GEP 2012–02

Nonparametric prediction of stock returns guided by prior knowledge

Michael Scholz, Jens Perch Nielsen, Stefan Sperlich

May 2012

Nonparametric prediction of stock returns guided by prior knowledge

Abstract

One of the most studied questions in economics and finance is whether equity returns or premiums can be predicted by empirical models. While many authors favor the historical mean or other simple parametric methods, this article focuses on nonlinear relationships. A straightforward bootstrap-test confirms that non- and semiparametric techniques help to obtain better forecasts. It is demonstrated how economic theory directly guides a model in an innovative way. The inclusion of prior knowledge enables for American data a further notable improvement in the prediction of excess stock returns of 35% compared to the fully nonparametric model, as measured by the more complex *validated* R^2 as well as using classical out-of-sample validation. Statistically, a bias and dimension reduction method is proposed to import more structure in the estimation process as an adequate way to circumvent the *curse of dimensionality*.

Keywords and Phrases: Prediction of Stock Returns, Cross-Validation, Prior Knowledge, Bias Reduction, Dimension Reduction

JEL Classification Numbers: C14, C53, G17

Michael Scholz
Department of Economics
University of Graz
Universitätstraße 15/F4
A-8010 Graz, Austria

Michael.Scholz@uni-graz.at

Jens Perch Nielsen
Faculty of Actuarial Science and Insurance
Cass Business School
106 Bunhill Row
London, EC1Y8TZ, UK

Jens.Nielsen.1@city.ac.uk

Stefan Sperlich
Département des sciences économiques
Université de Genéve
Bd du Pont d'Arve 40 1211 Genévre 4, Switzerland

Stefan.Sperlich@unige.ch

1 Introduction and Overview

One of the most studied questions in economics and finance is whether equity returns or premiums are predictable. Until the mid-1980's, the view of financial economists was that returns are not predictable, at least not in an economically meaningful way, see for example Fama (1970), and that stock market volatility does not change much over time. Tests of predictability were motivated by *efficient capital markets* and it was common to assume that predictability would contradict to constant expected returns, the efficient markets paradigm.

However, the empirical research in the late twentieth century suggests that excess returns (over short-term interest rates) are predictable, especially over long horizons, as pointed out in Cochrane (1999). For example, Fama and French (1988b) or Poterba and Summers (1988) take only past returns in an univariate mean-reverting sense into account and find rather weak statistical significance, which seems stronger by the inclusion of other predictive variables. In the vast literature, among others, short term interest rates (Fama and Schwert (1977) or Campbell (1991)), yield spreads (Keim and Stambaugh (1986), Campbell (1987), or Fama and French (1989)), stock market volatility (French et al. (1987) or Goyal and Santa-Clara (2003)), book-to-market ratios (Kothari and Shanken (1997) or Ponti and Schall (1998)), and price-earnings ratios (Lamont (1998) or Campbell and Shiller (1988a)) are proposed. There are also numerous articles which examine the predictive power of the dividend yield and, particularly, the dividend ratio on excess stock returns over different horizons. The most influential of them are Fama and French (1988a, 1989), Campbell and Shiller (1988a,b), and Nelson and Kim (1993). For the economic interpretation and the question what drives this predictability, we refer to the discussion in Rey (2004).

By the recent progress in asset pricing theory and the still growing number of publications that report empirical evidence of return predictability it seems that the paradigm of constant expected returns was abandoned. In this spirit, conditional and dynamic asset pricing models (e.g. Campbell and Cochrane (1999)) as well as models that analyze the implications of return predictability on portfolio decisions, when expected returns are time-varying (e.g. Campbell and Viceira (1999)), are proposed. But, certain aspects of the empirical studies cast doubt on the predicting ability of price-based variables and should be considered with caution. While, for example, Fama and French (1988a), Campbell (1991) or Cochrane (1992) find that the aggregate dividend yield strongly predicts excess returns, with even stronger predictability at longer horizons, in contrast, Boudoukh et al. (2008) criticize these findings as an illusion based on the fact that the R^2 of the model is roughly proportional to the considered horizon. Also Ang and Bekaert (2007) find only short-horizon predictability. On the other hand, Rapach et al. (2010) recommend a combination of individual forecasts, including this way the information provided from different variables and reducing forecast volatility. Goyal and Welch (2008) favor the

historical average in forecasting excess stock returns, which gives better results than predictive regressions with different variables, but then again Campbell and Thompson (2008) respond that many of them beat the historical mean by imposing weak restrictions on the signs of coefficients and return forecasts, or by imposing restrictions of steady-state valuation models. Thus, the evidence for stock market predictability is still controversial debated and less transparent than previous work may have suggested.

The most popular model in the economic and financial literature is the discounted-cash-flow or present-value model, which relates the price of a stock to its expected future cash flows, namely, its dividends, discounted to the present value using a constant or time-varying discount rate (e. g. Rozeff (1984), Campbell and Shiller (1987, 1988a,b) or West (1988)). The model assumes the efficient market paradigm of constant expected returns and is based on the well-known discrete-time perfect certainty model (Gordon growth model) and its dynamic generalization. Hence, stock prices are high when dividends are discounted at a low rate or when dividends are expected to grow rapidly. Limitations of this linear model like the apparently exponential growth of stock prices or dividends over time makes it less appropriate than a nonlinear model which can better capture the properties of returns over time as mentioned by Chen and Hong (2009). For example, Froot and Obstfeld (1991) introduce a dividend model with intrinsic bubbles which are nonlinearly driven by exogenous fundamental determinants of asset prices. An other possible extension to the simple model is the use of a log-linear approximation of the present-value relation, see, for example, in Campbell (1991) or Ang and Bekaert (2007). Thus, the asset price behavior can be modeled without imposing restrictions on expected returns. Following these studies and their results that expected asset returns and dividend ratios are time-varying and highly persistent, it is important to model the relationships between equity returns and dividend ratios, interest rates, excess returns, or cash flows in a nonlinear fashion.

In the most empirical studies, the linear predictive regression is applied. Even though this type of model is rather simple, the econometric problems that appear in forecasting asset returns, in testing predictability, and in evaluating the predictive power of the model are numerous. First, the fact that several predictor variables like valuation ratios are highly persistent might cause the found predictability to be spurious. Stambaugh (1999) points out that, although an OLS estimate would be consistent, it is biased and has sampling distributions that differ from those in the standard setting. Also Nelson and Kim (1993) mention that biases affect inference and should be accounted for in practice when studying predictability. These problems become even more serious if data-mining is used. Ferson et al. (2003) show that spurious regression and data-mining effects reinforce each other such that many regressions, based on single predictor variables, may result in spurious conclusions. Possible solutions can be found in Amihud and Hurvich (2004), where an augmented regression is used, in Chiquoine and Hjalmarsson (2009), where an jackknifing procedure is proposed, or in Jansson and Moreira (2006), where infer-

3

ence in a bivariate regression is conducted. Second, an additional source of bias in predictive regressions is the *error-in-variables* problem coming from the fact that, for example, yields contain forecasts of future returns and dividend growth (cf. the discussion in Fama and French (1988a), Goetzmann and Jorion (1995) or Lettau and Ludvigson (2010)), and thus, the explanatory variable is not properly exogenous. Kothari and Shanken (1992) examine the extent to which aggregated stock return variation is explained by variables, chosen to reflect revisions in expectations of future dividends, and provide evidence that the error-in-variables problem is a major one. Third, the main concern in long-horizon predictive regression follows from the use of overlapping data such that error terms are caused to be strongly serially correlated, particularly when the time horizon is relatively large compared to the sample size. Hodrick (1992) examines the statistical properties of different methods for conducting inference in long-horizon regression and his simulations indicate that the test statistics can be substantially biased, but he still concludes with some predictability for U.S. stock market returns. Also Nelson and Kim (1993) analyze small-sample biases in their simulations of a VAR system for returns and dividend yields. Under the null hypothesis of no predictability, they find that the simulated distributions of t-statistics are biased upward by an amount that increases with the horizon and, nevertheless, report predictability of post-war U.S. stock returns. In another simulation, Goetzmann and Jorion (1993) use a bootstraping approach to illustrate how inference may be affected and report only marginal evidence of predictability. More recently, Wolf (2000) uses subsampling for finding reliable confidence intervals—for regression parameters in the context of dependent and possibly heteroscedastic data—and does not find convincing evidence for the predictability of stock returns. Valkanov (2003) shows that, in finite samples where the forecasting horizon is a nontrivial fraction of the sample size, the t-statistics do not converge to a well-defined distribution, and reports only weak predictive power of the dividend yield. Also Ang and Bekaert (2007) find that, at long-horizons, excess return predictability by the dividend yield is not statistically significant using a structural model of equity premiums and accounting for small sample properties. Alternative econometric methods or new statistical tests for conducting valid inference and bias correction can be found in the literature. These studies emphasize that the usual corrections to standard errors are only valid asymptotically and pose the question whether "asymptotic" should be measured in terms of years, decades, or centuries, particularly for long-horizon forecasts. Fourth, Rey (2004) notes that recent theoretical econometric results indicate that these methods fail to provide an asymptotically valid inference when the predictive variable has a near unit root. Lewellen (2004), Torous et al. (2004) or Campbell and Yogo (2006) show that incorporating information about the order of integration can result in large efficiency gains and therefore have a significant effect on inferences. Fifth, while previous studies usually review the inclusion of financial and macroeconomic variables in the linear regression framework (using for example a finite-order VAR system), the functional form

of the regression is not verified. Chen and Hong (2009) mention that, for example, a VAR model cannot fully capture the nonlinear dynamics of dividend yields implied by the present value model. Thus, for a linear regression, one cannot conclude that the null hypothesis of no predictability holds, because there may exist a disregarded nonlinear relationship. Campbell and Shiller (1998) point out that it is quite possible that the true relation between valuation ratios and long-horizon returns is nonlinear. In this case a linear regression forecast might be excessively bearish. But, more and more articles in the literature address this topic. For example, Qi (1999) uses a neural network to examine U.S. stock market return predictability, or Perez-Quiros and Timmermann (2000) apply a Markov switching model for returns of large and small U.S. firms. However, for all of them the functional form is known, while McMillan (2001) examines the relationship between U.S. stock market returns and various predictive variables with a model-free nonparametric estimator. Also Harvey (2001) or Drobetz and Hoechle (2003) analyze conditional expectations of excess returns with nonparametric techniques, but fail to improve forecasts. In contrast, Nielsen and Sperlich (2003) obtain improvements compared to parametric models using a local-linear kernel-based estimator and Danish stock market data. Sixth, different authors, for example, Goyal and Welch (2003, 2008), Butler et al. (2005) or Campbell and Thompson (2008), criticize that most linear predictive regressions have often performed poorly out-of-sample. Particularly, during the bull market of the late 1990's, low valuation ratios predicted extraordinarily low stock returns that did not materialize until the early 2000's (Campbell and Shiller (1998)). It is well-known that useful information on possible misspecified models can be revealed by in-sample diagnostics, while in this way overfitting can be caused or spurious predictability captured. Out-of-sample evaluation could be a possibility to solve these problems and capture the true predictability of a model or a data generating process. For example, Clark (2004) shows with Monte Carlo simulations that out-of-sample forecast comparisons can help prevent overfitting, but in contrast, Inoue and Kilian (2004) conclude that results of in-sample tests of predictability will be more credible due to more power than results of out-of-sample tests. Thus, an overall assessment of return predictability remains difficult, and the question, whether the reason for poor out-of-sample performance of linear prediction models is due to possible nonlinear relations or due to the unpredictability of returns, persists unclear. Numerous studies that use out-of-sample tests have focused on valuation ratios. While, for Fama and French (1988a), the out-of-sample performance of the dividend yield has been a success, Bossaerts and Hillion (1999) discover that even the best prediction models have no out-of-sample forecasting power. Torous and Valkanov (2000) study predictive regressions with a small signal-noise ratio and find that in this case spurious regression is unlikely to be a problem. They further argue that the excessive noisy nature of returns, relative to the explanatory variables, can explain both the apparent in-sample predictability as well as the failure to find out-of-sample forecasting power. Rapach and Wohar (2006) test stock return predictability with

a bootstrap procedure and find that certain financial variables display significant in-sample and out-of-sample forecasting ability. Goyal and Welch (2008) systematically analyze the in-sample and out-of-sample performance of mostly linear regressions and find that the historical average return almost always gives better return forecasts. However, Campbell and Thompson (2008) show that most of the variables used by Goyal and Welch (2008) perform better out-of-sample than the forecast produced with the historical average return, if weak restrictions on the signs of coefficients and return forecasts are imposed. Despite the small out-of-sample explanatory power, they conclude that it is still economically meaningful for investors.

In this paper, we propose a new way to include prior knowledge in the prediction of stock returns. Economic theory directly guides the modelling process. The immediate consequence of that is a dimension and bias reduction, both to import more structure as a proper way to circumvent the curse of dimensionality. First, we start with a fully nonparametric approach which allows the modeling of nonlinearities and interactions of predictive variables. Here, we estimate the model by a local-linear kernel regression smoother which already improves the predictive power in contrast to simple linear versions of the model. The long-lasting popularity of simple predictive regression models justifies the usefulness of the linear method for stock return prediction. However, a model (statistical or from financial theory) can only be an approximate to the real world and thus a linear model can only be seen as a first step in the representation of the unknown relationship in mathematical terms. Second, we include in a semiparametric fashion the available prior information, where the former nonparametric estimator is multiplicatively guided by the prior. This could be, for example, a standard regression model or likewise a good economic model provided by the clever economist. This approach helps to reduce bias in the nonparametric estimation procedure and thus to improve again the predictive power. An economist might provide an economic model better then our structured one. A good economic model should then be validated along the lines of this article. A nonparametric smoother guided by this economic model might be an excellent predictor. Third, we propose a simple bootstrap test to evidence that our method works and does not give better results just by chance. Forth, we apply the proposed technique to American data. For the empirical part of our article, we use the annual data provided by Robert Shiller that include, among other variables, long term stock, bond and interest rate data since 1871 to examine long term historical trends in the US market. It is an updated and revised version of Chapter 26 from Shiller (1989), where a detailed description of the data can be found. Note further that the application to this data set is not meant as a comprehensive study rather as an illustration of the auspicious and potential use of the strategy developed in our article.

Our scope is to show that linear predictive regression models suffer from neglected non-linear relationships and that the inclusion of prior information further improves out-of-sample performance of nonlinear prediction models. Moreover, we evidence that our predictor-based re-

gression models beat the historical average excess stock return. For this purpose, we apply for all models the validated R^2 of Nielsen and Sperlich (2003). This quality measure of the prediction allows directly the comparison of the cross-validated proposed model with the cross-validated historical mean in an out-of-sample fashion. Note further that we also use this instrument to find the optimal bandwidth in non- and semiparametric regression as well as to select the best model.

Note that we do not control and thus allow for non-stationarity, i.e. unit roots, in the predictive variables. Here, we follow the arguments of Torous et al. (2004). They show that due to rational expectations non-stationarities in predictive variables as functions of asset prices, for example dividend by price or earnings by price, can occur. While Park (2010) indicates the impact of a unit root for the predictability of stock returns using linear models, Wang and Phillips (2009) give a convenient basis for inference in a structural nonparametric regression with nonstationary time series when there is a single integrated or near-integrated regressor.

For the American data we find that, due to our bootstrap test, nonlinear models are more adequate than linear regressions, and that the inclusion of prior knowledge greatly improves the prediction quality. With our best prediction model for one-year excess stock returns we not only beat the simple historical mean but we also obtain an essentially improved validated R^2 of 18.5, a relative increase of 35% compared to the best nonparametric model without prior, or a relative increase of 131% compared to the simple regression.

The remainder of the article is structured as follows. Section 2 describes the prediction framework and the used measure of validation. Furthermore, the bootstrap test is introduced and first results of linear and nonlinear models are provided. Section 3 considers the nonparametric prediction that is guided in a new way by prior knowledge. Among others, the dimension reduction approach is evolved. Finally, Section 4 outlines wider results, summarizes the article and gives a short outlook.

2   Preliminaries and First Steps

We consider excess stock returns defined as

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

where D_t denotes the (nominal) dividends paid during year t, P_t the (nominal) stock price at the end of year t, and r_t the short-term interest rate, which is

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

using the discount rate R_t. In our article, we concentrate on forecasts over the one-year horizon, but also longer periods can easily be included with $Y_t = \sum_{i=0}^{T-1} S_{t+i}$, the excess stock return at time t over the next T years.

In the following, we study the prediction problem

$$Y_t = g(X_{t-1}) + \xi_t, \tag{1}$$

where we want to forecast excess stock returns Y_t using lagged predictive variables X_{t-1}, like the dividend-price ratio, d_{t-1}, earnings by price, e_{t-1}, the long-term interest rate, L_{t-1}, the risk-free rate, r_{t-1}, inflation, inf_{t-1}, the bond, b_{t-1}, or also the stock return, Y_{t-1}. The functional form of g is fixed for the simple parametric relationship, but remains fully flexible for the non- and semiparametric counterpart. The error terms ξ_t are mean zero variables given the past. Basically, we address the regression problem of estimating the conditional mean function $g(x) = E(Y|X = x)$ using n i.i.d. pairs (X_i, Y_i) observed from a smooth joint density and its multivariate generalization.

2.1 Out-of-Sample Validation and the more Complex Validated R^2 Measure

Since we use non- and semiparametric techniques, we need an adequate measure for the predictive power. Classical in-sample measures like R^2 or adjusted R^2 cannot be used because various problems occur. For example, the classical R^2 favors always the most complex model or is also inconsistent, if the estimator is inconsistent, as shown by Valkanov (2003). Furthermore, the usual penalization for complexity via a degree-of-freedom adjustment gets meaningless in nonparametrics because it is still unclear what degrees-of-freedom are in this setting. Moreover, in prediction we are not interested in how well a model explains the variation inside the considered sample but, in contrast, would like to know how well it works out-of-sample. For this reasons, we use the validated R^2 of Nielsen and Sperlich (2003) which has some nice features and is defined as

$$R_V^2 = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2}, \tag{2}$$

and we also evaluate our final choice using classical out-of-sample validation methods. Note that in (2) cross-validated values $\hat{g}_{-t}$ and $\bar{Y}_{-t}$ are used, i.e. the (parametric or nonparametric) function g and the historical mean $\bar{Y}$ are predicted at t without the information contained in this point in time, and hence, the R_V^2 is an out-of-sample measure. The validated R^2 is independent of the amount of parameters (in the simple parametric case of g) and measures how well a given model and estimation principal predicts compared to the cross-validated historical mean. This means for positive R_V^2 values, that the predictor-based regression model (1) beats the historical average excess stock return.

Moreover, cross-validation not only punishes overfitting, i.e. pretending a functional relationship which does not really exist, but also allows us to find the optimal (prediction) bandwidth for the non- and semiparametric estimators (cf. Gyöfri et al. (1990)) This means that we use the validated R_V^2 for both, model selection and optimal bandwidth choice.

Note further that in standard out-of-sample tests, which estimate the model up to some year and test on the next years data, the underlying amount of data changes in size for different years. But the standard variance-bias trade-off is extremely dependent on the underlying amount of data. Due to cross-validation, our approach with the R_V^2 has almost the exact correct underlying size of data so that the variance-bias trade-off of our validation is therefore expected to be more accurate than current methods. In other words, we use the R_V^2 measure for our search, because it gives a more correct trade-off of complexity of the model versus available information. In the empirical study, we have 137 years of information. If we cut off those 137 years to say 100 years and use the remaining 37 years for the out-of-sample, our optimization criteria will favour less complex models corresponding to less information (100 years) than the available information (137 years). We tested this standard assumption from model selection by running our methodology by a number of subsamples of varies sizes. As expected smaller samples led to less complex models. More often that not, we ended with one-dimensional regression models when considering sub-samples. Our approach has therefore been to use the R_V^2 measure for our search and then to evaluate our final choice from a classical out-of-sample validation. Moreover, for a stationary process it should not matter, if we skip only the information in point t or all following points in time. The only difference would be that the remaining size of data is to small for the application of non- or semiparametric methods.

2.2 A Bootstrap Test

To show that our method works and does not give better results as the cross-validated historical mean just by chance, we propose a simple bootstrap test. In this, we test the parametric null that the true model is the cross-validated historical mean against a non/semiparametric alternative, i.e. that the true model is our proposed fully nonparametric (5) or semiparametric model with (8). In detail, we estimate the model under the null and under the alternative, and calculate the R_V^2 as well as

$$\tau = \frac{1}{T} \sum_t \left(\hat{g}_{-t} - \bar{Y}_{-t} \right)^2 . \tag{3}$$

The intention is now to simulate the distribution of R_V^2 and τ under the null. Since we do not know the distribution of the underlying random variables, the excess stock returns, we cannot directly sample from them and thus apply the wild bootstrap. It is a stylized fact that stock returns are not normally distributed. Using the wild bootstrap, we avoid this poor

9

Table 1: Predictive power of the simple linear model (4).

	S	d	e	r	L	inf	b
R_V^2	-1.0	1.0	8.0	2.7	-1.1	-1.4	-0.4
R_{adj}^2	0.2	1.7	8.8	3.6	-0.6	-0.4	-0.1

NOTE: Lagged explanatory variables: S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

approximation. For this, we construct B bootstrap samples $\{Y_1^b, \ldots, Y_T^b\}$ using the residuals under the null

$$\hat{\varepsilon}_t^0 = Y_t - \bar{Y}_{-t}$$

and independent and identically distributed random variables with mean zero and variance one, for example, $u_t^b \sim N(0,1)$, such that

$$Y_t^b = Y_t + \hat{\varepsilon}_t^0 \cdot u_t^b.$$

In each bootstrap iteration b, we calculate now the cross-validated mean $\bar{Y}_{-t}^b$ of the Y_t^b, $t = 1, \ldots, T$, as well as the estimates of the alternative model $\hat{g}_{-t}^b$, and, finally, $R_V^{2,b}$ and τ^b like in (2) and (3) with this new estimates. To decide, if we reject or not, we use critical values from corresponding quantiles of the empirical distribution function of the B bootstrap analogues $R_V^{2,b}$ or τ^b, for example, from

$$F^*(u) = \frac{1}{B} \sum_b \mathbb{I}_{\{\tau^b \leq u\}}.$$

This is a well-known testing procedure, which has proved to be consistent in numerous tests, and has therefore been applied, of cause with certain modifications, to many non- or semiparametric testing problems.

2.3   The Simple Predictive Regression

For the sake of illustration, we develop our strategy step by step and start with the simple model. In empirical finance, often the linear predictive regression model

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t \tag{4}$$

is used to evidence predictability of excess stock returns. We are fully aware of the in the introduction mentioned problems with this model, nevertheless, we use it in this basic form, not only as starting point of our empirical study but also as a straightforward possibility to generate a simple prior.

For the American data, Table 1 shows both, the usual adjusted and the validated R^2. More

or less the same values appear, whereas the adjusted R^2 is always greater than the validated R^2. But already Fama and French (1988a) note that the usual in-sample R^2 tend to overstate explanatory power due to possible bias. More important, both measures evidence the earnings yield as the variable with the most explanatory power, i. e. we start our analysis with a validated R^2 of 8.0 and will concentrate on the behavior of models which include this covariate.

Our findings directly confirm to the results of Lamont (1998), who mentions the additional power of the earnings-price ratio for the prediction of excess stock returns in his study using postwar U.S. data. Interestingly, the often used dividend-price ratio gives only poor results.

2.4   The Nonparametric Model

Following the growing evidence of nonlinear behavior in asset returns documented in the literature, we examine the relationship of excess stock returns and the financial variables of the last section using a flexible, because model-free, nonparametric estimator. The model

$$Y_t = g(X_{t-1}) + \xi_t \tag{5}$$

is estimated with a local-linear kernel smoother using the quartic kernel and the optimal bandwidth chosen by cross-validation, i. e. by maximizing the R_V^2 as described in Section 2.1. Note again, that no functional form is assumed. One should further keep in mind that the nonparametric method can estimate the linear function without any bias, since we apply a local-linear smoother. Thus, the simple linear model is automatically embedded in our approach (what is also the case for all of the non- and semiparametric models proposed in the rest of this work). Table 2 shows the results, the validated R^2 and the estimated p-values of the bootstrap test. Remember that we test the parametric null hypothesis, i. e. the true model is the cross-validated historical mean, against the nonparametric alternative, i. e. model (5) holds. The estimated p-value gives the probability that under the null a R_V^2 value can be found which is greater or equal to the observed one. We focus here on the R_V^2 and its estimated p-values, since no essential differences occur between the decisions made for R_V^2 and τ. Nevertheless, we show the τ statistics and its estimated p-values in the corresponding tables, since the distinction of both is the fact that τ basically measures only the variation between the estimates of two procedures, while the R_V^2 compares the fit of them. Using the usual significance levels, we find only the earnings variable with a p-value of 0.005 to be able to forecast stock returns better than the historical mean. We further find an almost factor 1.5 increase in the validated R^2 from 8.0 to 11.8, compared to the simple regression. Note also, that at a 10% level the risk-free rate has small predictive power with a R_V^2 value of 2.5, which is smaller than the one obtained with the linear model.

Figure 1 shows for both variables the estimated linear and nonlinear functions. While for

Table 2: Predictive power of the nonparametric model (5) and corresponding estimated p-values of the bootstrap test.

	S	d	e	r	L	inf	b
R_V^2	-1.2	0.9	11.8	2.5	-0.8	-1.6	-0.7
p-value	0.596	0.193	0.005	0.079	0.571	0.759	0.573
τ	0.029	0.078	0.384	0.139	0.008	0.013	0.026
p-value	0.543	0.253	0.047	0.062	0.643	0.645	0.482

NOTE: Lagged explanatory variables: S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.
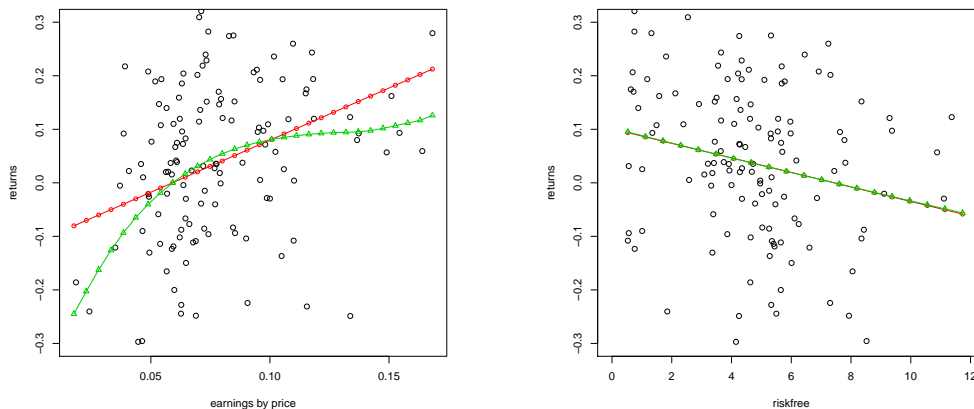
Figure 1: Left: stock returns and earnings by price, Right: stock returns and risk-free; both estimated with linear model (4) (circles) and nonlinear model (5) (triangles)

risk-free an almost identical linear relationship is found, for earnings by price, nonlinearities appear. Economic theory predicts that the short-term interest rate has a negative impact on stock returns. Figure 1 confirms this relationship, since it shows an almost linear declining stock return for an increasing risk-free rate. An increase in the interest rate could raise financial costs, followed by a reduce of future corporate profitability and stock prices. Also the findings for earnings by price agree with the theory. A growing earnings-price ratio makes firms more interesting for investors, and thus stock returns should also increase, as can bee seen in the left part of Figure 1.

Motivated by this results that both, earnings and risk-free, explain to some extent stock returns, we broaden in the next subsection our model to the multivariate case.

Table 3: Predictive power of the two-dimensional linear model (6).

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	6.8	6.9	12.2	7.3	9.2	8.8
R_{adj}^2	8.5	8.7	13.9	8.7	10.7	10.0

NOTE: Lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

2.5 The Multivariate Parametric Model

The natural extension of model (4) is

$$Y_t = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X}_{t-1} + \varepsilon_t, \tag{6}$$

where $\boldsymbol{X}_{t-1}$ can be a vector of different explanatory variables, higher order terms, interactions of certain variables, or a combination of them. But again, we concentrate on the simple case, i.e. we use only two different regressor variables in (6) for creating a simple prior. Table 3 shows the results, the validated and the adjusted R^2, for the regression of lagged earnings by price together with an other variable on stock returns. We find again that the size of both measures is comparable. Moreover, the additional variables inflation, bond yield, and risk-free rate further improve the prediction, compared to the simple model (4) with earnings by price as unique explanatory variable, due to R_V^2 values greater than 8.0. In particular, even the one-dimensional nonparametric model (5) with earnings by price as covariate is outperformed by the multivariate linear model (6) using earnings by price and the risk-free rate as regressors. Here we find a R_V^2 of 12.2 instead of 11.8 for the former one.

3 Nonparametric Prediction Guided by Prior Knowledge

3.1 The Fully Nonparametric Model

To allow the use of more than one explanatory variable in a flexible nonparametric way, we consider the conditional mean equation

$$Y_t = g(\boldsymbol{X}_{t-1}) + \xi_t, \tag{7}$$

where the vector $\boldsymbol{X}_{t-1}$ includes now different regressor variables. Table 4 gives the results, the validated R^2 and the estimated p-value of the proper bootstrap test, using again earnings by price together with another explanatory variable. Here, we find evidence that the appropriate functional form is nonlinear. For all these models we reject at the usual significance levels the null hypothesis that the true model would be the simple historical mean. Moreover, we

Table 4: Predictive power of the fully two-dimensional nonparametric model (7) and corresponding estimated p-values of the bootstrap test.

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	8.5	12.6	13.7	11.0	11.0	11.3
p-value	0.003	0.003	0.000	0.004	0.000	0.000
τ	0.307	0.503	0.485	0.383	0.452	0.430
p-value	0.045	0.018	0.005	0.021	0.011	0.001

NOTE: Lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

find again for all models improved stock return predictions compared to the multivariate linear model (6) because all R_V^2 values are significantly higher. The best model at the moment is the fully two-dimensional one using earnings by price and the risk-free rate, resulting in a R_V^2 value of 13.7, what is a remarkable increase in predictive power of 12% compared to the parametric counterpart.

Here, we only apply two-dimensional models because more complex nonparametric models would not end in better results. Typically, such settings are faced with essential difficulties, like the curse of dimensionality, boundary or bandwidth problems. We will see in the following how it is possible to circumvent or at least to reduce them in the combination of strategies that are usually applied individually.

3.2   Improved Smoothing Through Prior Knowledge

In this subsection, we include prior information in our analysis. This could be, for example, a regression model coming from empirical data analysis or statistical modeling, or likewise a good economic model provided by the clever economist. We limit ourselves to the former because already the use of such simple pilot estimates helps to improve the prediction of stock returns as we will demonstrate in the following. Note that the use of a linear regression model for the prior is in line with a wide array of prominent prediction models from the literature, see, for example, Campbell and Thompson (2008). Moreover, a possible linear trend (driven by one ore even more variables) which non-trivially distorts the data generating process can be detected and (multiplicatively) corrected in our approach.

The basic idea—see, for example, the well written paper of Glad (1998)—is the combination of the parametric pilot from model (4) or (6) and the nonparametric smoother from Subsections 2.4 or 3.1 in a semiparametric fashion, where the latter nonparametric estimator is multiplica-

tively guided by the former parametric and builds on the simple identity

$$g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)}. \tag{8}$$

Remember that we address the regression problem of estimating the conditional mean function $g(x) = E(Y|X = x)$, utilizing its standard solution, the fit of some parametric model $g_\theta(x)$, with the parameter θ, to the data. The essential fact is that if the prior captures some of the characteristics of the shape of $g(x)$, the second factor in (8) becomes less variable than the original $g(x)$ itself. Thus a nonparametric estimator of the *correction factor* $\frac{g(x)}{g_\theta(x)}$ gives better results with less bias.

Note again, that the global pilot could be generated by any parametric technique including simple linear methods, by more complex approaches like nonparametric regression (for the multiplicative bias correction in nonparametric regression, see Linton and Nielsen (1994)) or regression splines with few knots, but also by well-founded economic theory. However, very often even a simple and rough parametric guide is enough to improve the estimate.

From (8) it is obvious that local problems for the above guided approach can occur if the prior itself crosses the x-axis one or more times. Two possible solutions are usually described in the literature. First, a suitable truncation is proposed, i.e. clipping the absolute value of the correcting factor, for example, below 1/10 and above 10 makes the estimator more robust. Second, one could shift all response data Y_i a distance c in such a way that the new prior $g_\theta(x)+c$ is strictly greater than zero and does not anymore intersect the x-axis:

$$g(x) + c = (g_\theta(x) + c) \cdot \frac{g(x) + c}{g_\theta(x) + c}. \tag{9}$$

Note that the estimator becomes for increasing size of c more and more equal to the usual local polynomial which is invariant to such shifts, so that large values of c resolve the intersection problem, but diminish the effect of the guide.

Of course, parameter estimation variability also affects the result, but Glad (1998) shows that there is actually no loss in precision caused by the prior. Even for clear misleading guides she reports the tendency of ignoring the incorrect information and to end up with results similar to that one produced by the fully nonparametric estimator. Also in small samples the guided estimator has strong bias reducing properties. In her experiments, all not too unreasonable guides significantly reduce the bias for all sample sizes and level of noise.

Mainly in the multivariate version, this approach can improve prediction. The reason for it lies in the fact that traditional nonparametric estimators, like the in Section 3.1 presented one, have a rather slow rate of convergence in higher dimensions. Also for a guided multivariate kernel estimator the possibility for bias reduction is essential if the parametric guide captures important

Table 5: Predictive power for model (9).

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	6.6	13.5	12.1	12.8	9.5	8.0

NOTE: In both steps, the prior and estimation of the correction factor, used lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

features of $g(x)$. Note that in the conditional asymptotic bias of the multivariate local-linear estimator the hessian of the true function appears. But for a "quasi linear" correction factor produced by a very good prior, the second derivatives should be very small and thus also the bias. Thus, the idea of guided nonparametric regression turns out to be even more helpful in such a setting.

It is also possible to interpret equations (8) or (9) as an optimal transformation of the nonparametric estimation problem. The subsequent nonparametric smoother of the transformed variables, i. e. of the correction factor, is characterized by less bias. For simple transformation techniques that improve nonparametric regression, see, for example, Park et al. (1997).

Table 5 shows the results, i. e. the validated R_V^2, of models based on (9) which use earnings by price together with another explanatory variable. The same variables are used to generate the simple linear prior with model (6) and to estimate the correction factor. We find that for earnings by price together with dividend by price as well as long term interest rate our strategy helps to improve the prediction power. Compared to the fully two-dimensional model (7) with the same variables, for the former we find a still not satisfying increase of the validated R^2 of 7%, and a notable one of 16% for the latter. For the other variables, our quality measure for the prediction decreases slightly. The reason for this lies in a poor prior or in the fact that the fully two-dimensional smoother already estimates the unknown relationship between stock returns and the used explanatory variables adequately. Note that we skip in Table 5 and in the rest of this article the results of the bootstrap test for the models guided by a prior because we will see that those models result with further improved R_V^2 than the fully nonparametric models (we have already seen in the applied bootstrap tests that the fully nonparametric models are significantly better than the simple historical mean).

3.3 Prior Knowledge for Dimension Reduction

As discussed in previous subsections, fully nonparametric models suffer in several aspects, with increasing number of dimensions, from the curse of dimensionality, and are faced with bandwidth or boundary problems. Since this type of estimator is based on the idea of local weighted averaging, the observations are sparsely distributed in higher dimensions causing unsatisfactory

Table 6: Predictive power for dimension reduction using identity (10).

	S	d	r	L	inf	b
e	8.8	7.6	15.8	10.7	11.4	11.8

NOTE: The prior is generated by a one-dimensional linear regression (4) and uses as lagged explanatory variables S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor is estimated as in model (5) using only e earnings by price.

performance. To circumvent this, it is often proposed to import more structure in the estimation process, like additivity (cf. Stone (1985)) or semiparametric modelling. But, these are not the only possible solutions. Here, our in Section 3.2 proposed approach can also help to import more structure and reduce dimensionality in a multiplicative way. For example, instead of using for both, prior and nonparametric smoother of the correction factor, a two dimensional model, we reduce both to one-dimensional problems, but with different explanatory variables. For this, we first generalize (9) and concentrate on the analog identity

$$g(x_1) + c = (g_\theta(x_2) + c) \cdot \frac{g(x_1) + c}{g_\theta(x_2) + c}. \tag{10}$$

Please keep in mind that this is a separable model of x_1 and x_2. The results of this approach can be found in Table 6. Here, we use the simple linear parametric model (4) with different variables for the prior step. After that we estimate the correction factor with the one-dimensional nonparametric model (5) and earnings by price as covariate. Four of the six in Table 6 presented models improve stock return prediction, as we can observe an increased R_V^2 compared to the fully two-dimensional models from Subsection 3.1. For example, a simple linear prior with the risk-free rate and nonparametric smoother with earnings by price gives a validated R^2 of 15.8, a remarkable increase of 15% compared to our best model so far, the fully two-dimensional one with exact the same variables.

The estimated functions for both models, the fully two-dimensional one (7) and the model guided by prior with (10), as well as for the simple parametric counterpart are shown in Figure 2. Note that we fix one variable at a certain level and plot the relationship of stock returns with the remaining variable. On the left hand side of Figure 2, we fix the risk-free rate at values of 1.0, 6.0, and 12.0. For example, we see that the estimated function, which is guided by the prior, always forecasts negative stock returns for very high earnings by price. In contrast, the parametric and fully nonparametric fit show positive increasing stock returns for earnings by price from a value of 0.11. On the right hand side of Figure 2, we fix earnings by price at 0.03, 0.05, and 0.13. All displayed estimates are more or less linear and find at all levels of earnings by price a linear relationship between stock returns and the risk-free rate. Again, the negative impact of the risk-free rate on stock returns can bee seen. Only for a small earnings-price ratio,

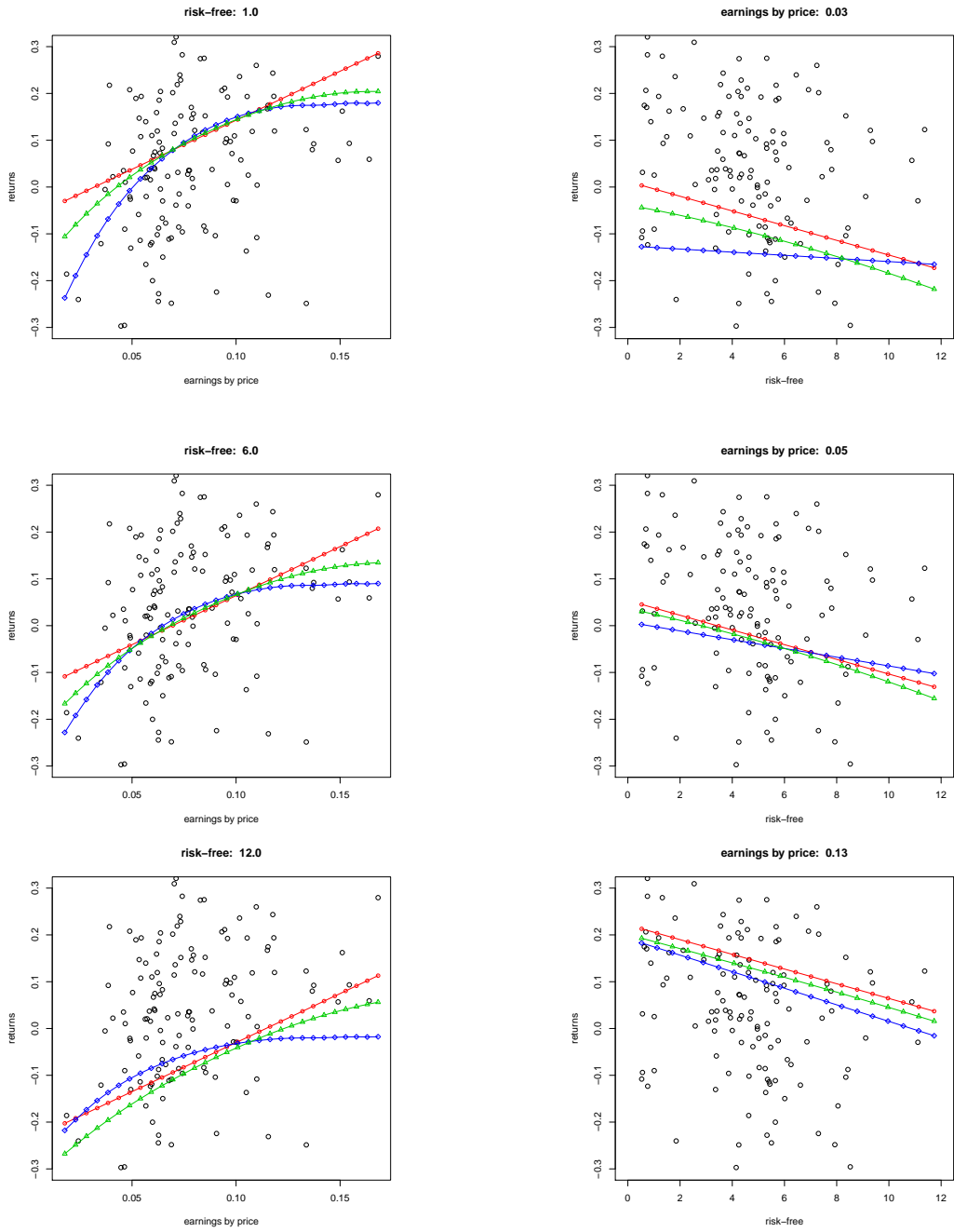**Figure 2:** Left: stock returns and earnings by price at different levels of risk-free, Right: stock returns and risk-free at different levels of earnings by price; both estimated with simple linear model (6) (circles), fully nonparametric model (7) (triangles), and the model guided by prior (10) (diamonds). The simple linear model (4) with the risk-free rate as regressor is used to generate the prior.

Table 7: Predictive power for dimension reduction using identity (10).

	e, S	e, d	e, r	e, L	e, inf	e, b
e	7.5	8.5	16.1	9.9	10.2	10.1

NOTE: The prior is generated by a two-dimensional linear regression (6) and uses as lagged explanatory variables e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor is estimated as in model (5) using only e earnings by price.

the estimator guided by the prior results in an almost constant line, what means that for small earnings by price the risk-free rate has no, or only a small, impact on stock returns.

Note further, that the approach with the prior results in a better fit in the boundary region compared to the fully nonparametric one and thus in more reliable results. The reason for this lies again in the different number of dimensions used for the nonparametric part of the estimators.

3.4 Extensions to Higher Dimensional Models

The above approach can easily be extended in several ways. Here, we consider higher dimensions for x_1 and x_2 in (10) with possible overlapping covariates. For example, we could also use a two-dimensional linear prior in (10) and still estimate the correction factor with a one-dimensional nonparametric model. This results again in an improvement because we find a validated R^2 of 16.1 for the model that uses earnings by price and the risk-free rate for the simple linear prior and only earnings in the nonparametric step, as can be seen in Table 7. This is again a notable increase in predictive power of 18% compared to the best fully nonparametric model.

The other way around is possible too. We use the simple one-dimensional parametric prior (4) together with a fully two-dimensional nonparametric smoother. In the application of this method, we find the results presented in Table 8. For example, using in the simple linear prior step the risk-free rate and in the nonparametric smoother earnings by price and the long-term interest rate, we find an R_V^2 of 18.5, an improvement of impressive 35% compared to the nonparametric model without prior, or an increase of 131% compared to the simple predictive regression, the starting point of our analysis. Also a simple linear prior with the long-term interest rate, together with earnings by price and again long-term interest rate in the nonparametric step, improves the prediction power by remarkable 29% compared to the fully nonparametric version of the model. This results are in accordance with economic theory since the most important part of the stock return is related to the change in interest rates and earnings.

In the above examples, we have seen that the simple extension to identity (10) combines transformation, bias and dimension reduction techniques in a new way and in a single approach, in contrast to the usual proposed separable or additive structures. Thus, boundary and band-

19

Table 8: Predictive power for dimension reduction using identity (10).

	S	d	e	r	L	inf	b
e, L	9.9	13.1	14.2	18.5	13.3	13.3	11.4

NOTE: The prior is generated by a one-dimensional linear regression (4) and uses as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor is estimated as in model (7) using e earnings by price and L long-term interest rate as covariates.
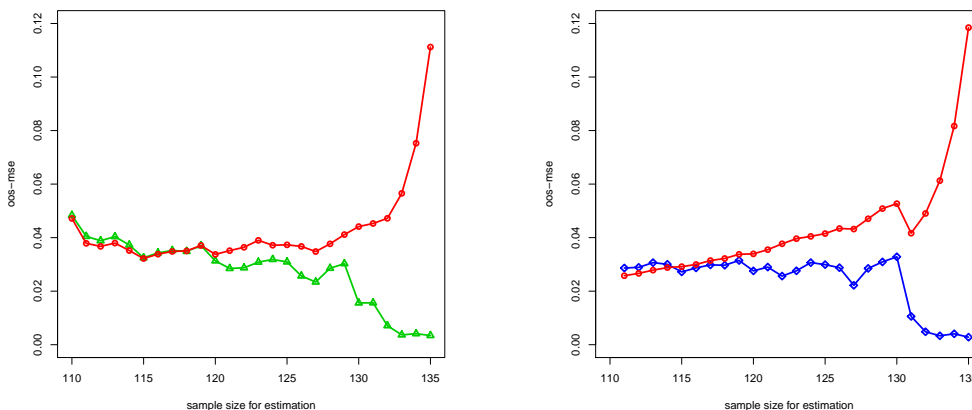

Figure 3: Out-of-sample MSE for different sample sizes. Left: earnings and risk-free in (i) linear model (6) (circles) and (ii) nonlinear model (7) (triangles), Right: earnings, risk-free, and long-term in (i) linear model (6) (circles) and (ii) model guided by prior (10) (diamonds).

width problems are easily alleviated and the curse of dimensionality circumvented.

3.5 Out-of-sample Validation

As already mentioned in Section 2.1, we evaluate our final model choice (i.e. the model based on the full information set) using a classical out-of-sample validation. For this we estimate for different sample sizes the best fully nonparametric model (7), the model guided by prior (10) as well as the corresponding linear models (6). Afterwards we calculate for all of them the out-of-sample MSE and compare the results of the linear model to the non- or semiparametric model in Figure 3. We see that with increasing sample size, i.e. with growing information, the more complex methods achieve lower out-of-sample MSE than the simple regression and that the difference, especially for the new approach with prior, is considerably large. These findings confirm that the validated R^2 is an adequate out-of-sample measure as well as that the innovative idea of including prior information helps to improve forecasting stock returns, not only in-sample but also, and more important, out-of-sample.

20

4 Further Remarks and Conclusions

4.1 Wider Results

Up to now, we concentrated in our article on models which involved the variable earnings by price. Of course, we used other explanatory variables too. The results of such models can be found on the analogy to previous representations in Table 10–14 in the appendix. There, we also give a short overview of the used data. Table 9 presents summary statistics of the available variables. Note that we calculate the inflation variable as the percentage change of the consumer price index and the bond variable as the difference of the ten-year government bond.

As Table 10 and 11 indicate, it is hard to find a model that can better predict than the simple historical mean. But it is not surprising that, once we find such a model, the risk-free rate is an important part of it. For example, we find for the fully nonparametric model, risk-free rate together with dividend by price ($R_V^2 = 3.0$) or long-term interest rate ($R_V^2 = 8.5$), validated R^2 values that are significantly different from zero. However, these models do not have the predictive power found before for the model that uses earnings by price and risk-free ($R_V^2 = 13.7$).

In Table 12–14, we include the already shown results (for earnings by price) for reasons of clarity and comparability. We find that earnings by price consistently gives the best results (in the sense of the largest R_V^2 value), together with the interest rates. Moreover, we see that more complex models do not automatically imply better results. For example, if we use the simple linear prior (4) with the risk-free rate and estimate the correction factor along (10) with model (5) and earnings by price as covariate (see third line in Table 12), we obtain a validated R^2 of 15.8. On the other side, if we also include the risk-free rate when we estimate the correction factor, i. e. with the more complex model (7), we get only a R_V^2 of 10.7 (see third line in Table 14). Furthermore, we stress again that the choice of the prior is crucial. This can bee seen, for example, in line three of Table 13, where we estimate the correction factor with model (5) and earnings by price as covariate. The use of the simple prior (6) with earnings by price and dividend by price gives a R_V^2 of 8.5, while we nearly double ($R_V^2 = 16.1$) the result if we take the same prior but the risk-free rate instead of dividend by price.

4.2 Summary and Outlook

The objective of our article is to show that the prediction of excess stock returns can essentially be improved by the approach of flexible non- and semiparametric techniques. We start with a fully nonparametric model and estimate this with a standard local-linear kernel regression, whereas we maximize the validated R^2 for the choice of the best model and the bandwidth. We further propose a simple wild-bootstrap test which allows us to decide whether we can accept

the parametric null hypothesis, that the historical mean is the right model, or whether we prefer the non- or semiparametric alternative. After we have seen the usefulness of the nonparametric approach, we introduce a possibility to include prior knowledge in the estimation procedure. This can be, for example, a good economic model or likewise a simple parametric regression. We indicate, that even the inclusion of the latter in a semiparametric fashion, more precisely, in a multiplicative way, can enormously improve the prediction of stock returns. To illustrate the potential of our method, we apply it to annual American stock market data, which are provided by Robert Shiller and used for several other articles. Our results confirm to economic theory, namely that the most important part of stock returns is related to the change in interest rates and earnings.

To deliver a statistically insight into our method, we mention that, mainly in higher dimensions, a nonparametric approach would suffer from the curse of dimensionality, bandwidth or boundary problems. A possible adjustment for this problem is the imposition of more structure. Our method contributes to this strategy due to its new and innovative idea—a model directly guided by economic theory. We achieve by a simple transformation the combination of bias and dimension reduction, i. e. more structure to circumvent the curse of dimensionality. This means in our case that a reliable prior captures some of the characteristics of the shape of the estimating function, and thus a multiplicative correction can cause a bias and dimension reduction in the remaining nonparametric estimation process of the correction factor. Thus, we present here a method which greatly improves nonparametric regression in combination with a simple parametric technique.

An other possibility to impose more structure in the prediction process of excess stock returns could be the use of same years covariates. Usually, economic theory says that the price of a stock is driven by fundamentals and investors should focus on forward earnings and profitability. Thus, information on same years, instead of last years, earnings or interest rates can improve prediction. The problem which obviously occurs is that this information is unknown and must also be predicted in some way. One possibility could be the two-step approach of Scholz et al. (2011) proposed for the inclusion of the same years bond yield, which is related to the change in interest rates. Furthermore, one should take into account calendar effects or structural breaks, as described for linear models in Paye and Timmermann (2006). As already mentioned, also longer horizons are easily included in the analysis. Here, a possible improvement could be an appropriate error-correction method.

References

Ang, A., Bekaert, G. (2007). Stock Return Predictability: Is It There? Review of Financial Studies 20, 651–707.

Amihud, Y., Hurvich, C.M. (2004). Predictive Regression: A Reduced-Bias Estimation Method. Journal of Financial and Quantitative Analysis 39, 813–841.

Bossaerts, P., Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? The Review of Financial Studies 12, 405–428.

Boudoukh, J. M., Richardson, M. P., Whitelaw, R. F.(2008). The Myth of Long-Horizon Predictability. The Review of Financial Studies 21, 1577–1605.

Butler, A. W., Grullon, G., Weston, J. P. (2005). Can Managers Forecast Aggregate Market Returns? Journal of Finance 60, 963–986.

Campbell, J. Y. (1987). Stock Returns and the Term Structure. Journal of Financial Economics 18, 373–399.

Campbell, J. Y. (1991). A Variance Decomposition for Stock Returns. Economic Journal 101, 157–179.

Campbell, J. Y., Cochrane, J. H. (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. Journal of Political Economy 107, 205–251.

Campbell, J. Y., Shiller, R. J. (1987). Cointegration and Tests of Present Value Models. Journal of Political Economy 95, 1067–1087.

Campbell, J. Y., Shiller, R. J. (1988a). Stock Prices, Earnings, and Expected Dividends. Journal of Finance 43, 661–676.

Campbell, J. Y., Shiller, R. J. (1988b). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. The Review of Financial Studies 1, 195–227.

Campbell, J. Y., Shiller, R. J. (1998). Valuation Ratios and the Long-Run Stock Market Outlook. Journal of Portfolio Management 1, 195–227.

Campbell, J. Y., Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? The Review of Financial Studies 21, 1455–1508.

Campbell, J. Y., Viceira, L. M. (1999). Consumption and Portfolio Decisions when Expected Returns are Time-Varying. Quarterly Journal of Economics 144, 433–495.

Campbell, J. Y., Yogo, M. (2006). Efficient Tests of Stock Return Predictability. Journal of Financial Economics 81, 27–60.

Chen, Q., Hong, Y. (2009). Predictability of Equity Returns over Different Time Horizons: A Nonparametric Approach. Cornell University/Department of Economics. Working Paper.

Chiquoine, B., Hjalmarsson, E. (2009). Jackknifing Stock Return Predictions. Journal of Empirical Finance 16, 793–803.

Clark, T. E. (2004). Can Out-of-Sample Forecast Comparisons Help Prevent Overfitting? Journal of Forecasting 23, 115–139.

Cochrane, J. H. (1992). Explaining the Variance of Price-Dividend Ratios. The Review of Financial Studies 5, 243–280.

Cochrane, J. H. (1999). New Facts in Finance. Economic Perspectives 23, 36–58.

Drobetz, W., Hoechle, D. (2003). Parametric and Nonparametric Estimation of Conditional Return Expectations. WWZ/Department of Finance, Working Paper No. 9/03.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance 25, 383–417.

Fama, E. F., French, K. R. (1988). Dividend Yields and Expected Stock Returns. Journal of Financial Economics 22, 3–25.

Fama, E. F., French, K. R. (1988). Permanent and Temporary Components of Stock Prices. Journal of Political Economy, 96, 246–273.

Fama, E. F., French, K. R. (1989). Business Conditions and Expected Returns on Stocks and Bonds. Journal of Financial Economics 25, 23–49.

Fama, E. F., Schwert, G. W. (1977). Asset Returns and Inflation. Journal of Financial Economics 5, 115–146.

Ferson, W. E., Sarkissian, S., Simin, T. T. (2003). Spurious Regression in Financial Economics? Journal of Finance 58, 1393–1413.

French, K. R., Schwert, G. W., Stambaugh, R. F. (1987). Expected Stock Returns and Volatility. Journal of Financial Economics 19, 3–29.

Froot, K., Obstfeld, M. (1991). Intrinsic Bubbles: The Case of Stock Prices. American Economic Review 81, 1189–1217.

Glad, I. K. (1998). Parametrically Guided nonparametric Regression. Scandinavian Journal of Statistics 25, 649–668.

Goetzmann, W. N., Jorion, P. (1993). Testing the Predictive Power of Dividend Yields. Journal of Finance 48, 663–679.

Goetzmann, W. N., Jorion, P. (1995). A Longer Look at Dividend Yields. Journal of Business 68, 483–508.

Goetzmann, W. N., Santa-Clara, P. (2003). Idiosyncratic Risk Matters!. Journal of Finance 58, 975–1007.

Goyal, A., Welch, I. (2003). The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium? Management Science 49, 639–654.

Goyal, A., Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies 21, 1455–1508.

Gyöfri, L., Härdle, W., Sarda, P., Vieu, Ph. (1990). Nonparametric Curve Estimation from Time Series (Lecture Notes in Statistics). Heidelberg: Springer.

Harvey, C. R.(2001). The Specification of Conditional Expectations. Journal of Empirical Finance 8, 573–638.

Hodrick, R. J.(1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. The Review of Financial Studies 5, 357–386.

Inoue, A., Kilian, L. (2004). In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? Econometric Reviews 23, 371–402.

Jansson, M., Moreira, M. J. (2006). Optimal Inference in Regression Models With Nearly Integrated Regressors. Econometrica 74, 681–714.

Keim, D. B., Stambaugh, R. F. (1986). Prediction of Stock Returns in the Stock and Bond Markets. Journal of Financial Economics 17, 357–390.

Kothari, S., Shanken, J. (1992). Stock Return Variation and Expected Dividens: A Time-Series and Cross-Sectional Analysis. Journal of Financial Economics 31, 177–210.

Kothari, S., Shanken, J. (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 53, 1563–1587.

Lewellen, J. (2004). Predicting Returns With Financial Ratios. Journal of Financial Economics 74, 209–235.

Lettau, M., Ludvigson, S. (2010). Measuring and Modeling Variation in the Risk-Return Trade-off. In: Aït-Sahalia, Y., Hansen, L. P. (Eds.). Handbook of Financial Econometrics, Vol. I. Amsterdam: North-Holland, 617–690.

Linton, O., Nielsen, J. P. (1994). A multiplicative bias reduction method for nonpametric regression. Statistics and Probability Letters 19, 181–187.

McMillan, D. G. (2001). Nonlinear Predictability of Stock Market Returns: Evidence from Non-parametric and Treshold Models. International Review of Economics and Finance 10, 353–368.

Nelson, C., Kim, M. (1993). Predictable Stock Returns: Reality or Statistical Illusion? Journal of Finance 48, 641–661.

Nielsen, J. P., Sperlich, S. (2003). Prediction of Stock Returns: A new way to look at it. Astin Bulletin 33, 399–417.

Park, C. (2010). When does the dividendprice ratio predict stock returns? Journal of Empirical Finance 17, 81-101.

Park, B. U., Kim, W. C., Ruppert, D., Jones, M. C., Signorini, D. F., Kohn, R. (1997). Simple Transformation Techniques for Improved Nonparametric Regression. Scandinavian Journal of Statistics 24, 145–163.

Paye, B. S., Timmermann, A. (2006). Instability of Return Prediction Models. Journal of Empirical Finance 13, 274-315.

Perez-Quiros, G., Timmermann, A. (2000). Firm Size and Cyclical Variations in Stock Returns. Journal of Finance 55, 1229–1262.

Ponti, J., Schall, L. D. (1998). Book-to-Market Ratios as Predictors of Market Returns. Journal of Financial Economics 49, 141–160.

Poterba, J., Summers, L. (1988). Mean Reversion in Stock Returns: Evidence and Applications. Journal of Financial Economics 22, 27–60.

Qi, M.(1999). Nonlinear Predictability of Stock Returns Using Financial and Economic Variables. Journal of Business and Economic Statistics 17, 419–429.

Rapach, D. E., Wohar, M. E. (2006). In-Sample vs. Out-of-Sample Tests of Stock Return Predictabilty in the Context of Data Mining. Journal of Empirical Finance 13, 231–247.

Rapach, D. E., Strauss, J. K., Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. The Review of Financial Studies 23, 821–862.

Rey, D. (2004). Stock Market Predictability: Is it There? A Critical Review. WWZ/Department of Finance, Working Paper No. 12/03.

Rozeff, M. (1984). Dividend Yields Are Equity Risk Premiums. Journal of Portfolio Management 11, 68–75.

Scholz, M., Sperlich, S., Nielsen, J. P. (2011). Nonparametric Prediction of Stock Returns With Generated Bond Yields. Working Paper.

Shiller, R. J. (1989). Market Volatility. MIT-Press, Cambridge.

Stambaugh, R. (1999). Predictive Regressions. Journal of Financial Economics 54, 375–421.

Stone, C. J. (1985). Additive Regression and Other Nonparametric Models. Annals of Statistics 13, 689–705.

Torous, W., Valkanov, R. (2000). Boundaries of Predictability: Noisy Predictive Regressions. Working Paper, UCLA.

Torous, W., Valkanov, R., Yan, S. (2004). On Predicting Stock Returns with Nearly Integrated Explanatory Variables. Journal of Business 77, 937–966.

Valkanov, R. (2003). Long-Horizon Regressions: Theoretical Results and Applications. Journal of Financial Economics 68, 201–232.

Wang, Q., Phillips, P. (2009). Structural Nonparametric Cointegrating Regression. Econometrica 77, 1901–1948.

West, K. (1988). Dividend Innovations and Stock Price Volatility. Econometrica 56, 37–61.

Wolf, M. (2000). Stock Returns And Dividend Yields Revisited: A New Way To Look At An Old Problem. Journal of Business and Economic Statistics 18, 18–30.

Appendix: Tables of Additional Results

Table 9: US market data (1872-2009).

	Max	Min	Mean	Sd
S&P Stock Price Index	1479.22	3.25	165.08	345.39
Dividend Accruing to Index	28.39	0.18	3.96	6.27
Earnings Accruing to Index	81.51	0.16	8.69	15.54
Stock Returns	0.44	-0.62	0.04	0.18
Dividend by Price	0.09	0.01	0.05	0.01
Earnings by Price	0.17	0.02	0.08	0.03
Short-term Interest Rate	17.63	0.53	4.77	2.77
Long-term Interest Rate	14.59	1.95	4.67	2.27
Inflation	0.21	-0.16	0.02	0.06
Bond	2.03	-4.13	-0.02	0.77

NOTE: Downloadable from `http://www.econ.yale.edu/~shiller/data.htm`.

Table 10: Predictive power of the two-dimensional linear model (6).

	S,d	S,r	S,L	S,inf	S,b	d,r	d,L	d,inf
R^2_V	0.7	1.1	-2.2	-2.3	-1.6	3.5	-0.1	-0.4
R^2_{adj}	2.5	3.4	-0.4	-0.1	-0.1	5.0	1.0	1.3

	d,b	r,L	r,inf	r,b	L,inf	L,b	inf,b
R^2_V	0.8	7.7	1.2	1.5	-2.5	-1.5	-1.9
R^2_{adj}	1.8	8.6	2.9	2.9	-1.1	-0.8	-0.7

NOTE: Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

Table 11: Predictive power of the fully two-dimensional nonparametric model (7) and corresponding estimated p-values of the bootstrap test.

	S,d	S,r	S,L	S,inf	S,b	d,r	d,L	d,inf
R_V^2	0.3	0.5	-2.3	-2.4	-2.1	3.0	-0.3	-0.7
p-value	0.192	0.206	0.686	0.625	0.556	0.043	0.226	0.312
τ	0.123	0.165	0.039	0.050	0.052	0.202	0.079	0.086
p-value	0.240	0.185	0.589	0.574	0.499	0.069	0.233	0.369

	d,b	r,L	r,inf	r,b	L,inf	L,b	inf,b
R_V^2	0.3	8.5	0.7	1.9	-2.5	-1.6	-2.4
p-value	0.131	0.002	0.161	0.071	0.811	0.625	0.718
τ	0.123	0.319	0.140	0.144	0.014	0.026	0.031
p-value	0.082	0.013	0.186	0.101	0.818	0.659	0.665

NOTE: Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

Table 12: Predictive power for dimension reduction using identity (10).

	S	d	e	r	L	inf	b
S	-4.8	-1.0	5.2	-0.5	-3.8	-3.9	-3.2
d	-1.0	-2.1	5.6	1.9	-1.8	-2.0	-0.6
e	8.8	7.6	9.3	15.8	10.7	11.4	11.8
r	-0.6	1.8	10.8	-0.8	-1.2	-1.0	-0.1
L	-3.5	-1.7	5.9	2.3	-3.7	-3.8	-2.6
inf	-3.9	-2.0	8.8	-0.5	-4.3	-4.9	-3.6
b	-3.4	-1.0	7.6	0.7	-3.3	-3.8	-3.3

NOTE: The prior (columns) is generated by a one-dimensional linear regression (4) and the correction factor (rows) is estimated as in model (5). Both use as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield.

Table 13: Predictive power for dimension reduction using identity (10).

	e, S	e, d	e, r	e, L	e, inf	e, b
S	3.5	3.9	9.0	4.4	6.9	5.8
d	4.2	4.1	10.7	5.3	8.0	6.5
e	7.5	8.5	16.1	9.9	10.2	10.1
r	9.0	10.5	9.1	7.5	10.3	10.1
L	4.5	5.4	11.2	4.3	6.7	6.4
inf	7.9	8.8	11.0	7.4	7.8	8.3
b	6.2	6.9	11.4	6.6	7.9	6.0

NOTE: The prior (columns) is generated by a two-dimensional linear regression (6) and uses as lagged explanatory variables e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor (rows) is estimated as in model (5) using only one of the explanatory variables.

Table 14: Predictive power for dimension reduction using identity (10).

	S	d	e	r	L	inf	b
e, S	5.3	2.9	5.8	12.2	7.1	7.9	8.7
e, d	9.6	11.0	10.2	17.1	12.0	12.6	12.5
e, r	9.9	6.8	11.0	10.7	10.1	12.9	12.6
e, L	9.9	13.1	14.2	18.5	13.3	13.3	11.4
e, inf	7.9	5.7	8.9	13.0	9.5	8.7	10.2
e, b	8.3	5.6	8.5	15.9	9.9	10.6	9.1

NOTE: The prior (columns) is generated by a one-dimensional linear regression (4) and uses as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor (rows) is estimated as in model (7) using e earnings by price together with another covariate.

Graz Economics Papers

For full list see:

http://www.uni-graz.at/vwlwww/vwlwww_forschung/vwlwww_gep.htm

Address: Department of Economics, University of Graz,

Universitätsstraße 15/F4, A-8010 Graz

02–2012   **Michael Scholz, Jens Perch Nielsen, Stefan Sperlich**: Nonparametric prediction of stock returns guided by prior knowledge

01–2012   **Ronald Wendner**: Ramsey, Pigou, heterogenous agents, and non-atmospheric consumption externalities