City Research Online

City, University of London Institutional Repository

Citation: Scholz, M., Nielsen, J. P. and Sperlich, S. (2015). Nonparametric Prediction of Stock Returns Based on Yearly Data: The Long-Term View. Insurance: Mathematics and Economics, 65, pp. 143-155. doi: 10.1016/j.insmatheco.2015.09.011

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/12535/

Link to published version: http://dx.doi.org/10.1016/j.insmatheco.2015.09.011

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Nonparametric Prediction of Stock Returns Based on Yearly Data: The Long-Term View

Michael Scholz[a], Jens Perch Nielsen[b,*], Stefan Sperlich[c]

[a] *Department of Economics, University of Graz, Universitaetsstrasse 15/F4, 8010 Graz, Austria*
[b] *Faculty of Actuarial Science and Insurance, Cass Business School, 106 Bunhill Row, London, EC1Y8TZ, UK*
[c] *Geneva School of Economics and Management, Université de Genève, Bd du Pont d'Arve 40, 1211 Genève 4, Switzerland*

Abstract

One of the most studied questions in economics and finance is whether empirical models can be used to predict equity returns or premiums. In this paper, we take the actuarial long-term view and base our prediction on yearly data from 1872 through 2014. While many authors favor the historical mean or other parametric methods, this article focuses on nonlinear relationships between a set of covariates. A bootstrap test on the true functional form of the conditional expected returns confirms that yearly returns on the S&P500 are predictable. The inclusion of prior knowledge in our nonlinear model shows notable improvement in the prediction of excess stock returns compared to a fully nonparametric model. 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: Prediction of Stock Returns, Cross-Validation, Prior Knowledge, Bias Reduction, Dimension Reduction

JEL: C14, C53, C58, G17, G22

*Corresponding author at: Faculty of Actuarial Science and Insurance, Cass Business School, 106 Bunhill Row, London, EC1Y8TZ, UK. Tel.: +44 (0)20 7040 0990.

Email addresses: Michael.Scholz@uni-graz.at (Michael Scholz), Jens.Nielsen.1@city.ac.uk (Jens Perch Nielsen), Stefan.Sperlich@unige.ch (Stefan Sperlich)

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 commonly assumed that predictability would contradict the efficient markets paradigm. In this paper, we take the long-term actuarial view and base our predictions on annual data of the S&P500 from 1872 through 2014. Clearly there are not many historical years in our records and data sparsity is an important issue in our approach. It could be argued that it would be better to use monthly, weekly or even daily data to the extend that more data is available. However, in this context, it cannot be overlooked that the logistics of prediction are very different for yearly, monthly, weekly and daily data. Clearly volatility is key. On the one hand, bias becomes less of an issue when predicting daily data; but on the other hand, bias might be of great importance when predicting yearly data. In the timeframe of a year, it might play a role similar to that of volatility. In other words, the classical trade-off of variance and bias depends on the horizon. A good model for monthly data might be a bad model for yearly data and vice versa. We take the long-term view using yearly data and predict at a one-year horizon. Our reason for doing so is that we are really interested in actuarial models of long-term savings and potential econometric improvements to such models (see, for example, Guillen et al. (2013a); Guillen et al. (2013b); Owadally et al. (2013); Bikker et al. (2012); Guillen et al. (2014); or Gerrard et al. (2014)). It is, therefore, perhaps not a surprise that our favored methodology for validating our sparse long-term yearly data originates from the actuarial literature (see Nielsen and Sperlich (2003)).

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 by Cochrane (1999). For example, Fama and French (1988b) only take into account past returns in a univariate mean-reverting sense and, in doing so, find only rather weak statistical significance, which seems stronger when other predictive variables are included. In the vast array of pertinent literature, short-term interest rates (Campbell (1991)); yield spreads (Fama and French (1989)); stock market volatility (Goyal and Santa-Clara (2003)); book-to-market ratios (Kothari and Shanken (1997)); price-earnings ratios (Lamont (1998)); and consumption-wealth ratios (Lettau and Ludvigson (2001)) are proposed. Numerous other articles examine the predictive power of the dividend yield and, particularly, the dividend ratio on excess stock returns over different horizons. To date, Fama and French (1988a, 1989); Campbell and Shiller (1988a,b); and Nelson and Kim (1993) have published the most influential articles on the subject matter. With regard to the economic interpretation of the driving force behind predictability, we refer to the discussion by Rey (2004). In more literature, Neely et al. (2014) use technical indicators and compare predictive ability with that of macroeconomic variables.

Given the recent progress in asset pricing theory and the continually growing number of publications reporting on empirical evidence for return predictability, it seems that the paradigm of constant expected returns has been abandoned. In that same spirit, conditional and dynamic asset pricing models (e. g. Campbell and Cochrane (1999)) as well as models that analyse the implications of return predictability on portfolio decisions, when expected returns are time-varying (e. g. Campbell and Viceira (1999)), are proposed. Nevertheless, certain aspects of the empirical studies cast doubt on the predictive ability of price-based variables and should be examined with caution. While, for example, Fama and French (1988a) or Campbell (1991) find that the aggregate dividend yield strongly predicts excess returns with even stronger predictability on longer horizons, in contrast, Boudoukh et al. (2008) criticize those same findings as an illusion based on the fact that the R^2 of the model is roughly proportional to the considered horizon. On the other hand, Rapach et al. (2010) recommend a combination of individual forecasts. Goyal and Welch (2008) favor the historical average in forecasting excess stock returns, which yields better results than predictive regressions with different variables. Then again, Campbell and Thompson

(2008) respond that many of them outperform the historical mean either by imposing weak restrictions on the signs of coefficients and return forecasts or by imposing restrictions of steady-state valuation models. Elliott et al. (2013) suggest a new method for combining linear forecasts based on subset regressions and showing improved performance over other classical linear prediction methods. Bollerslev et al. (2009) demonstrate that the variance risk premium is able to explain a nontrivial fraction of the time-series variation in post-1990 aggregate stock market returns, and Bandi and Perron (2008) illustrate the long-run dependence between expected excess market returns and past market variance.

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, which is discounted to the present value using a constant or time-varying discount rate (e. g. Campbell and Shiller (1987, 1988a,b)). 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. The limitations of this linear model, such as the apparently exponential growth of stock prices or dividends over time, make it less suitable than a nonlinear model – which, as mentioned by Chen and Hong (2009), can better capture the properties of returns over time. For example, Froot and Obstfeld (1991) introduce a dividend model featuring intrinsic bubbles that are nonlinearly driven by exogenous fundamental determinants of asset prices. The linear model can also be extended by using a log-linear approximation of the present-value relation (see, for example, Campbell (1991)). Thus, the asset price behavior can be modeled without imposing restrictions on expected returns. In light of results of studies indicating that expected asset returns and dividend ratios are time-varying and highly persistent, it is essential that the relationships between equity returns and dividend ratios, interest rates, excess returns, or cash flows be modeled in a nonlinear fashion.

In this paper, we consider the annual American data provided by Robert Shiller. Shiller's dataset includes, among other variables, long-term stock and bond price changes and interest rate data for the years ranging from 1872 through 2014. It is an updated and revised version of Shiller's (1989) own Chapter 26 which provides a detailed description of the data. It should be noted that the application to this data set is not meant to serve as a comprehensive study but rather as an illustration of the auspicious and potential use of the strategy developed in our article.

We use bootstrap techniques to test the null hypothesis of the non-predictability of returns when using information such as earnings. We then consider predictive models for the returns. We estimate both linear and nonlinear models and use a cross-validated measure of fit to rank the prediction methods. The nonlinear model is estimated by a local-linear kernel regression smoother. This nonlinear smoother has significantly better predictive properties than classical linear models often used for prediction. The long-lasting popularity of 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 approximation to the real world. As such, a linear model can only be seen as a first step in the representation of the unknown relationship in mathematical terms. In a semi-parametric fashion, we then include the available prior information, where the former nonparametric estimator is multiplicatively guided by the prior. This prior could be, for example, a standard regression model or likewise an appropriate economic model. This approach helps to reduce bias in the nonparametric estimation procedure and thus to improve again the predictive power.

The cross-validated measure of performance used here is a generalized version of the validated R^2 of Nielsen and Sperlich (2003). This measure of prediction allows for direct comparisons between proposed models. Furthermore, it should also be noted that we use this measure of performance not only to find the optimal bandwidth in non- and semi-parametric regression but also to select the best model. While we find our cross-validated measure of prediction to work well for our specific application, cross-validation and other data driven measures of model selection may lead to models that are too complex (see, for example, Shao (1993) or Racine (2000)).

We point out that our predictor-based regression models outperform the historical average excess stock return. Moreover, the best prediction model for one-year excess stock returns not only outperforms the historical mean but also obtains an improved validated R^2 of 20.9 – which corresponds to a relative increase of 42% compared to the best nonparametric model without prior and represents a relative increase of 62% compared to the linear regression. We also carry out purely out-of-sample predictions to verify that our conclusions based on validated R^2 do indeed hold true when predicting future yearly returns.

The remainder of the article is structured as follows. Section 2 describes the prediction framework and the applied measure of validation. Here, the bootstrap test is introduced; and the first results of linear and nonlinear models are also provided. Section 3 explores nonparametric prediction as guided in a new way by prior knowledge. Approaches such as the dimension reduction approach are specified, and an out-of-sample validation is carried out. Section 4 serves as a conclusion. Finally, the Appendix contains results for all models and all variables not discussed in the main text.

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. Using the discount rate R_t, the short-term interest rate is expressed as

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

In our article, we concentrate on forecasts covering the one-year horizon. Nevertheless, longer periods can also 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

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

whereby we want to forecast excess stock returns (Y_t), using lagged predictive variables (X_{t-1}) – such as 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 set for the parametric relationship, while remaining fully flexible for non- and semi-parametric counterparts. The error terms (ξ_t) are mean zero variables given the past. 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. We do not assume any explicit distribution for asset returns. An explicit understanding of this distribution (see, for example, Eling (2014)) could perhaps enhance the efficiency of our estimation.

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

Since we use non- as well as semi-parametric techniques, we need an adequate measure of 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 the most complex model and is often inconsistent (see Valkanov (2003)). Furthermore, the usual penalization for complexity via a degree-of-freedom adjustment becomes meaningless in nonparametrics because it remains 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. To that end, we follow the classical approach of estimating the prediction error directly. However, prediction error per se is difficult to interpret intuitively. Hence, in the same vein of R^2, we use a generalized version of Nielsen and Sperlich's (2003) validated R^2 based on leave-k-out cross-validation, which is suitable in a time series context. The validated R^2 is defined as

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

where $\hat{g}_{-t}$ and $\bar{Y}_{-t}$ are leave-k-out estimators of the (parametric or nonparametric) function g and the unconditional mean of Y_t. Both are computed removing k observations around the t-th point in time. Here we use $k = 1$, the classical leave-one-out estimator. Nevertheless, it is well-known that cross-validation often demands the omission of more than one data point and possibly requires some extra correction when the omitted fraction of data is non-negligible (see, for example, Burman et al. (1994)).

The validated R^2 is independent of the amount of parameters (in the parametric case of g) and measures the predictive power of a given model and estimation principal as compared to the cross-validated historical mean. For positive R_V^2 values, that means, that the predictor-based regression model (1) outperforms the historical average excess stock return. Moreover, cross-validation not only punishes instances of overfitting – such as, for example, feigning the existence of a functional relationship that does not really exist – but also allows us to find the optimal (predictive) bandwidth for non- and semi-parametric estimators (cf. Gyöfri et al. (1990)), which means that we use the validated R_V^2 for both model selection and optimal bandwidth choice.

In standard out-of-sample tests, the variance-bias trade-off is extremely dependent on the underlying amount of data. Due to cross-validation, our approach featuring R_V^2 has almost the exactly 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 in our search because it provides for a more correct trade-off of complexity of the model versus available information. In the empirical study, we analyse subsample stability, cut the full information set of 142 years down to 84 years, and estimate the models. In that context, our optimization criteria will favor less complex models corresponding to less information. We tested the standard assumption of model selection by running our methodology on a number of subsamples of various sizes. As expected, smaller samples led to less complex models. More often than not, when considering sub-samples, we ended with one-dimensional regression models. Our approach has therefore been to use the R^2_V measure for our search and then to evaluate our final choice using a classical out-of-sample validation.

2.2. Bandwidth Choice for Nonparametric Estimation

In addition to the R^2_V measure, various other methods for the bandwidth choice in nonparametric estimation are also proposed in the relevant literature. For example, Bandi et al. (2011) consider bandwidth selection for the nonparametric estimation of potentially non-stationary regressions. Although still partly unpublished, there is some literature on nonparametric estimation with non-stationary time-series data. To the best of our knowledge, the current main references are Karlsen and Tjostheim (2001) and Guerre (2004). Both articles rely on ideas of decomposing the process into a recurrent part plus a remainder in order to make consistency statements. Here it should be kept in mind that nonparametric estimators are local, and therefore, require that each "location" be re-visited infinitely many times. The latter article does not restrict the dynamics of the time series to beta-null recurrence and avoids technical smoothness conditions on its invariant measure. Bandi et al. (2011) prefer to stick to the case of the decomposition and assumptions introduced by Karlsen and Tjostheim (2001). As their estimator is a moment estimator based on the first two moments, the bandwidth selector simply looks at the same two (estimated) moment expressions. Thus, the bandwidth is based on first order bias and variance estimators. Recently, Wilhelm (2014) presented a theoretical refinement of this method by adding some higher order approximations – but only for stationary time-series – allowing for some autocorrelation and heteroscedasticity. Again, it is for nonparametric GMM estimation and a theoretical, asymptotically optimal bandwidth. Reviews for almost all the other (relevant) bandwidth selection methods include Heidenreich et al. (2013) for density estimation, and Köhler et al. (2014) for regression.

2.3. A Bootstrap Test

To demonstrate that our method works and does not yield better results than the cross-validated historical mean merely by chance, we propose a bootstrap test. We test the parametric null that the true model is the cross-validated historical mean against a non- or semi-parametric alternative (i. e. that the true model is our proposed fully nonparametric (5) or semi-parametric model with (8)). In detail, we estimate the model not only under the null hypothesis but also 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. We, therefore, apply the wild bootstrap. It is a stylized fact that stock returns are not normally distributed. Using the wild bootstrap, we avoid poor approximation. 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

$$Y_t^b = \bar{Y}_{-t} + \hat{\varepsilon}_t^0 \cdot u_t^b.$$

In each bootstrap iteration b, we now calculate 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 these new estimates. In order to decide, whether to reject or retain, we use critical values from corresponding quantiles of the empirical distribution function of the B bootstrap analogues – that is, from

$$F^*(u) = \frac{1}{B} \sum_b \mathbb{1}_{\{r^b \leq u\}},$$

where r^b is $R_V^{2,b}$ or τ^b, respectively, and $\mathbb{1}_A$ denotes the indicator function of an appropriate set A. It is a standard bootstrap testing procedure for non- as well as semi-parametric testing problems (for details see the survey by Gonzales-Manteiga and Crujeiras (2013) and the citations therein).

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

	S	d	e	r	L	inf	b
R_V^2	-1.4	-0.1	7.1	4.0	-0.8	-1.5	-0.2
R_{adj}^2	-0.2	0.6	7.9	4.8	-0.3	-0.5	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.

2.4. The Linear Predictive Regression

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

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

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

For the American data Table 1 displays both the usual adjusted and the validated R^2, whereby the adjusted R^2 is always greater. Fama and French (1988a) already indicate that the classical in-sample R^2 tend to overstate explanatory power due to possible bias. However, the validated R^2 evidences the earnings yield as the variable with the most explanatory power. Therefore, we will concentrate on the behavior of models that include that covariate and take the R_V^2 of 7.1 as a first reference value.

Our findings directly conform 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 renders only poor results (cf. Boudoukh et al. (2007)).

2.5. The Nonparametric Model

In accordance with the growing evidence of nonlinear behavior in asset returns documented in the relevant literature, we examine the relationship of excess stock returns and the financial variables of the

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

	S	d	e	r	L	inf	b
R_V^2	-1.5	0.8	11.5	3.8	-0.1	-1.5	-0.4
p-value	0.714	0.240	0.008	0.022	0.270	0.772	0.404
τ	0.019	0.106	0.368	0.174	0.016	0.014	0.035
p-value	0.634	0.168	0.034	0.016	0.516	0.648	0.356

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.

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 – that is, by maximizing the R_V^2 as described in Section 2.1. Once again, it should be noted here that no functional form is assumed. Moreover, it should be kept in mind that the nonparametric method can estimate linear functions without any bias, given that we apply a local-linear smoother. Thus, the linear model is automatically embedded in our approach, which is also the case for all of the non- and semi-parametric models proposed in the rest of this article. Table 2 shows the results, the validated R^2, and the estimated p-values of the bootstrap test. It should not be overlooked that we test the parametric null hypothesis – that is, the true model is the cross-validated historical mean against the nonparametric alternative. As such, the model (5) holds. The estimated p-value yields 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 τ statistics and the 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 the earnings variable (p-value 0.008) as well as the risk-free rate (p-value 0.022) to be capable of forecasting

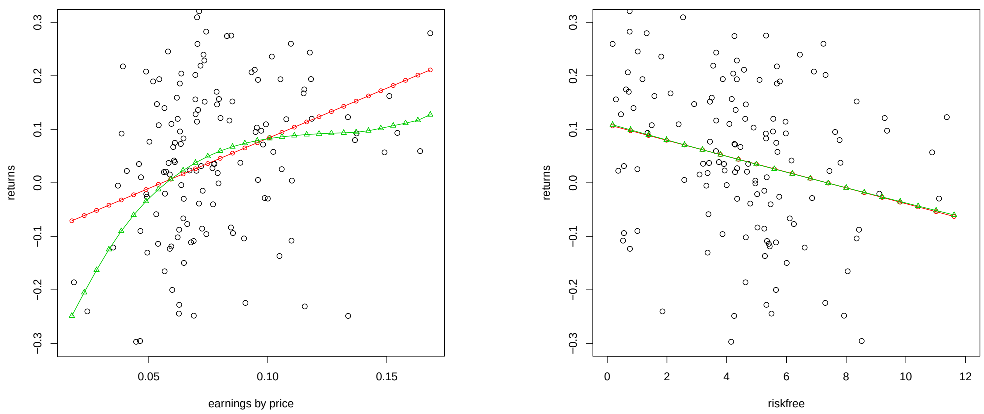

Figure 1: Left: stock returns and earnings by price, Right: stock returns and risk-free; both estimated with linear model (4) (circles) and nonlinear model (5) (triangles)

stock returns better than the historical mean. For earnings by price the validated R^2 increases by 62% from 7.1 to 11.5 (compared to the linear regression).

Figure 1 shows the estimated linear and nonlinear functions for both variables. For risk-free, an almost identical linear relationship is found, while nonlinearities appear for earnings by price. 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 reduction of future corporate profitability and stock prices. Also the findings for earnings by price aline 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 the results indicating that both (earnings and risk-free) explain stock returns to some extent, we expand our model to the multivariate case in the next subsection.

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

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	5.6	6.9	12.9	6.9	7.5	8.2
R_{adj}^2	7.4	8.7	14.5	8.3	9.0	9.5

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.

2.6. 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. Nevertheless, we concentrate on the simple case. More specifically, we use only two different regressor variables in (6) to create 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 another variable on stock returns. We find again that the size of both measures is comparable. Moreover, the additional variables of inflation, bond yield, and risk-free rate further improve prediction power relative to the simple model (4) with earnings by price as a unique explanatory variable due to R_V^2 values greater than 7.1. In particular, the multivariate linear model (6) using earnings by price and the risk-free rate as regressors even outperforms the one-dimensional nonparametric model (5) with earnings by price as covariate. Here we find a R_V^2 of 12.9 instead of 11.5 for the former.

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

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

Table 4: Predictive power (in percent) 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	7.6	12.9	14.7	14.5	11.0	12.1
p-value	0.020	0.004	0.002	0.000	0.000	0.002
τ	0.272	0.504	0.499	0.822	0.686	0.388
p-value	0.064	0.018	0.004	0.000	0.002	0.028

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.

where the vector $\boldsymbol{X}_{t-1}$ now includes different regressor variables. Table 4 shows the results, the validated R^2 and the estimated p-value of the proper bootstrap test, once again using earnings by price together with another explanatory variable. Here, we find evidence that the appropriate functional form is nonlinear. For all these models, at the usual levels of significance, we reject the null hypothesis that the true model would be the historical mean. Moreover, once again for all models, we find improved stock return predictions compared to those of their multivariate linear counterparts (6) because all R_V^2 values are significantly higher. The best model at the moment is the fully two-dimensional one that uses earnings by price and the risk-free rate, resulting in a R_V^2 value of 14.7, which represents an increase in predictive power of 14% compared to that of its parametric counterpart.

Here we only apply two-dimensional models. According to our validation results, three-dimensional nonparametric models all yield worse predictions than those of the best two-dimensional nonparametric models. Typically, such settings are faced with essential difficulties – such as the curse of dimensionality and/or, boundary or bandwidth problems. In the following section, we will see how it is possible to circumvent or at least reduce the influence of those difficulties by using a combination of strategies that are usually applied individually.

3.2. Improved Smoothing Through Prior Knowledge

The basic idea (see Glad (1998)) is the combination of the parametric pilot from model (4) or (6) and the nonparametric smoother from Subsections 2.5 or 3.1 in a semi-parametric fashion, where the latter nonparametric estimator is multiplicatively 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* $\frac{g(x)}{g_\theta(x)}$ yields better results with less bias.

Once again, it should be noted that the global pilot could be generated by any parametric technique including linear methods; either by relying on regression splines with few knots or by using more complex approaches such as nonparametric regressions (for multiplicative bias correction in nonparametric regression, see Linton and Nielsen (1994)); or even by economic theory. However, even a simple and rough parametric guide is quite often enough to improve the estimate.

Judging from identity (8), it is obvious that local problems for the guided approach outlined above can occur if the prior itself crosses the x-axis at least once or several times. Two possible solutions are usually described in the literature. First of all, a suitable truncation is proposed. For example, the absolute value of the correcting factor can be clipped below 1/10 and above 10, making the estimator more robust. Second, all response data Y_t could be shifted a distance of c in such a way that the new prior $g_\theta(x) + c$ is strictly greater than zero and no longer intersects the x-axis:

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

It should be kept in mind that, as c increases in size, the estimator draws continually closer to the usual local polynomial, which is invariant to such shifts. As such, large values of c resolve the intersection problem but diminish the effect of the guide. In the practical application, we choose c according to our validation criterion.

Of course, parameter estimation variability also affects results; but Glad (1998) demonstrates that the prior actually causes no loss in precision. Even for clearly misleading guides, Glad (1998) reports the tendency to ignore incorrect information and, thereby end up with results similar to those yielded by fully nonparametric estimators. The guided estimator even has strong bias reducing properties in small samples. In experiments, Glad (1998) finds that all passably reasonable guides significantly reduce the bias for all sample sizes and the level of noise.

This approach can improve prediction mainly in the multivariate version. The reason for its effectiveness lies in the fact that traditional nonparametric estimators, such as the one presented in Section 3.1, have a rather slow rate of convergence in higher dimensions. Moreover, for a guided multivariate kernel estimator, the possibility for bias reduction is also essential if the parametric guide captures important features of $g(x)$. Here it should be noted that, in the conditional asymptotic bias of the multivariate local-linear estimator, the hessian of the true function appears. However, for a "quasi linear" correction factor produced by a very good prior, not only the second derivatives but also, consequentially, the bias should be very small. Therefore, 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 (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. 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 linear prior with model (6) and to estimate the correction factor. For all models in Table 5, our quality measure for the prediction decreases slightly. The reason lies in a poor prior or in the fact that the fully two-dimensional

Table 5: Predictive power (in percent) for model (9).

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	2.9	9.0	10.4	12.8	8.7	8.4

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.

smoother already estimates the unknown relationship between stock returns and the used explanatory variables adequately. In Table 5 and in the rest of this article, the results of the bootstrap test for the models guided by a prior are omitted given that we will see that those models result in more improved R_V^2 than the fully nonparametric models. In the applied bootstrap tests, we have already seen that the fully nonparametric models are significantly better than the 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 also 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 performance. To circumvent such problems, the importation of more structure in the estimation process – such as additivity (cf. Stone (1985)) or semi-parametric modelling – is often proposed. However, these are not the only possible solutions. Here, our approach proposed in Section 3.2 can aid in the importation of more structure and in the reduction of dimensionality in a multiplicative way. For example, instead of using a two dimensional model for both the prior and the nonparametric smoother of the correction factor, we reduce both to one-dimensional problems with different explanatory variables. Thus, we first generalize (9) and concentrate on the analog identity

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

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

	S	d	r	L	inf	b
e	8.5	8.4	16.5	10.9	11.2	11.7

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.

It should be kept in mind that this model is separable in x_1 and x_2. The results of this approach can be found in Table 6. Here, we use the linear parametric model (4) with different variables for the prior step. Afterwards, we estimate the correction factor with the one-dimensional nonparametric model (5) and earnings by price as covariate. Four of the six models presented in Table 6 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 linear prior with the risk-free rate and nonparametric smoother with earnings by price yields a validated R^2 of 16.5 – which represents an increase of 12% compared to our best model so far, the fully two-dimensional model 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 parametric counterpart are shown in Figure 2. Note that we set 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 set the risk-free rate at values of 1.0, 6.0, and 12.0. For example, we see in the bottom left-hand part of Figure 2 that the estimated function – which is guided by the prior (in diamonds) – always forecasts negative stock returns. In contrast, the parametric and fully nonparametric fit show positive increasing stock returns for earnings by price from a value of 0.12. On the right-hand side of Figure 2, we set earnings by price at 0.03, 0.05, and 0.13. All the 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, the estimator guided by the prior results in an almost constant line – which means that, for small earnings by price, the risk-free rate has no, or only a small, impact on stock

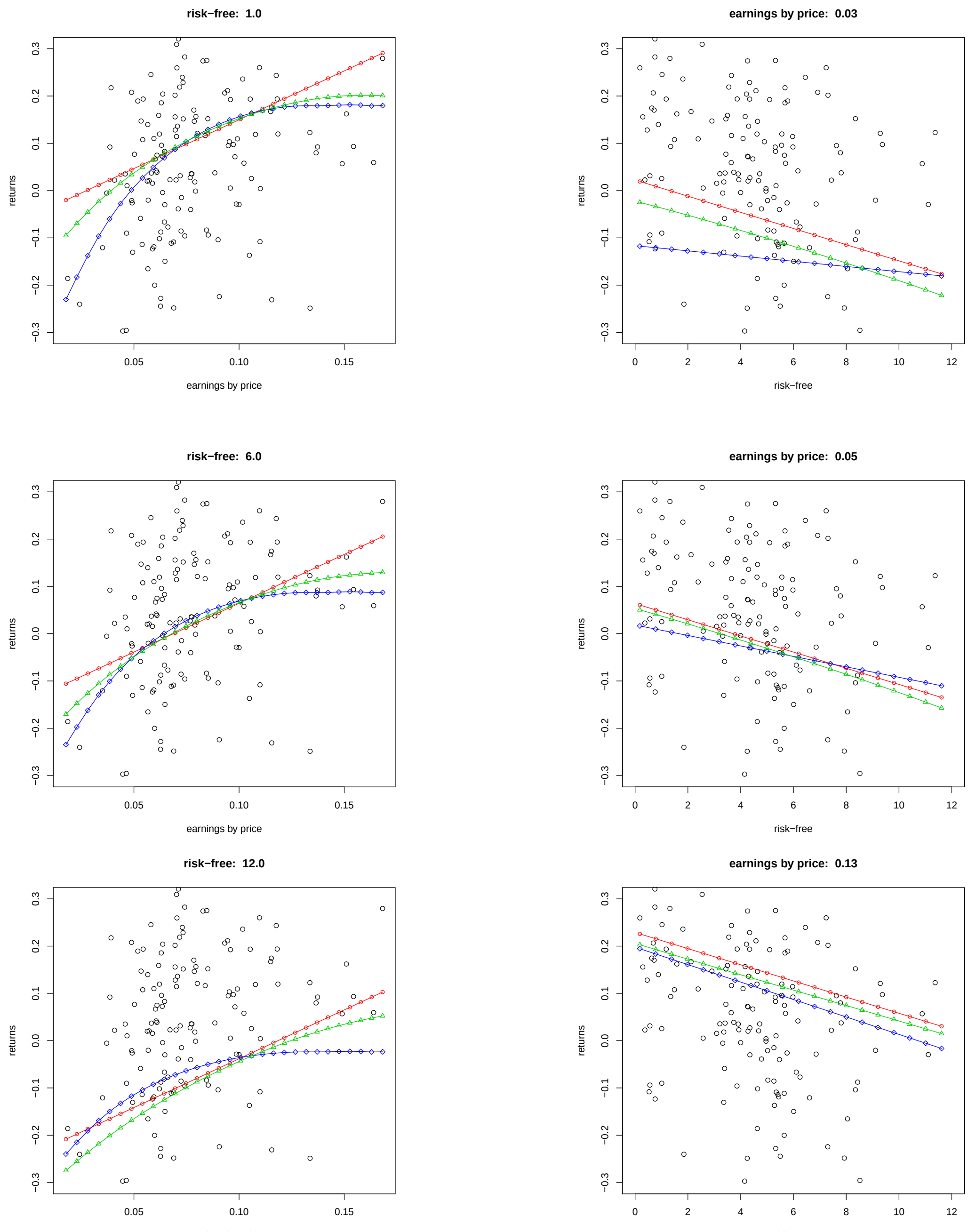

Figure 2: Left: stock returns and earnings by price at different levels of risk-free, Right: stock returns and risk-free at different levels of earnings by price; both estimated with linear model (6) (circles), fully nonparametric model (7) (triangles), and the model guided by prior (10) (diamonds).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

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

	e, S	e, d	e, r	e, L	e, inf	e, b
e	7.5	9.1	16.0	10.0	9.9	10.2

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.

returns.

Moreover, we must point out that the approach featuring the prior does indeed result in a better fit in the boundary region compared to the fully nonparametric one and thus in more reliable results. The reason 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. In doing so, we find a validated R^2 of 16.0 for the model that uses earnings by price and the risk-free rate for the linear prior and only earnings in the nonparametric step, as can be seen in Table 7.

The other way around is also feasible. We use the one-dimensional parametric prior (4) together with a fully two-dimensional nonparametric smoother. The results of the application of this method are presented in Table 8. For example, by using the risk-free rate in the linear prior step and using earnings by price along with the long-term interest rate in the nonparametric smoother, we find an R^2_V of 20.9 – which represents an impressive improvement of 42% compared to the best nonparametric model without the prior or an increase of 62% compared to the predictive regression (the starting point of our analysis). These results are in accordance with economic theory, given that 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 one single approach, in contrast to

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

	S	d	e	r	L	inf	b
e, L	9.8	12.8	14.4	20.9	13.0	13.2	11.5

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.

the commonly proposed separable or additive structures. Thus, boundary and bandwidth 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. We defined an initial estimation sample of 110 observations (from 1872 through 1981). This was used for the purpose of chosing not only the smoothing parameters (bandwidths) in the non- and semi-parametric model but also the variables to be included in the linear model, both based on validated R^2. The remaining sample of 33 observations corresponding to the years from 1982 through 2014, was used to evaluate out-of-sample performance in terms of prediction mean square error (oos-mse) and out-of-sample R^2 (oos-R^2) as defined, for example, by Campbell and Thompson (2008)

$$\text{oos-}R^2 = 1 - \frac{\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{T}(Y_t - \bar{Y}_t)^2},$$

where $\hat{Y}_t$ is the fitted value of the chosen model through period $t-1$, and $\bar{Y}_t$ is the historical mean return also estimated through period $t-1$. Every time a new data point is available, we re-estimate the parameters of the linear model as well as the non- or semi-parametric estimator. It should be noted that the smoothing is kept fixed and at the values determined in the estimation sample. The re-estimation is carried out both using an expanding window as well as a rolling window. We present the results in Figure 3 only for an expanding window since no important differences appeared.

In order to analyse the effect of the length of the estimation sample on out-of-sample performance, we

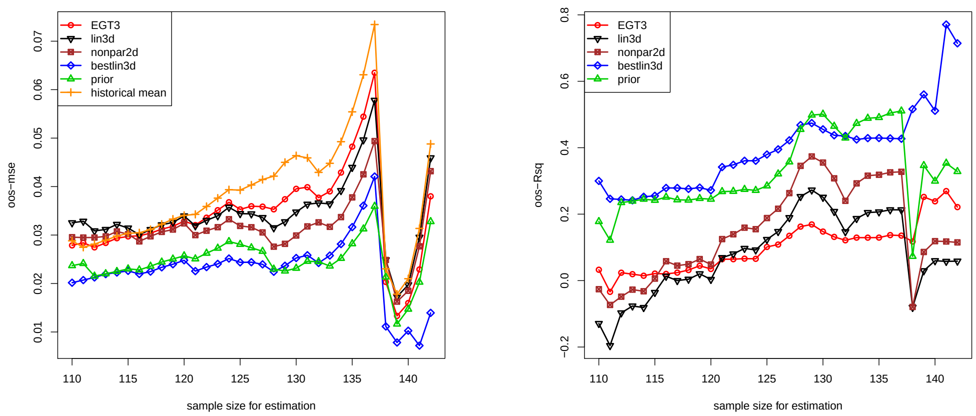

Figure 3: Left: oos-mse for different sample sizes. Right: oos-R^2 for different sample sizes. EGT3: complete subset regression of Elliott et al. (2013) with $k = 3$ (circles), lin3d: linear model (6) on $\{e, r, L\}$ (upside down triangles), nonpar2d: fully nonlinear model (7) on $\{e, L\}$ (squares), bestlin3d: 3-dim. linear model which performs best with hindsight (diamonds), prior: model guided by prior (10) with linear prior on $\{r\}$ and nonparametric correction factor on $\{e, L\}$ (triangles), historical mean (pluses)

then increased the number of observations used for estimation step by step from 110 to 142. In Figure 3, the following six models are compared:

1. EGT3 is based on the complete subset regression by Elliott et al. (2013), where $k = 3$ (circles);
2. lin3d is the linear model (6) on $\{e, r, L\}$ (upside down triangles);
3. nonpar2d is the fully nonlinear model (7) on $\{e, L\}$ (squares);
4. bestlin3d is the three-dimensional linear model that performs best (in terms of oos-mse) with hindsight at each period in time (that is to say that, in hindsight, it is the best combination of 3 variables that one could choose at each period (diamonds));
5. prior is the model guided by prior (10) with a linear prior on $\{r\}$ and nonparametric correction factor on $\{e, L\}$, as chosen with our validation criterion (triangles); and
6. the historical mean.

We have included the best possible historical three-dimensional linear estimator (bestlin3d) in our table.

It is a goodness-of-fit estimator knowing the past and it is therefore infeasible when predicting. We have included it for comparative reasons to illustrate that our feasible predictor based on prior information is close to and sometimes even better than the infeasible linear model. These results provide us with strong arguments in favor of our nonparametric prior-based predictor. Here, it should be noted that the best combination of 3 variables in bestlin3d can also change over time. However, we have found that they are relatively stable. For example, for estimation sample sizes ranging from 111 to 137, the combination $\{d, e, inf\}$ was best, while $\{d, e, r\}$ was best for larger estimation samples. Furthermore, with increasing estimation sample size – that is, with growing information – we see that the more complex prior method even achieves a lower oos-mse than the three-dimensional linear model, which performs best with hindsight. That particular result is mainly driven by the good performance of the prior model during the financial crises in 2008, as can be seen in Figure 4. In that context, the predicted annual excess stock returns of the different models are shown in comparison to the realized annual excess stock returns of the S&P500.

Meanwhile, it should be noted that, in comparison to the fully nonparametric model, we have observed that the inclusion of prior information allows for the use of smaller bandwidths in the nonparametric estimation part in (10). Hence, this new method might better capture nonlinearities.

Predictions around the business-cycle

As a next step, we attempt to analyse the behavior of the predictions of the different models around business-cycle peaks and troughs. Table 9 shows the oos-mse's from 1982 through 2014 (including peaks and troughs respectively). It should be noted that, during the given time period, four years with negative growth rates in real gross domestic product can be observed: 1982, 1991, 2008, and 2009. During the recessions, the prior model performs best. It is also worth mentioning that the historical mean seems to render reasonable predictions around the peaks of the business-cycle, while performing at its worst around the troughs.

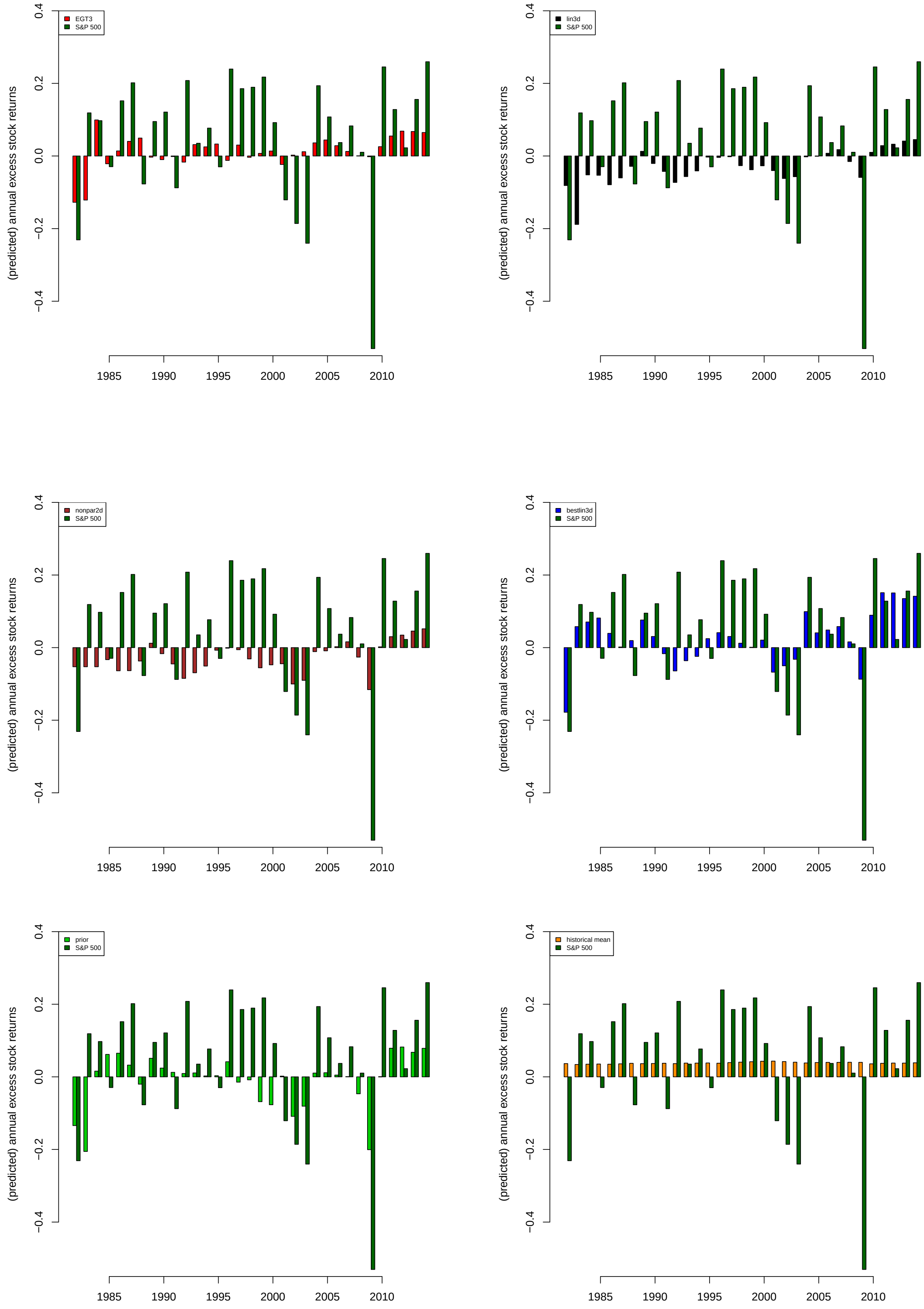

Figure 4: Predicted annual excess stock returns of the different models (cf. Figure 3) in comparison to the realized annual excess stock returns of the S&P500 (First row: EGT3 and lin3d, Second row: nonpar2d and bestlin3d, Third row: prior and historical mean). Note: The data consists of January values such that we observe the largest drawdown caused by the financial crises in 2009.

Table 9: Oos-mse of predictions along the business cylce, oos-period: 1982–2014

model	whole period	business-cycle peaks	business-cycle troughs
EGT3	0.028	0.021	0.074
lin3d	0.033	0.028	0.062
nonpar2d	0.030	0.026	0.052
bestlin3d	0.020	0.016	0.051
prior	0.024	0.022	0.033
historical mean	0.029	0.019	0.103

Performance of simple prediction-based portfolios

Here we are interested in the performance of a portfolio based on the predictions of the different models. We limit ourselves to three simple trading strategies (and ignore transaction costs etc.): (a) buy-and-hold without reinvesting dividends; (b) simple market timing (whereby for positive predictions, we invest 100% of our capital in stocks and 0% in a risk-free asset; and for negative predictions, we invest 0% in stocks and 100% in a risk-free asset); and (c) more conservative market timing (whereby for predictions with an expected return greater than 5%, we invest 100% of our capital in stocks and 0% in a risk-free asset; for predictions with an expected return between 0% and 5%, we invest 50% in stocks and 50% in a risk-free asset; and for negative predictions, we invest 0% in stocks and 100% in a risk-free asset). We repeat this analysis for three different horizons of 33, 15, and 7 years: 1982–2014, 2000–2014, 2008–2014 (based on the predictions displayed in Figure 4). Table 10 shows the compound annual growth rate (cagr) – that is, the year-over-year growth rate of the portfolio over the chosen period – and the maximum drawdown (mdd) of the investment. For the horizon of 33 years and the simple market timing we find, that the prior-based portfolio performs best among the feasible models and is the only one that can even outperform the historical mean. A similar result also holds true for the more conservative strategy. For the shorter periods, portfolios based on prior and EGT3 have similar cagr while in most cases the prior has a significantly lower mdd and should be preferred.

Moreover, it should also be noted that a drawdown in the portfolio only happens when the price of the S&P500 decreases. We analyse excess stock returns and thus it is possible to observe a negative excess

Table 10: Portfolio performance

	1982–2014		2000–2014		2008–2014	
(a) buy-and-hold						
	cagr	mdd	cagr	mdd	cagr	mdd
	9.1	40.7	3.7	40.7	5.1	40.7
(b) simple strategy						
	cagr	mdd	cagr	mdd	cagr	mdd
EGT3	10.9	30.5	11.2	30.5	17.2	0.0
lin3d	9.1	0.0	10.3	0.0	16.5	0.0
nonpar2d	9.1	0.0	10.3	0.0	16.7	0.0
bestlin3d	14.8	2.0	16.1	0.0	17.2	0.0
prior	12.5	5.5	11.3	5.5	16.7	0.0
historical mean	11.0	38.8	8.0	38.8	8.8	38.8
(c) conservative strategy						
	cagr	mdd	cagr	mdd	cagr	mdd
EGT3	9.6	11.7	9.8	11.7	14.4	0.0
lin3d	7.7	0.0	7.5	0.0	10.6	0.0
nonpar2d	8.1	0.0	8.3	0.0	12.3	0.0
bestlin3d	12.6	0.0	13.8	0.0	16.8	0.0
prior	10.5	0.2	9.7	0.2	14.2	0.0
historical mean	9.0	18.1	6.9	18.1	7.3	18.1

NOTE: cagr is the compound annual growth rate and mdd the maximum drawdown (both in percent).

stock return, even when the price increases (but not enough to exceed the risk-free investment), as was the case for example, in the year 1985. The consequence is that, even for a positive prediction, there is no drawdown in the portfolio, which can be observed (e. g., for the bestlin3d model in the conservative strategy).

Subsample Stability

We complete our out-of-sample validation with the issue of subsample stability. Campbell and Thompson (2008) analyse the out-of-sample performance of different models for three subsamples. One source of motivation for this investigation are Goyal and Welch's (2008) findings that valuation ratios have the strongest ability to forecast stock returns in the oos-period 1927–1956, which includes the Great Depression and World War II. We repeat this analysis with the six models introduced above and three

Table 11: Subsample stability

	1927–1956		1956–1985		1985–2014	
model	oos-mse	oos-R^2	oos-mse	oos-R^2	oos-mse	oos-R^2
EGT3	0.055	10.6	0.018	34.5	0.028	1.9
lin3d	0.060	2.9	0.017	35.2	0.025	12.7
nonpar2d	0.052	16.3	0.024	12.0	0.030	-2.6
bestlin3d	0.049	21.4	0.016	41.6	0.024	16.3
prior	0.052	15.9	0.018	34.8	0.024	16.8
historical mean	0.062		0.027		0.029	

similar out-of-sample periods of 30 years: 1927–1956, 1956–1985, and 1985–2014. The results (oos-mse and oos-R^2) can be found in Table 11. In terms of oos-mse, the first oos-period yields the worst results, while the second oos-period has the lowest. Nevertheless, in all subsamples, the prior model performs best; and, in the last period, it even has a higher oos-R^2 than the non-feasible model bestlin3d.

Once again, we emphasize that, especially for shorter estimation periods other (and often less complex) models would have been chosen by our validation criterion for the fully nonparametric and prior models. For example, a fully nonparametric model on $\{r\}$ over the out-of-sample period 1956–1985 would have reached an oos-R^2 of 41.1% and an oos-mse of 0.016.

4. Further Remarks and Conclusions

4.1. Broader Results

Up until this point, in our article, we have concentrated on models involving the variable earnings by price. Of course, we also used other explanatory variables. The results of such models can be found in the analogy to previous representations in Table 13–17 in the appendix. There, we also provide a short overview of the data used in Table 12. Here, it should be noted 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 13 and 14 indicate, it is hard to find a model that can predict better than the historical mean. Nonetheless, it is not surprising that, once we find such a model, the risk-free rate is an important part of it. For example, for the fully nonparametric model, risk-free rate together with dividend by price ($R^2_V = 3.8$) or long-term interest rate ($R^2_V = 10.0$), we find 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 ($R^2_V = 14.7$).

In Table 15–17, we include the previously demonstrated results (for earnings by price) for reasons of clarity and comparability. We find that the variable earnings by price consistently yields the best results (in the sense of the largest R^2_V 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 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 the third line of Table 15), we obtain a validated R^2 of 16.5. On the other hand, 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^2_V of 11.9 (see the third line of Table 17). Furthermore, we must emphasise once again that the choice of the prior is crucial. This can bee seen, for example, in line three of Table 16, where we estimate the correction factor with model (5) and earnings by price as covariate. The use of the prior (6) together with earnings by price and dividend by price yields a R^2_V of 9.1, while we nearly double ($R^2_V = 16.0$) 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 demonstrate that the prediction of excess stock returns can essentially be improved by an approach using flexible non- and semi-parametric techniques. We start with a fully nonparametric model and estimate using a standard local-linear kernel regression, whereby we maximize the validated R^2 for the choice of the best model and bandwidth. We further propose a 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 semi-parametric alternative. After we have seen the usefulness of the nonparametric approach, we introduce a method to incorporate prior knowledge in the estimation procedure. Examples are parametric regression or appropriate economic models. We indicate, that even the inclusion of the latter in a semi-parametric 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 provided by Robert Shiller and used for several other articles (see, for example, Campbell and Thompson (2008) or Chen et al. (2012)). Our results conform to economic theory, namely that the most important part of stock returns is related to the change in interest rates and earnings.

In order to deliver a statistical insight into our method, we mention that, mainly in higher dimensions, a nonparametric approach would suffer from the curse of dimensionality and/or bandwidth or boundary problems. A possible adjustment to counteract this problem would be to impose more structure. Our method contributes to this strategy due to its new and innovative idea – a model directly guided by economic theory. By means of a simple transformation, we achieve the combination of bias and dimension reduction (i. e. more structure to circumvent the curse of dimensionality), which means that, in our case, a reliable prior captures some of the characteristics of the shape of the estimating function; and as such, a multiplicative correction can cause a bias and dimension reduction in the remaining nonparametric estimation process of the correction factor. Thus, we propose here a method that greatly improves nonparametric regression in combination with a parametric technique.

Another way to impose more structure in the prediction process of excess stock returns could be the use of same-year 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 the same years' – instead of the 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. For example, Scholz et al. (2014) propose a two-step approach for the inclusion of the same years' bond yield, which is related to the change in interest rates. Furthermore, calendar effects or structural breaks, as described for linear models by Paye and Timmermann (2006) should be taken into consideration. While many financial prediction models are intended for the short term, the one-year view taken in this paper is natural and omnipresent in actuarial science. Clearly, pension savings is for the long-term, while the forecasting of balance sheet numbers like reserves in non-life insurance are mostly for one-year (see, for example, Kuang et al. (2011)). We believe that the long-term view provided by this paper could play an important role in building financial prediction models to be used in actuarial saving models in the future.

Bandi, F., Corradi, V., and Wilhelm, D. (2011). Data-Driven Bandwidth Selection for Nonparametric Nonstationary Regressions. Mimeo.

Bandi, F. and Perron, B. (2008). Long-run risk-return trade-offs. Journal of Econometrics 143, 349– 374.

Bikker, J., Broeders, D., Hollanders, D., and Ponds, E., 2012, Pension Funds Asset Allocation and Participant Age: A Test of the Life-Cycle Model, *Journal of Risk and Insurance* 79(3), 595–618.

Bollerslev, T., Tauchen, G., , and Zhou, H., 2009, Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* 22, 4463–4492.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M., 2007, On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing, *Journal of Finance* 62(2), 877–915.

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

Burman, P., Chow, E., and Nolan, D., 1994, A Cross-Validatory Method for Dependent Data, *Biometrika* 81, 351–358.

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

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

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

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

Chen, L., Da, Z., and Priestly, P., 2012, Dividend Smoothing and Predictability, *Management Science* 58(10), 1834–1853.

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

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

Eling, M., 2014, Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models?, *Insurance: Mathematics and Economics* 59, 45–56.

Elliott, G., Gargano A., and Timmermann, A., 2013, Complete Subset Regressions, *Journal of Econometrics* 177, 357–373.

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

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

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

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

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

Gerrard, R., Guillen, M., Nielsen, J. P., and Perez-Marin, A. (2014). Long-run savings and investment strategy optimization. *Scientific World Journal* Article ID 510531.

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

Gonzales-Manteiga, W. and Crujeiras, R., 2013, An updated review of Goodness-of-Fit tests for regression models, *Test* 22, 361–411.

Goyal, A., and Santa-Clara, P., 2003, Idiosyncratic Risk Matters! *Journal of Finance* 58, 975–1007.

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

Guerre, E., 2004, Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series, SSRN-working paper.

Guillen, M., Jarner, S., Nielsen, J. P., and Perez-Marin, A. (2014). Risk-Adjusted Impact of Administrative Costs on the Distribution of Terminal Wealth for Long-Term Investment. *Scientific World Journal* Volume 2014 (2014), Article ID 521074, 12 pages.

Guillen, M., Konicz, A., Nielsen, J. P., and Perez-Marin, A., 2013a, Do not pay for a Danish interest guarantee. The law of the triple blow, *Annals of Actuarial Science* 7(2), 192–209.

Guillen, M., Perez-Marin, A., Nielsen, J. P., and Petersen, K., 2013b, Performance measurement of pension strategies: a case study of Danish life cycle products, *Scandinavian Actuarial Journal* 1, 49-68.

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

Heidenreich, N., Schindler, A., Sperlich, S., and Perez-Marin, A., 2013, Bandwidth Selection Methods for Kernel Density Estimation: a review of fully automatic selectors, *AStA - Advances in Statistical Analysis* 97, 403–433.

Karlsen, H. A., and Tjostheim, D., 2001, Nonparametric Estimation in Null Recurrent Time Series, *The Annals of Statistics* 29, 372–416.

Köhler, M., Schindler, A., and Sperlich, S., 2014, A Review and Comparison of Bandwidth Selection Methods for Kernel Regression, *International Statistical Review* 82, 243–274.

Kothari, S., and Shanken, J., 1997, Book-to-market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis, *Journal of Financial Economics* 44, 169–203.

Kuang, D., Nielsen, B., and Nielsen, J.P., 2011, Forecasting in an Extended Chain-Ladder-Type Model, *Journal of Risk and Insurance*, 78(2), 345–359.

Lamont, O., 1998, Earnings and Expected Returns, *Journal of Finance* 53, 1563–1587.

Lettau, M., and Ludvigson, S., 2001, Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 56, 815–849.

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

Neely, Ch., Rapach, D., Tu, J., and Zhou, G., 2014, Forecasting the Equity Risk Premium: The Role of Technical Indicators, *Management Science* 60(7), 1772–1791.

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

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

Owadally, I., Haberman, S., and Gómez Hernández, D., 2013, A Savings Plan With Targeted Contributions, *Journal of Risk and Insurance* 80(4), 975–1000.

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

Paye, B. S., and Timmermann, A., 2006, Instability of Return Prediction Models, *Journal of Empirical Finance* 13, 274–315.

Rapach, D. E., Strauss, J. K., and 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.

Racine, J., 2000, Consistent Cross-Validatory Model-Selection for Dependent Data: hv-Block Cross-Validation, *Journal of Econometrics* 99, 39–61.

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

Scholz, M., Sperlich, S., and Nielsen, J. P., 2014, Nonparametric Long Term Prediction of Stock Returns With Generated Bond Yields, Mimeo.

Shao, J., 1993, Linear Model Selection by Cross-Validation, *Journal of the American Statistical Association* 88, 486–495.

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

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

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

Wilhelm, D., 2014, Optimal Bandwidth Selection for Robust Generalized Method of Moments Estimation, Mimeo.

Appendix: Tables of Additional Results

Table 12: US market data (1872-2014).

	Max	Min	Mean	Sd
S&P Stock Price Index	1807.78	3.25	207.04	407.55
Dividend Accruing to Index	39.44	0.18	4.91	8.00
Earnings Accruing to Index	105.96	0.16	11.59	21.66
Stock Returns	0.44	-0.62	0.04	0.18
Dividend by Price	0.09	0.01	0.04	0.02
Earnings by Price	0.17	0.02	0.08	0.03
Short-term Interest Rate	17.63	0.13	4.61	2.85
Long-term Interest Rate	14.59	1.91	4.60	2.26
Inflation	0.21	-0.16	0.02	0.06
Bond	2.03	-4.13	-0.02	0.78

NOTE: We base our empirical study on the annual data from the file `chap26.xls` which was extended to the full period of interest using monthly data in `ie_data.xls` (both files are downloadable from `http://www.econ.yale.edu/~shiller/data.htm`) and short-term interest rate data from the Federal Reserve.

Table 13: Predictive power (in percent) of the two-dimensional linear model (6).

	S,d	S,r	S,L	S,inf	S,b	d,r	d,L	d,inf
R_V^2	-1.3	2.2	-2.3	-3.0	-1.7	4.1	-1.0	1.6
R_{adj}^2	0.7	4.3	-0.5	-0.7	-0.1	5.5	0.2	2.8

	d,b	r,L	r,inf	r,b	L,inf	L,b	inf,b
R_V^2	-0.1	8.8	2.6	2.9	-2.1	-1.1	-1.8
R_{adj}^2	0.9	9.6	4.3	4.1	-0.5	-0.4	-0.5

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 14: Predictive power (in percent) 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	-1.6	1.7	-2.3	-2.8	-2.1	3.8	-1.0	-1.7
p-value	0.464	0.116	0.660	0.688	0.590	0.020	0.422	0.526
τ	0.072	0.190	0.036	0.040	0.052	0.213	0.048	0.050
p-value	0.428	0.118	0.612	0.614	0.448	0.052	0.464	0.524

	d,b	r,L	r,inf	r,b	L,inf	L,b	inf,b
R_V^2	-0.6	10.0	2.1	3.5	-2.1	-0.8	-2.1
p-value	0.316	0.000	0.070	0.012	0.668	0.352	0.648
τ	0.094	0.345	0.175	0.180	0.020	0.040	0.039
p-value	0.248	0.004	0.108	0.034	0.690	0.450	0.584

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 15: Predictive power (in percent) for dimension reduction using identity (10).

	S	d	e	r	L	inf	b
S	-5.1	-2.7	4.2	0.7	-3.8	-4.2	-3.1
d	-1.6	-2.1	5.3	2.9	-1.6	-2.0	-0.2
e	8.5	8.4	9.1	16.5	10.9	11.2	11.7
r	0.6	2.5	11.5	0.9	-0.3	0.5	1.5
L	-3.1	-2.2	5.7	4.8	-2.9	-3.0	-1.7
inf	-4.2	-2.7	8.5	0.8	-3.9	-4.8	-3.3
b	-3.4	-1.8	7.2	2.4	-2.7	-3.4	-1.4

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 16: Predictive power (in percent) for dimension reduction using identity (10).

	e, S	e, d	e, r	e, L	e, inf	e, b
S	2.5	3.9	9.8	3.9	5.4	5.3
d	3.6	4.2	11.9	5.6	7.6	6.5
e	7.5	9.1	16.0	10.0	9.9	10.2
r	9.6	11.9	10.0	7.5	10.9	10.9
L	4.1	6.4	12.0	4.1	6.2	6.4
inf	7.0	9.6	11.6	7.2	7.6	8.2
b	5.6	7.4	12.8	6.8	7.5	5.6

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 17: Predictive power (in percent) for dimension reduction using identity (10).

	S	d	e	r	L	inf	b
e, S	4.2	2.0	4.9	12.6	6.8	7.0	8.4
e, d	10.3	11.6	10.6	17.9	13.1	13.1	13.1
e, r	10.9	9.0	12.1	11.9	10.3	14.0	13.8
e, L	9.8	12.8	14.4	20.9	13.0	13.2	11.5
e, inf	7.8	7.4	8.1	13.7	9.3	8.4	10.0
e, b	9.1	7.6	9.5	17.4	10.9	11.5	10.3

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.