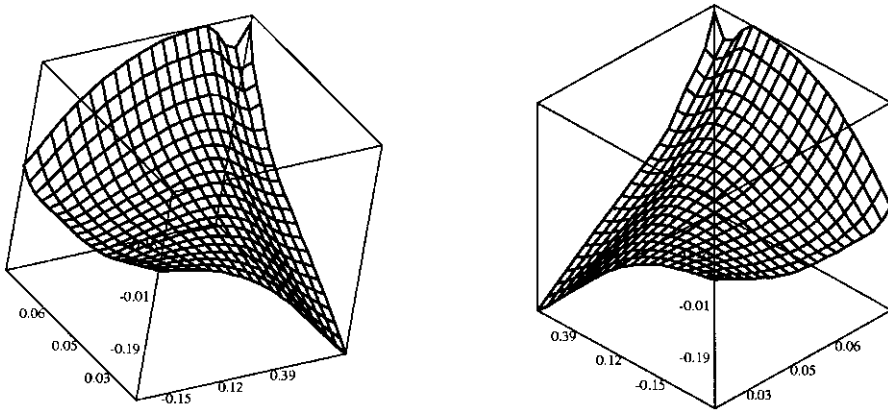


FIGURE 5: Two different views on the nonparametric regression fit of stock excess on D/P and stock excess lagged for the period 1923-1996.

stock return. However, this graph does show that Danish investors should have kept away for new investments in stocks in 2001, since they were just about to finish a magnificent year with a general Danish excess return on stocks above 30% resulting in a historical low dividend-price ratio of around 1.5%. A more detailed picture leading to the same type of conclusions can be found in the three dimensional plot in Figure 5.

7. EMPIRICAL RESULTS FOR 1922-2001

The statistical evidence based on the updated data set does not change the estimated curves and variables very much. However, the estimated predictive power of the updated data set leaves a much less optimistic impression of the

possibility of predicting stock returns than the corresponding results based on the original data set. It is perhaps not surprising for followers of the stock market that the last five years, 1997-2001, have been unusual. Based on the updated data all considered linear models break down – they simply do not predict. This is in contradiction to classical studies like Fama and French (1988), Wilkie (1993) and others and need serious consideration in further work.

However, our main statements still hold when we extend the data set up to 2001. The relationship between the classical R^2 and the validated version, R_V^2 , still play the same role and the full, not additive, nonparametric models still have predictive power.

The optimal R_V^2 is reached for $T = 4$ (time horizon) when only including $d = D/P$. Looking at $T = 1$, the best model for the period 1922-2001 now only uses last year stock return. The model based on both last years stock return and the dividend yield does, however, predicts almost as well. Their predictive powers are respectively, 1.0% and 0.9%. For the time period 1948-2001, we get a relatively impressive R_V^2 of 3.1% while including dividend yield and last years stock return. Optimal bandwidths are between $4\sigma_W$ and $4.8\sigma_W$, so the models are by far not linear. Here, σ_W is the vector of the standard deviations of the different regressors.

Let us first consider Table 5, the corresponding one to Table 1 in Section 5 where we looked at the classical R^2 values for T -year excess stock returns on δ_t , model (8). As one can see clearly, the model fits terribly bad compared to the results obtained for the time period 1922-1996. All R_V^2 values are negative for the linear and log-linear models. We therefore skip here the analogs for Tables 2 and 3 from Section 5.

TABLE 5
CLASSICAL R^2 VALUES FOR T -YEAR EXCESS STOCK RETURNS ON δ_t , MODEL (8)

horizon T	$\delta_t = d_t$		$\delta_t = \ln(d_t)$	
	1923-2001	1949-2001	1923-2001	1949-2001
1	1.6%	3.3%	0.8%	1.8%
2	1.8%	3.3%	0.3%	0.9%
3	1.8%	2.9%	0.2%	0.6%
4	2.6%	3.0%	0.2%	0.2%
5	1.8%	1.2%	0.1%	0.0%
5	1.2%	1.1%	0.0%	0.0%

Table 6 is the corresponding one to Table 4 in Section 5, i.e. we have drawn for the different time horizons T the R_V^2 obtained for the nonparametric model with d_t being the only regressor. The highest R_V^2 is given for $T = 4$ years, but now only with 6.7% for the whole, respectively 12.5% for the post war period.

TABLE 6

PREDICTABILITY OF T -YEAR EXCESS STOCK RETURNS
 USING NONPARAMETRIC MODELS AND MEASURED IN R_V^2 .
 EXPLANATORY VARIABLE WAS DIVIDEND YIELD, d_t

T	1923-2001	1949-2001
1	-1.9%	-0.3%
2	0.0%	2.1%
3	-1.6%	3.5%
4	6.7%	12.5%
5	-10.7%	-5.9%
6	-20.4%	-11.9%

Finally, the Figures 2 to 4 showing the impact of dividend by price ratio $d_t = D_t/P_t$ on the real excess stock returns, stay quite the same when we include the years 1997-2001 into the estimation.

8. CONCLUSIONS

There are mainly three points we make. We first look for a reasonable measure for prediction power (the R_V^2). Second, we use this measure to evaluate the power of prediction of classical as well as more flexible methods. It turns out that the use of nonparametrics methodology and the inclusion of last years stock return significantly improve the level of prediction. Third, fixing the time horizon ($T = 1$) and using flexible methods, we ask for the best prediction model. Finally we illustrate how this can help us for a better understanding of the considered process (discussion of the Figures).

9. APPENDIX

Appendix 1. Local linear kernel regression

In this appendix we give a brief insight into the algorithms of nonparametric flexible function regression. In particular we explain the local linear smoothing. The basic idea is to construct an estimator that lays a smooth surface (or hyperplane), e.g. in the one dimensional case a smooth line, into the point cloud that presents its functional form. The smoothness of that surface can be (pre-) determined by choosing a respectively large *smoothing parameter* h , called bandwidth. Actually, often this parameter can also be data driven, see Appendix 3.

First, it is important to understand that this estimator works locally, e.g. we estimate the wanted function, the hyperplane, at each point we are interested in separately. This is, using the notation $E[Y|W] = g(w)$, $Y \in \mathbb{R}$, $W, w \in \mathbb{R}^d$ with $g(\bullet) : \mathbb{R}^d \rightarrow \mathbb{R}$, an unknown smooth function we estimate

$g(w_0)$ for some point $w_0 \in \mathbb{R}^d$. Having $(W_i, Y_i)_{i=1}^n$ observed, this is done by local least squares:

$$\left(\frac{\hat{g}(x_0)}{\widehat{\nabla g}(x_0)} \right) = \underset{a_0, a_1}{\operatorname{argmin}} \sum_{i=1}^n \{Y_i - a_0 - a_1^T (W_i - w_0)\}^2 K_h(W_i - w_0), \quad (11)$$

$a_0 \in \mathbb{R}, a_1 \in \mathbb{R}^d$ and $\nabla g(\bullet)$ being the gradient of $g(\bullet)$. Further, $K_h(v) = \prod_{j=1}^d \frac{1}{h} K\left(\frac{v_j}{h}\right)$ is a $\mathbb{R}^d \rightarrow \mathbb{R}$ weight function. In our calculations we chose $K(v) = \frac{15}{16}(1 - u^2)^2 \mathbb{1}\{|u| \leq 1\}$. So we used a weighted least squares estimator for linear regression that becomes a local estimator due to the weights K_h giving a lot of weight to points (W_i, Y_i) where W_i is close to w_0 but zero weights to points far from w_0 .

Here, in the weighting function comes the smoothing parameter h in: the larger h and consequently the environment with positive weighting, the smoother gets the resulting hyperplane, i.e. $h \rightarrow \infty$ gives a linear function whereas $h = 0$ yields interpolation of the Y_i 's. Consistency, asymptotic theory and properties are well known and studied for the multivariate case in Ruppert and Wand (1994), for a general introduction see Fan and Gijbels (1996).

An often discussed question is how to choose bandwidth h in practice. As we are concerned about prediction, we take that bandwidth that is minimizing the “out of sample” prediction error using the Cross Validation measure, see Appendix 3.

Appendix 2. Local linear additive regression

We speak of a nonparametric additive model if $g(w), w \in \mathbb{R}^d$ is of the form

$$g(w) = c + g_1(w_1) + g_2(w_2) + \dots + g_d(w_d), \quad \text{with } c = E[Y], \quad (12)$$

and $g(\bullet) : \mathbb{R} \rightarrow \mathbb{R}$ unknown smooth functions with $E[g_j(W_j)] = 0$ for identification. This is the natural extension of the classical linear regression model relaxing the restriction of linear impacts to arbitrary (but smooth) ones. Several procedures are known in the literature, see Sperlich (1998). In this article we focus only on the backfitting by Hastie, Tibshirani (1990). If $g(w)$ is really of additive form, then, under some regularity conditions, this gives us consistent estimators; if not, it tries to estimate the additive model that fits our data best. This is done by iteration: start with some initials $\hat{g}_j^{[0]}(\bullet)$ and $\hat{c} = \frac{1}{n} \sum_{i=1}^n Y_i$.

Then regress $Y - \hat{c} - \sum_{j \neq k}^d \hat{g}_j^{[r-1]}(W_j)$ on W_k to get $\hat{g}_k^{[r]}$ until convergence.

For the regression we applied the local linear kernel estimator. Again, bandwidths can be chosen using Cross Validation, presented in Appendix 3.

Appendix 3. Cross Validation

A typical question of interest, not only in prediction problems, is how to evaluate the different models. This concerns the model or variable selection as well as the bandwidth choice. In general, a natural way to evaluate an estimator is to look on the mean squared error or the expected squared difference between estimate and observation Y : $E[\{Y - \hat{g}(W)\}^2]$. This certainly has to be estimated. Additionally, as we speak about prediction, we would like to know how well the estimator works outside the considered sample. Both aspects are taken into account in the so called Cross Validation (CV) values, defined as

$$\text{CV - value} = \frac{1}{n} \sum_{l=1}^n \{y_l - \hat{g}^{(l)}(w_l)\}^2, \quad (13)$$

where $\hat{g}^{(l)}(w_l)$ is the considered estimator evaluated at point w_l but determined without observation (w_l, y_l) . This CV-value is an approximation for the mean squared error (also for prediction) and a quite common used validation measure in nonparametric regression. For time series context and more references see e.g. Gyöfri, Härdle, Sarda, and Vieu (1990).

Remark: It is important to eliminate always all information that is aimed to predict from the estimation of $g(\bullet)$. So, if we predict the increase of assets over a period of 4 years, the estimator $\hat{g}^{(l)}$ is calculated not only without the l^{th} observation but also without the three years before and after year l .

How can it be used for bandwidth or model selection?

We give an example for bandwidth selection: write \hat{g} as a function of the bandwidth (\hat{g}_h) and look for that h that minimizes

$$\text{CV}(h) = \frac{1}{n} \sum_{l=1}^n \{y_l - \hat{g}_h^{(l)}(w_l)\}^2.$$

This has been shown to give the optimal bandwidth in nonparametric regression, we again refer to Gyöfri et al. (1990).

Note finally, that minimizing the CV-value is equivalent to maximizing the R_V^2 . So CV is directly used to find both: the optimal h for prediction and the optimal model for prediction.

REFERENCES

- BARBER, B.M. and TERRANCE, O. (2000) Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* **55**, 773-806.
- CAIRNS, A.J.G. (2000) A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313-330.
- CAMPBELL, J.Y. and SHILLER, R.J. (1988) The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* **1**, 195-228.
- CAMPBELL, J.Y., LO, A.W. and MACKINLAY, A.C. (1997) *The Econometrics of financial markets*. Princeton University Press, Princeton, New Jersey.
- DIMSON, E., MARSH, P. and STAUNTON, M. (2002) *Triumph of the Optimists*. Princeton University Press.

- ENGSTED, T. and TANGGÅRD, C. (2000) The Danish stock and bond markets: Comovement, return predictability and variance decomposition. *Journal of Empirical Finance* **8**, 243-271.
- EUN, C.S. and RESNICK, B.G. (1988) Exchange rate uncertainty, forward contracts, and international portfolio selection. *Journal of Finance* **43**, 197-215.
- EUN, C.S. and RESNICK, B.G. (1994) Exchange rate uncertainty, forward contracts, and international portfolio selection. *Management Science* **40**, 140-161.
- FAMA, E. and FRENCH, K. (1988) Dividend yields and expected stock returns. *Journal of Financial Economics* **22**, 3-25.
- FAN, J. and GJEBELS, I. (1996) *Local polynomial regression*. Chapman and Hall, London.
- GOETZMANN, W.N. and JORION, P. (1993) Testing the predictive power of dividend yields. *Journal of Finance* **48**, 663-679.
- GORDON, M. and SHAPIRO, P. (1956) Capital Equilibrium analysis: The required rate of profit. *Management Science* **3**, 102-110.
- GYÖFRI, L., HÄRDLE, W., SARDA, P. and VIEU, Ph. (1990) *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.
- HANSEN, S. (1974) *Økonomisk vækst i Danmark [Economic growth in Denmark]*, Akademisk Forlag, Copenhagen, Denmark.
- HARRIS, G.R. (1999) Markov chain Monte Carlo estimation of regime switching vector autoregressions. *ASTIN Bulletin* **29**, 47-79.
- HASTIE, T.J. and TIBSHIRANI, R.J. (1990) *Generalized Additive Models*. Chapman and Hall.
- HODRICK, R.J. (1992) Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* **5**, 357-386.
- HURVICH, C.M., SIMONOFF, J.S. and TSAI, C.-L. (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, B* **60**, 271-293.
- JAGANNATHAN, R. and KOCHERLAKOTA, N.R. (1996) Why should older people invest less in stocks than younger people. *Federal Reserve Bank of Minneapolis. Quarterly Review, Summer 1996*. 11-20.
- KVÅLSETH, T. (1985) Cautionary note about R^2 . *The American Statistician* **39**, 279-285.
- LUND, J. and ENGSTED, T. (1996) GMM and present value tests of the C-CAPM: Evidence from the Danish, German, Swedish, and UK stock markets. *Journal of International Money and Finance* **15**, 497-521.
- MÜLLER M. (1998) Computer-assisted Statistics Teaching in Network Environments. *COMPSTAT '98 Proceedings*, Bristol, UK.
- RICHARDSON, M. and STOCK, J.H. (1989) Drawing inferences from statistics based on multi-year asset returns. *Journal of Financial Economics* **25**, 323-348.
- RUPPERT, D. and WAND, M.P. (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* **22**, 1346-1370.
- SIEGEL, J.J. (1998) *Stocks for the long run*, 2nd ed. New York: McGraw-Hill.
- SHILLER, R.J. (1981) Do stock prices move too much to be justified by subsequent changes in dividends. *American Economic Review* **71**, 421-436.
- SHILLER, R.J. (2000) *Irrational exuberance*. Princeton University Press, Princeton, New Jersey.
- SPERLICH, S. (1998) *Additive Modelling and testing Model Specification*. Shaker Verlag, Aachen.
- SPERLICH, S., LINTON, O., and HÄRDLE, W. (1999) Integration and backfitting methods in additive models-Finite sample properties and comparison. *Test* **8**, 419-458.
- WILKIE, A.D. (1993) Can dividend yields predict share price changes? *Transactions of the 3rd International AFIR Colloquium, Rome*, **1**, 335-347.
- WILKIE, A.D. (1995) More on a stochastic asset model for actuarial use. *British Actuarial Journal* **5**, 777-964.
- WILLIAMS, J.B. (1938) *The Theory of Investment Value*. Harvard University Press, Cambridge.