Graz Economics Papers – GEP

GEP 2019–06

Machine Learning for Forecasting Excess Stock Returns The Five-Year-View

Ioannis Kyriakou, Parastoo Mousavi, Jens Perch Nielsen and Michael Scholz

August 2019

Department of Economics
Department of Public Economics
University of Graz

An electronic version of the paper may be downloaded
from the RePEc website: http://ideas.repec.org/s/grz/wpaper.html

MACHINE LEARNING FOR FORECASTING EXCESS STOCK RETURNS – THE FIVE-YEAR VIEW

A PREPRINT

Ioannis Kyriakou
Faculty of Actuarial Science and Insurance
Cass Business School
City, University of London
London, United Kingdom
ioannis.kyriakou@city.ac.uk

Parastoo Mousavi
Faculty of Actuarial Science and Insurance
Cass Business School
City, University of London
London, United Kingdom
parastoo.mousavi.2@cass.city.ac.uk

Jens Perch Nielsen
Faculty of Actuarial Science and Insurance
Cass Business School
City, University of London
London, United Kingdom
jens.nielsen.1@city.ac.uk

Michael Scholz*
Department of Economics
University of Graz
Graz, Austria
michael.scholz@uni-graz.at

July 31, 2019

ABSTRACT

In this paper, we apply machine learning to forecast stock returns in excess of different benchmarks, including the short-term interest rate, long-term interest rate, earnings-by-price ratio, and the inflation. In particular, we adopt and implement a fully nonparametric smoother with the covariates and the smoothing parameter chosen by cross-validation. We find that for both one-year and five-year returns, the term spread is, overall, the most powerful predictive variable for excess stock returns. Differently combined covariates can then achieve higher predictability for different forecast horizons. Nevertheless, the set of earnings-by-price and term spread predictors under the inflation benchmark strikes the right balance between the one-year and five-year horizon.

Keywords: benchmark; cross-validation; prediction; stock returns; long-term forecasts; overlapping returns; autocorrelation

JEL: C14; C53; C58; G17; G22

1 Introduction

Machine learning (ML) for predicting asset returns is one of the in vogue topics in empirical finance. It is often seen as "i) a diverse collection of high-dimensional models for statistical prediction, combined with ii) so-called 'regularization' methods for model selection and mitigation of overfit, and iii) efficient algorithms for searching among a vast number of potential model specifications" (Gu et al., 2018). Mostly one of the following methods is well suited to address the three challenges mentioned earlier: linear models for regression (including regularization via shrinkage methods with penalization, such as Ridge Regression, Lasso, or Elastic Nets), dimension reduction via principal components regression and partial least squares, regression trees and forests (including boosted trees and random forests), (deep) neural networks, and

*Correspondence: michael.scholz@uni-graz.at; Tel.: +43-316-380-7112

boosting[2]. Hastie et al. (2017) describe the prediction setup in a very general way:

$$\min_{f\in\mathcal{H}}\left\{L\big(y_{t+h},f(Z_t)\big)+p(f,\tau)\right\},\quad t=1,\dots,T, \tag{1}$$

where y_{t+h} is the variable to be predicted h periods ahead, Z_t the vector of predictors, $\mathcal{H}$ a space of possible functions f that combine the data to form the prediction, p a penalty on f, τ a set of hyper-parameters (for example, the λ in the Lasso), and L a loss function that defines the optimal forecast.

In contrast to the methods mentioned above, in this paper we use local-linear smoothing techniques in combination with a leave-k-out cross-validation for the following reasons. First, we take the long-term actuarial perspective and are interested in longer-horizon stock returns based on annual observations. Thus, we are not in the big-data context where the number of observations is huge and the set of possible predictive variable combinations is enormous (and, thus, dimension reduction or shrinkage are indispensable), and do not need a penalization in (1). Our data set is, instead, sparse and a careful imposition of structure to the statistical modelling process is much more promising, as shown, for example, by Nielsen and Sperlich (2003) and Scholz et al. (2015, 2016) . Second, the evidence of stock return predictability is much stronger once one allows for nonlinear functions f in (1) as documented, for example, in Lettau and Van Nieuwerburgh (2008), Chen and Hong (2010), or Cheng et al. (2019). Thus, the local-linear smoother based on the standard L_2-loss function is ideally suited as it can estimate a linear function – the classical benchmark in this context – without any bias. Finally, our procedures are analytically well studied, i.e., sound and rigorous, statistical tools which let us operate in a glass house, not in a black box – in contrast to other fancier but less clear ML methods.

The remaining of this paper is organized as follows. In Section 2, we present our framework for the purpose of long-term predictions. We define the underlying financial model, describe the adopted local-linear smoother with its theoretical properties, and present our validation criterion for the model selection. In Section 3, we provide a description of our data set and a preceding descriptive analysis. Subsequently, we exhibit our empirical findings from different validated scenarios: i) a single benchmarking approach with the dependent variable measured on the original nominal scale; and ii) the case where both the independent and dependent variables are adjusted according to the benchmark (full benchmarking approach). Finally, we take the long-term view and comment on real income pension prediction. Section 4 summarizes the key points of our analysis and concludes the paper.

2 A framework for long-term prediction

In this section, we focus on nonlinear predictive relationships between stock returns over the next T years in excess of a reference rate or benchmark and a set of explanatory variables. We aim to investigate different benchmark models and their predictability over horizons of one year and five years. We consider four different benchmarks: the short-term interest rate, the long-term interest rate, the earnings-by-price ratio, and the inflation.

2.1 One-year predictions

We investigate stock returns $S_t=(P_t+D_t)/P_{t-1}$, where D_t denotes the (nominal) dividends paid during year t and P_t the (nominal) stock price at the end of year t, in excess (log-scale) of a given benchmark $B_{t-1}^{(A)}$:

$$Y_t^{(A)}=\ln\frac{S_t}{B_{t-1}^{(A)}}, \tag{2}$$

where $A\in\{R,L,E,C\}$ with, respectively,

$$B_t^{(R)}=1+\frac{R_t}{100},\quad B_t^{(L)}=1+\frac{L_t}{100},\quad B_t^{(E)}=1+\frac{E_t}{P_t},\quad B_t^{(C)}=\frac{CPI_t}{CPI_{t-1}},$$

R_t is the short-term interest rate, L_t the long-term interest rate, E_t the earnings accruing to the index in year t, and CPI_t the consumer price index for year t. The predictive nonparametric regression model for a one-year horizon is then given by

$$Y_t^{(A)}=m(X_{t-1})+\xi_t, \tag{3}$$

[2] For an overview, see, for example, Athey and Imbens (2019) or Coulombe et al. (2019).

where

$$m(x) = \mathbb{E}(Y^{(A)}|X = x),\ x \in \mathbb{R}^q, \tag{4}$$

is an unknown smooth function and ξ_t is a martingale difference process, i.e., serially uncorrelated zero-mean random error terms, given the past, of an unknown conditionally heteroscedastic form $\sigma(x)$.

Our aim is to forecast the excess stock returns $Y_t^{(A)}$ using popular lagged predictive variables X_{t-1} including the: i) dividend-by-price ratio $d_{t-1} = D_{t-1}/P_{t-1}$; ii) earnings-by-price ratio $e_{t-1} = E_{t-1}/P_{t-1}$; iii) short-term interest rate $r_{t-1} = R_{t-1}/100$; iv) long-term interest rate $l_{t-1} = L_{t-1}/100$; v) inflation $\pi_{t-1} = (CPI_{t-1} - CPI_{t-2})/CPI_{t-2}$; vi) term spread $s_{t-1} = l_{t-1} - r_{t-1}$; and vii) excess stock return $Y_{t-1}^{(A)}$.

2.2 Longer-horizon predictions

Longer horizons are of fundamental interest to long-term investors, such as pension funds or market participants saving for distant payoffs. Note that long-term investors are generally willing to take on more risk for higher rewards. It is also shown, for example, in Rapach and Zhou (2013) that longer horizons tend to produce better estimates than shorter horizons. Furthermore, describing the Danish policy-making in the communication of pension forecasts for individuals based on longer-term predictions, Munk and Rangvid (2018) point out that major finance houses today use longer horizons – up to ten years – to stabilise and improve future predictions. In our paper, we concentrate on the five-year view as a first compromise for a comparison between the long five-year horizon and the shorter one-year horizon. What seems clear, however, is that shorter horizons based on monthly, weekly or even daily data do not seem to provide the pension saver good information nor his or hers future income as a pensioner, therefore these type of short-term predictions – sometimes called investment robots – are not suitable when a pensioner should define his or hers risk appetite.

More in detail, for longer horizons T we consider the sum of annual continuously compounded returns:

$$Z_t^{(A)} = \sum_{i=0}^{T-1} Y_{t+i}^{(A)}.$$

Note that we use here overlapping returns $Z_t^{(A)}$, which require a careful econometric modelling. For illustrative purposes, assume a linear relationship in (3) between $Y_t^{(A)}$ and X_{t-1}, as well as persistence of the forecasting variable (treating the variables as deviations from their means):

$$Y_t^{(A)} = \beta X_{t-1} + \xi_t \quad \text{and} \quad X_t = \gamma X_{t-1} + \eta_t,$$

with ξ_t as in (3) and η_t being white noise. The T-year regression problem that is implied by this pair of one-year regressions is now

$$\begin{aligned} Z_t^{(A)} &= Y_t^{(A)} + \ldots + Y_{t+T-1}^{(A)} = (\beta X_{t-1} + \xi_t) + \ldots + (\beta X_{t+T-2} + \xi_{t+T-1}) \\ &= \beta \sum_{i=0}^{T-1} \gamma^i X_{t-1} + \beta \sum_{i=0}^{T-1} \sum_{j=0}^{T-1-i} \gamma^j \eta_{t+i} + \sum_{i=0}^{T-1} \xi_{t+i} = \phi X_{t-1} + \nu_t, \end{aligned}$$

i.e., the excess stock return for year t over the next T years can be decomposed in a predictive part depending on the variable X_{t-1} and an unpredictable error term ν_t. To avoid functional misspecification due to our simplistic assumption, we allow for nonlinearity and set up our predictive nonparametric regression model in the same fashion as in (3)

$$Z_t^{(A)} = m(X_{t-1}) + \nu_t, \tag{5}$$

where

$$m(x) = \mathbb{E}(Z_t^{(A)}|X = x),\ x \in \mathbb{R}^q, \tag{6}$$

is an unknown smooth function. The important difference between (3) and (5) is now that ξ_t is a martingale difference process but ν_t will be serially correlated by construction. The predictive variables X under consideration are the dividend-by-price ratio d, earnings-by-price ratio e, short-term interest rate r, long-term interest rate l, inflation π, term spread s, and the one-year excess stock return $Y^{(A)}$.

In the next section, we address the problem of estimating the conditional mean function (4) and the consequences of the serial correlation of the error terms ν_t for the estimation of m in (5). We present consistency results and asymptotic normality for the local-linear (LL) smoother which we then implement in Section 3.

2.3 The local-linear smoother

In the description of the statistical properties of our preferred local-linear smoother used for estimation of the conditional mean function, we follow closely Kyriakou et al. (2019). Consider a sample of real random variables $\{(X_t, Y_t), t = 1, \dots, n\}$ which are strictly stationary and weakly dependent. To measure the strength of dependence in the time series, we limit to the *strong*- or *α-mixing* (e.g., see Doukhan, 1994), where

$$\alpha_\tau = \sup_{t\in\mathbb{N}} \sup_{A\in\mathcal{F}_{t+\tau}^{\infty}, B\in\mathcal{F}_{-\infty}^{t}} |P(A\cap B) - P(A)P(B)|,$$

$\mathcal{F}_i^j$ denotes the σ-algebra generated by $\{X_k, i \le k \le j\}$. In addition, α_τ approaches zero as $\tau \to \infty$. Note that weakly dependent data rule out, for example, processes with long-range dependence and nonstationary processes with unit roots. We further assume that the sequence $\{(X_t, Y_t), t = 1, \dots, n\}$ is *algebraic α-mixing*, i.e., $\alpha = O(\tau^{-(1+\epsilon)})$ for some $\epsilon > 0$.

Consider now the prediction problem (3)–(4) and the Nadaraya–Watson (NW) estimator (local-constant kernel method) for $m(x)$ given by

$$\hat{m}_{NW}(x) = \frac{\hat{p}(x)}{\hat{f}(x)}, \tag{7}$$

where the probability density function of X_t, $f(x)$, is estimated for a given fixed value of $x = (x_1, \dots, x_q)' \in \mathbb{R}^q$ by

$$\hat{f}(x) = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x)$$

and

$$\hat{p}(x) = \frac{1}{n}\sum_{t=1}^{n} Y_t K_h(X_t - x).$$

K_h denotes some kernel function, for example, the product kernel

$$K_h(X_t - x) = \prod_{s=1}^{q} \frac{1}{h_s} k\left(\frac{X_{ts} - x_s}{h_s}\right),$$

which depends on a set of bandwidths $(h_1, \dots, h_q)$ and higher-order kernels k (the order $\nu > 0$ of the kernel is defined as the order of the first nonzero moment), i.e., univariate symmetric functions satisfying $\int k(u)du = 1$, $\int u^l k(u)du = 0$ $(l = 1, \dots, \nu - 1)$, and $\int u^\nu k(u)du =: \kappa_\nu > 0$. X_{ts} denotes the sth component of X_t $(s = 1, \dots, q)$.

Under the standard assumptions of serial dependence with a required rate α as stated earlier, bounded density $f(x)$, controlled tail behaviour of conditional expectations, $h_s \to 0$ $(s = 1, \dots, q)$ and $nH_q = nh_1 \dots h_q \to \infty$ as $n \to \infty$, the following result of pointwise convergence holds (see Li and Racine, 2007).

Theorem 1. *Under the given assumptions,*

$$|\hat{m}_{NW}(x) - m(x)| = O_p\left(\sum_{s=1}^{q} h_s^2 + \frac{1}{\sqrt{nH_q}}\right).$$

Several generalizations of Theorem 1 have been proposed in the literature, including proof of uniform and almost sure convergence of the NW estimator (Hansen, 2006) and quasi-complete convergence of the estimator in the case of generated regressors and weakly dependent data (Scholz et al., 2016). Li and Racine (2007) further show the asymptotic normality of the estimator by calculating the bias term $B_s(x) = \frac{\kappa_2}{2}\left(f(x)m_{ss}(x) + 2f_s(x)m_s(x)\right)/f(x)$, where subscripts s and ss denote, respectively, first and second order derivatives, and $\kappa_2 = \int u^2 k(u)du$.

Theorem 2. *Under the given assumptions,*

$$\sqrt{nH_q}\left(\hat{m}_{NW}(x) - m(x) - \sum_{s=1}^{q} h_s^2 B_s(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa^q \sigma^2(x)}{f(x)}\right),$$

where $\kappa = \int k^2(u)du$.

The extension to the LL estimator $\hat{m}_{LL}(x)$ is almost straightforward. For notational convenience, we focus on the case $q = 1$. Then, upon defining

$$\begin{aligned} s_j(x) &= \sum_{t=1}^{n} K_h(X_t - x)(X_t - x)^j, \\ t_j(x) &= \sum_{t=1}^{n} Y_t K_h(X_t - x)(X_t - x)^j \end{aligned}$$

for $j = 0, 1, 2$, we get

$$\hat{m}_{LL}(x) := \frac{t_0(x)s_2(x) - t_1(x)s_1(x)}{s_0(x)s_2(x) - s_1^2(x)} = \frac{\sum_{t=1}^{n} Y_t C_h(X_t - x)}{\sum_{t=1}^{n} C_h(X_t - x)} \tag{8}$$

with the kernel function

$$C_h(X_t - x) = \frac{1}{nh} \sum_{s \neq t} K_h(X_t - x)(X_s - X_t)K_h(X_s - x)(X_s - x)$$

representing a discretized version of $C(u) := \int K(u)(v-u)K(v)v dv$. Note that (8) is of the same form as (7) and that the kernel C has similar properties to K. Applying Theorem 1 yields the pointwise convergence result for the LL estimator.

Theorem 3. *Under the given assumptions,*

$$|\hat{m}_{LL}(x) - m(x)| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right).$$

For a multivariate extension and asymptotic normality, refer, for example, to Masry (1996).

Consider now the prediction problem (5)–(6). As mentioned before, the important difference is the inherited serial correlation of the error terms ν_t in (5). It is well documented in the statistical literature that, in the presence of correlated errors, quite fundamental problems occur: i) while the consistency result from Theorem 3 still holds, the left-out information of the error dependency leads to less efficient estimators (Xiao et al., 2003; Su and Ullah, 2006); and ii) the commonly applied automatic smoothing parameter selection procedures, like cross-validation or plug-in, break down (De Brabanter et al., 2011; Bergmeir et al., 2018). The latter problem will be discussed in detail in the next section.

More efficient estimators are proposed in the literature. For example, Xiao et al. (2003) use a pre-whitening transformation of the dependent variable that has to be estimated from the data. More in detail, the residual process ν_t is assumed to be stationary, zero-mean with variance σ_ν^2, and has an invertible linear process representation

$$\nu_t := \sum_{i=0}^{\infty} c_i \varepsilon_{t-i},$$

where ε_t are i.i.d. with zero mean and variance σ_ε^2. Define $c(L) = \sum_{i=0}^{\infty} c_i L^i$ with the usual lag operator L. By inverting $c(L)$ one gets an autoregressive representation of ν_t of infinite order:

$$c(L)^{-1} = a(L) = a_0 - \sum_{i=1}^{\infty} a_i L^i$$

and thus $a(L)\nu_t = \varepsilon_t$. The transformed regression problem (5) is then

$$\begin{aligned} a(L)Z_t^{(A)} &= a(L)m(X_{t-1}) + \varepsilon_t, \quad \text{or} \\ \tilde{Z}_t^{(A)} := Z_t^{(A)} - \sum_{i=1}^{\infty} a_i\big(Z_{t-i}^{(A)} - m(X_{t-1-i})\big) &= m(X_{t-1}) + \varepsilon_t \end{aligned} \tag{9}$$

with an uncorrelated error term ε_t. Denoting by $\tilde{m}$ the NW estimator of the transformed regression problem (9), Xiao et al. (2003) established the asymptotic distribution presented next.

Theorem 4. *Under the given assumptions,*

$$\sqrt{nH_q}\left(\tilde{m}_{NW}(x) - m(x) - \sum_{s=1}^{q} h_s^2 B_s(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa^q \sigma_\varepsilon^2(x)}{f(x)}\right),$$

where $\kappa = \int k^2(u)du$.

Note that $\tilde{m}_{NW}$ is more efficient than $\hat{m}_{NW}$ because $\sigma_\varepsilon^2 < \sigma_v^2 := \sigma_\varepsilon^2 \sum_{i=0}^{\infty} c_i^2$. In practice, the transformed dependent variable $\tilde{Z}_t^{(A)}$ is unknown, thus $\tilde{m}_{NW}$ is infeasible. Xiao et al. (2003) propose to replace $\tilde{Z}_t^{(A)}$ by an approximation based on estimates of the coefficients $\{a_i\}$ and a truncation of the infinite sum. They also show that the resulting feasible estimator has the same asymptotic distribution as described in Theorem 4.

Other contributions which provide more efficient local-polynomial estimators under similar settings can be found in Su and Ullah (2006), Linton and Mammen (2008), or more recently Geller and Neumann (2018) (and citations therein).

We do not apply such techniques in our paper for several reasons. First, additional parameters $\{a_i\}$ have to be estimated and the infinite sum must be truncated at a meaningful value in (9), or the residual process has to be adequately modelled by some parametric ARMA process or even nonparametrically, where the appropriate lag-length has to be specified. Second, most examples and simulations given in the literature are one-dimensional. But, we adapt our local-linear smoother in a multidimensional problem and it is not clear what the efficiency gain would be in our scarce data environment. Finally, in a recent study, Scholz et al. (2015) show that, in a long-term set up with annual data, the reduction of the prediction bias is crucial as it contributes squared to the prediction mean square error. Our approach of imposing economic structure as discussed in Section 3.4 aims in a similar direction.

Thus, we think that the more severe problem caused by autocorrelation is the misleading smoothing parameter selection for methods like cross-validation or plug-in. These will be discussed in detail in the next section.

2.4 The principal of validation: model selection and smoothing parameter choice

As we use a nonparametric technique, we require an adequate measure of the predictive power, hence classical in-sample measures, such as the R^2 or adjusted R^2, are not appropriate; in prediction, we are not interested in how well a model explains the variation inside the sample but, instead, in its out-of-sample performance. Therefore, we aim to estimate the prediction error directly.

For the purpose of model as well as optimal bandwidth selection, we use the validated R_V^2 of Nielsen and Sperlich (2003) based on a leave-k-out cross-validation. This method of finding the smoothing parameter has shown to be suitable also in a time series context. For example, Bergmeir et al. (2018) show that, in the case of uncorrelated errors, cross-validation is preferred to out-of-sample evaluation where a section from the end of the series is withheld for evaluation. For the latter only one evaluation on a test data set is possible, whereas cross-validation performs various evaluations, which is, especially in small data sets such as ours, beneficial.

Our validation criterion is defined as

$$R_V^2 = 1 - \frac{\sum\limits_t (Y_t^{(A)} - \hat{m}_{-t})^2}{\sum\limits_t (Y_t^{(A)} - \bar{Y}_{-t}^{(A)})^2}, \tag{10}$$

where leave-k-out estimators are used: $\hat{m}_{-t}$ for the nonparametric function m and $\bar{Y}_{-t}^{(A)}$ for the unconditional mean of $Y_t^{(A)}$. Both are computed by removing k observations around the tth time point. Here we use $k = 2T - 1$ due to the construction of the dependent variable over a horizon of T years, i.e., for the one-year horizon the classical leave-one-out estimator, while, for example, for the five-year horizon the leave-nine-out estimator. R_V^2 measures the predictive power of a given model compared to the cross-validated historical mean; positive R_V^2 implies that the predictor-based regression model (3) or (5) outperforms the corresponding historical average excess stock return over T years.

It is well known that cross-validation often requires omission of more than one data point and, possibly, additional correction when the omitted fraction of data is considerable (see, for example, Burman et al.,

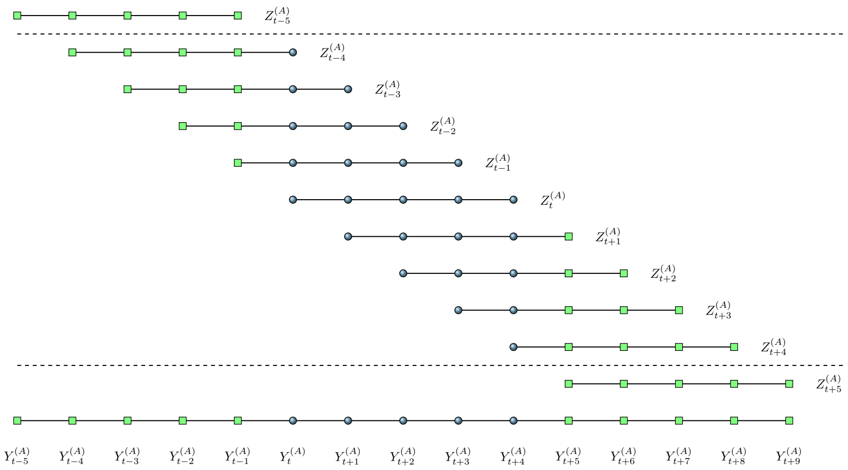

Figure 1: Illustration of the leave-nine-out set (between the dashed lines) of $Z_{t-4}^{(A)}, \ldots, Z_{t+4}^{(A)}$ which include at least one element of $Y_{t}^{(A)}, \ldots, Y_{t+4}^{(A)}$ (see bottom).

1994). Also, when serial correlation arises, as in our longer-horizon application, and the structure of the error terms is ignored, De Brabanter et al. (2011) show that automatic tuning parameter selection methods, such as cross-validation or plug-in, fail. The problem is that the chosen bandwidths become smaller for increasing correlations (Opsomer et al., 2001), and the corresponding model fits become progressively more undersmoothed. The bias of the predictor reduces this way and, as it contributes in a squared fashion to the prediction mean squared error – the numerator of the ratio in (10), R_V^2 increases (not because the fit is good but due to the ignored correlation structure). A misleading decision on the bandwidth or model specification, as well the set of preferred covariates is the consequence. To overcome those problems, Chu and Marron (1991) propose the use of bimodal kernel functions. Such functions are known to remove the correlation structure very effectively but the estimator $\hat{m}$ suffers from increased mean squared error, as discussed in De Brabanter et al. (2011). They also propose a correlation-corrected cross-validation which consists of, first, finding the amount of data k to be left out in the estimation process when a bimodal kernel function is used; and, second, applying the actual choice of the smoothing parameter using leave-k-out cross-validation with a unimodal kernel function. In our application, we can skip the first step because k is known by construction. For example, in the five-year case, we have $Z_t^{(A)} = Y_t^{(A)} + \ldots + Y_{t+4}^{(A)}$. Now we want to exclude the complete information included at time t, i.e., skip all $Z_s^{(A)}$ that include any of $Y_t^{(A)}, \ldots, Y_{t+4}^{(A)}$; it is easy to see from Figure 1 that this amounts to a leave-nine-out set of $Z_{t-4}^{(A)}, \ldots, Z_{t+4}^{(A)}$.

3 Empirical application: prediction of stock returns in excess of different benchmarks

3.1 The data

In this paper, we take the long-term actuarial view and base our predictions on annual US data provided by Robert Shiller. This data set, which is made available from `http://www.econ.yale.edu/~shiller/data.htm`, includes, among other variables, long-term changes of the Standard and Poor's (S&P) Composite Stock Price Index, bond price changes, consumer price index changes, and interest rate data from 1872 to 2019. This is an updated and revised version of Shiller (1989, Chapter 26), which provides a detailed description of the data. Note that the risk-free rate in this data set (based on the 6-month commercial paper rate until 1997 and afterwards on the 6-month certificate of deposit rate, secondary market) was discontinued in 2013. We follow the strategy of Welch and Goyal (2008) and replace it by an annual yield which is based on the 6-month Treasury-bill rate, secondary market, from `https://fred.stlouisfed.org/series/TB6MS`. This new series is only available from 1958 to 2019. In the absence of information prior to 1958, we had to estimate

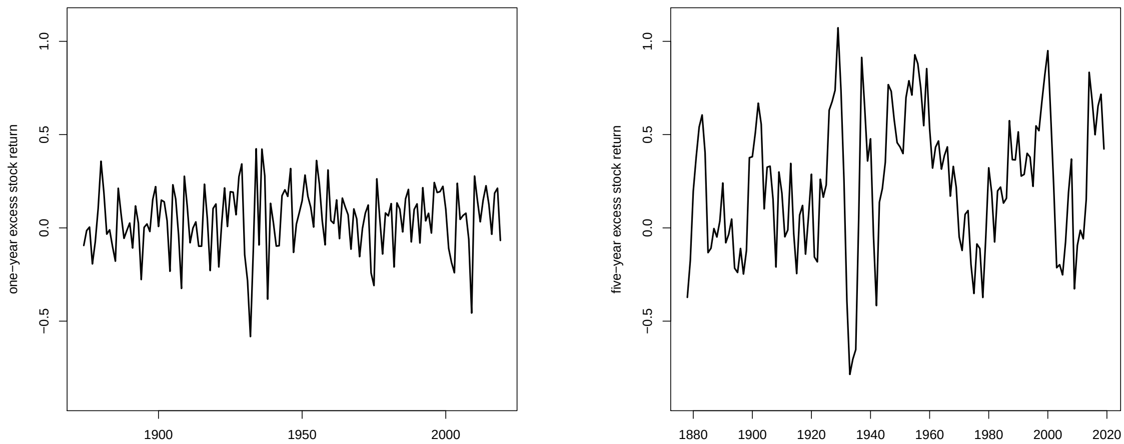

Figure 2: Left: one-year stock returns in excess of the risk-free benchmark. Right: five-year stock returns in excess of the risk-free benchmark. Period: 1872–2019. Data: annual S&P 500.

it. To this end, we regressed the Treasury-bill rate on the risk-free rate from Shiller's data for the overlapping period 1958 to 2013, which yielded

$$\text{Treasury-bill rate} = 0.0961 + 0.8648 \times \text{commercial paper rate}$$

with an R^2 of 98.6%. Therefore, we instrumented the risk-free rate from 1872 to 1957 with the predicted regression equation. The correlation between the actual Treasury-bill rate and the predictions for the estimation period is 99.3%.

In this paper, as we are interested in actuarial models of long-term savings, which also explains why the validation methodology we adopt originates from the actuarial literature (see Nielsen and Sperlich, 2003), we use yearly data and predict at a one-year and a five-year horizon. Clearly, we do not have many historical years and data sparsity can be an important issue in our approach. However, it is worth highlighting that prediction might vary for higher-frequency data and a good model, for example, for monthly data might not be for yearly data, and vice versa.

3.2 Descriptive analysis

As Fama and French (2018) point out, there is much research on predictability of returns, and a lot is known about the characteristics of short-horizon stock returns. For example, stylized facts about daily and monthly returns include excess kurtosis, distributions which are not normal, and volatility clustering. But less is known about distributions of long-horizon returns. However, such characteristics are of central interest to investors saving for distant payoffs.

Figure 2 shows the time series of the one-year returns (left) and five-year returns (right), both in excess of the risk-free benchmark, which are displayed on the same scale for the sake of comparison. The five-year series exhibits larger positive returns, which is not surprising as a longer period under risk should be paid off with a higher risk premium. Also the autoregressive structure of the five-year returns can be easily seen in comparison to the assumed weak dependence of one-year returns.

Figure 3 shows histograms of the one-year returns (left) and five-year returns (right) together with a kernel density estimate (red) and a fitted normal distribution (green). One notes again that the distribution of the five-year returns is shifted to the right but has a higher volatility. A Jarque–Bera test of the hypothesis of normality does reject for one-year returns (p-value = 0.013) but does not reject for five-year returns (p-value = 0.522). Furthermore, the density estimate for the five-year returns indicates more a mixture of normals than a single normal distribution which represents some evidence of a possible structural break in the data-generating process. Including structural changes in the modelling process is important. Higher-frequency

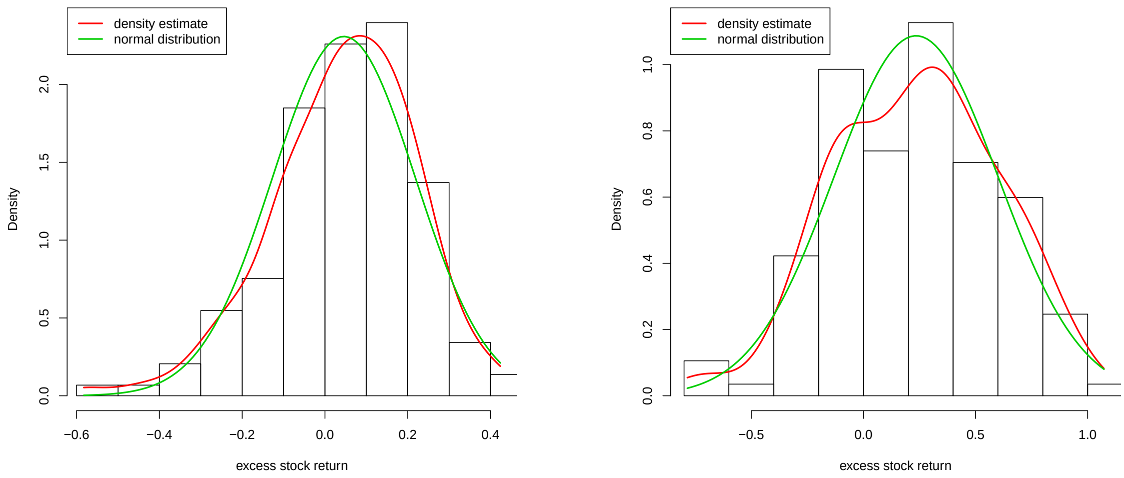

Figure 3: Histogram, kernel density estimate (red), and fitted normal distribution (green). Left: one-year stock returns in excess of the risk-free benchmark. Right: five-year stock returns in excess of the risk-free benchmark. Period: 1872–2019. Data: annual S&P 500.

returns with a certain number of observations allow for this. Rapach and Wohar (2004) find significant evidence of structural breaks in seven out of eight predictive regressions of S&P 500 returns, and three out of eight in CRSP (Center for Research in Security Prices) equal-weighted returns. Pesaran and Timmermann (2002) find that a linear predictive model that incorporates structural breaks improves the out-of-sample statistical forecasting power. We do not consider a structural break here as it is not clear at which point in time we should incorporate it (after the Great Recession, the second World War, the Global Financial Crisis, or during the Bretton Woods agreement). Also, accounting for this via sample split would result in an even smaller data set. From a statistical perspective, as we apply a fully nonparametric method, this would lead to mostly one-dimensional models because more observations are required to produce consistent estimates; the more information, the better to understand the underlying data-generating process.

This section is concluded with Table 1 which displays standard descriptive statistics for one-year and five-year returns as well as the available covariates. Both the one-year and five-year excess returns have a negative skewness, i.e., the left tail of the distribution (large negative returns) is longer or fatter than the right tail (large positive returns). Note that this is more pronounced in the case of one-year rather than five-year returns. While one-year returns are leptokurtic (positive excess kurtosis of 0.68), five-year returns exhibit a small negative excess kurtosis of -0.37.

Similar plots to those in Figures 2–3 and information as in Table 1 for the other benchmarks are available upon request by the authors. In the next sections, we analyze the predictability of one-year and five-year stock returns in excess of the different benchmarks.

3.3 Single benchmarking approach

In this section, we consider a single benchmarking approach as in Kyriakou et al. (2019) where only the variable S_t is adjusted according to some benchmark $B_{t-1}^{(A)}$, as shown in (2), while the independent variable(s) is (are) measured on the original nominal scale. The models (3) and (5) are estimated with a local-linear kernel smoother using the quartic kernel and the optimal bandwidth is chosen by cross-validation, i.e., by maximizing the R_V^2 given by (10). 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.

We study the empirical findings of R_V^2 values based on different validated scenarios shown for the one-year horizon in Table 2 and the five-year horizon in Table 3.

Table 1: US market data (1872–2019).

	Max	Min	Mean	Sd	Skew	Exc. kurt
S&P stock price index	2789.80	3.25	277.58	558.13	2.43	5.50
Dividend accruing to index	53.75	0.18	6.04	10.56	2.45	6.00
Earnings accruing to index	132.39	0.16	13.96	26.31	2.43	5.35
One-year excess stock returns $Y^{(R)}$	42.39	-58.26	4.58	17.28	-0.57	0.68
Five-year excess stock returns $Z^{(R)}$	107.27	-78.54	23.49	36.69	-0.14	-0.37
Dividend-by-price	9.88	1.17	4.31	1.71	0.46	0.25
Earnings-by-price	17.75	1.72	7.28	2.75	1.05	1.39
Short-term interest rate	14.93	0.07	3.97	2.50	0.96	2.34
Long-term interest rate	14.59	1.88	4.53	2.27	1.81	3.63
Inflation	20.69	-15.65	2.23	5.96	0.26	1.60
Spread	3.64	-3.71	0.56	1.32	-0.05	0.02

We find that in the case of the five-year returns, which is the focal point of this paper, the term spread s is, overall, the most powerful predictive variable for excess stock returns; this superior performance is also observed in the one-year case. More in details, with the prediction constrained to using only one-dimensional covariates, the term spread is the best predictor for the one-year[3] and five-year horizon under the short interest benchmark $B^{(R)}$ with, respectively, $R_V^2 = 9.7\%$ and 15.5%, but this also does quite well in the one-year case under the long rate and earnings-by-price benchmarks, $B^{(L)}$ and $B^{(E)}$, with $R_V^2 \in \{6.2\%, 7.5\%\}$ (for $B^{(C)}$ the best is π with 10.3%). In the five-year case under $B^{(L)}$ and $B^{(E)}$, the term spread s yields a high $R_V^2 \in \{8.0\%, 11.5\%\}$ whereas under $B^{(C)}$ the dividend-by-price ratio d gains ground with $R_V^2 = 7.6\%$.

In light of these remarks, we therefore focus the spotlight on the relationship between the spread and the excess stock returns. We present in the top panel of Figure 4 the estimated functions $\hat{m}$ (red solid line) for the one-year horizon (left) and the five-year horizon (right) under the single risk-free benchmark together with a corresponding linear model (dash-dotted green line), and a 45-degree line (dashed black line).[4] Both conform to the fact that an increase in the spread corresponds to an increase in the excess stock return. While a positive spread corresponds to a positive return for the one-year case, a spread larger than -1% gives on average a positive five-year return. This finding is in line with, for example, Resnick and Shoesmith (2002) who find that the value of the yield spread holds important information about the probability of a bear stock market. Regarding our validation procedure, Figure 4 also confirms that our approach correcting for autocorrelation in the five-year prediction problem was successful. The estimated functions are quite smooth indicating that the chosen bandwidth is not too small and that the resulting fit and validated R^2 are reasonable.

Back to our discussion of the results in Tables 2–3, in broad terms, five-year predictability improves over one-year: 67 out of 112 models achieve a larger R_V^2, and we observe 64 (five-year) versus 52 (one-year) models with nonnegative R_V^2, i.e., our proposed predictor-based regression model for the longer forecast horizon in this application surpasses the historical average excess return in the majority of cases. In addition, combining the term spread s with the dividend-by-price d results in uplifted predictability to 26.2%; this combination is, in fact, the best-performing one for 3 out of 4 benchmarks ($B^{(R)}, B^{(L)}, B^{(C)}$). In particular, imposing an additional covariate to s results in one-year R_V^2 in the range 6–10% under $B^{(R)}$; under other benchmarks, such as $B^{(L)}$, the one-year R_V^2 is in the range 3–6.5% (approx.); changing to the $B^{(E)}$ benchmark results in R_V^2 in the range 4.5–7.5%. Interestingly, for a five-year horizon, we observe a substantially improved predictive power with our cross-validated R_V^2 ranking some two-dimensional better than one-dimensional models, in fact,

[3] In the whole discussion that follows, note that the one-year predictions might differ from those originally reported in Kyriakou et al. (2019) due the updated data set and the replacement of the commercial paper rate by the Treasury-bill rate; nevertheless, the models remain similarly ranked.

[4] We present in Figure 4 the three single covariates with the largest R_V^2 ($T = 1$ and $T = 5$) for the single risk-free benchmark case: term spread (9.7% and 15.5%), short-term interest rate (3.0% and 7.8%), and long-term interest rate (0.0% and 1.4%).

Table 2: Predictive power for one-year excess stock returns $Y_t^{(A)}$: the single benchmarking approach. The prediction problem is defined in (3). The predictive power (%) is measured by R_V^2 as defined in (10). The benchmarks $B^{(A)}$ considered are based on the short-term interest rate ($A \equiv R$), long-term interest rate ($A \equiv L$), earnings-by-price ratio ($A \equiv E$), and consumer price index ($A \equiv C$). The predictive variables used are X_{t-1}, given by the dividend-by-price ratio d_{t-1}, earnings-by-price ratio e_{t-1}, short-term interest rate r_{t-1}, long-term interest rate l_{t-1}, inflation π_{t-1}, term spread s_{t-1}, excess stock return $Y_{t-1}^{(A)}$, or the possible different pairwise combinations as indicated.

Benchmark $B^{(A)}$	Explanatory variable(s) X_{t-1}						
	$Y^{(A)}$	d	e	r	l	π	s
Short-term rate	-1.6	-1.1	-0.6	3.0	0.0	-1.4	9.7
Long-term rate	-1.8	-0.8	-0.4	1.9	0.0	-1.4	6.2
Earnings-by-price	-1.7	-1.2	-1.4	0.0	-0.8	-1.2	7.5
Inflation	-1.4	-0.2	-1.5	0.8	-0.8	10.3	7.2
	$(Y^{(A)},d)$	$(Y^{(A)},e)$	$(Y^{(A)},r)$	$(Y^{(A)},l)$	$(Y^{(A)},\pi)$	$(Y^{(A)},s)$	
Short-term rate	-2.6	-2.4	0.9	-2.4	-2.9	6.3	
Long-term rate	-2.4	-2.3	-0.2	-2.4	-3.1	2.7	
Earnings-by-price	-3.5	-3.7	-2.0	-2.8	-2.8	4.5	
Inflation	-1.6	-3.4	-0.9	-2.5	9.7	4.8	
	(d,e)	(d,r)	(d,l)	(d,π)	(d,s)		
Short-term rate	-2.9	2.1	-1.6	-2.6	9.3		
Long-term rate	-2.7	1.3	-1.3	-2.3	5.8		
Earnings-by-price	-3.7	-1.4	-2.2	-2.4	6.0		
Inflation	-1.9	0.8	-1.2	9.5	7.9		
	(e,r)	(e,l)	(e,π)	(e,s)			
Short-term rate	4.0	-1.1	-1.6	9.1			
Long-term rate	3.2	-0.5	-1.3	5.5			
Earnings-by-price	-1.4	-2.3	-2.7	5.4			
Inflation	-0.4	-2.5	10.9	5.4			
	(r,l)	(r,π)	(r,s)				
Short-term rate	8.5	1.4	10.0				
Long-term rate	4.9	0.3	6.5				
Earnings-by-price	6.0	-1.5	7.2				
Inflation	5.2	9.5	7.4				
	(l,π)	(l,s)					
Short-term rate	-2.1	10.1					
Long-term rate	-2.0	6.6					
Earnings-by-price	-2.0	7.0					
Inflation	9.9	7.4					
	(π,s)						
Short-term rate	7.7						
Long-term rate	4.1						
Earnings-by-price	5.2						
Inflation	15.4						

more than for a one-year horizon: in particular, as possibly anticipated by the aforementioned performances of d and s, the two-dimensional covariate (d,s) boosts R_V^2 to 26.2% under $B^{(R)}$, performs best with 21% under $B^{(L)}$, and comes second with 12% under $B^{(E)}$ being beaten by $(Y^{(E)},s)$ with 14.1%.

In the one-year case, quite remarkable is the predictor π, either in itself or combined with covariates $Y^{(C)}, d, e, r, l$, under the inflation benchmark $B^{(C)}$ leading to R_V^2 in the range 9.5–15.4%. In addition, when

Table 3: Predictive power for five-year excess stock returns $Z_t^{(A)}$: the single benchmarking approach. The prediction problem is defined in (5). Additional notes: see Table 2.

Benchmark $B^{(A)}$	Explanatory variable(s) X_{t-1}						
	$Y^{(A)}$	d	e	r	l	π	s
Short-term rate	0.9	1.1	-1.5	7.8	1.4	-1.8	15.5
Long-term rate	1.1	4.6	0.5	3.9	1.0	-1.0	8.0
Earnings-by-price	1.4	-3.8	-3.6	-4.7	-1.4	-1.4	11.5
Inflation	1.3	7.6	-3.9	-6.7	-3.5	6.8	0.8
	$(Y^{(A)},d)$	$(Y^{(A)},e)$	$(Y^{(A)},r)$	$(Y^{(A)},l)$	$(Y^{(A)},\pi)$	$(Y^{(A)},s)$	
Short-term rate	-1.5	-2.8	8.2	2.2	-1.7	16.4	
Long-term rate	2.4	-0.3	4.4	1.6	-0.5	9.0	
Earnings-by-price	-4.5	-3.9	-4.9	-0.6	-0.4	14.1	
Inflation	5.9	-4.8	-5.7	-3.0	7.3	2.2	
	(d,e)	(d,r)	(d,l)	(d,π)	(d,s)		
Short-term rate	-3.1	6.4	-4.1	-2.3	26.2		
Long-term rate	1.0	5.8	-0.1	1.6	21.0		
Earnings-by-price	-7.5	-12.1	-6.6	-6.1	12.0		
Inflation	3.4	0.3	1.8	10.8	13.7		
	(e,r)	(e,l)	(e,π)	(e,s)			
Short-term rate	8.1	-4.5	-1.5	19.1			
Long-term rate	7.9	-3.3	1.7	12.6			
Earnings-by-price	-11.4	-8.8	-4.5	10.4			
Inflation	-9.5	-12.5	9.7	-1.1			
	(r,l)	(r,π)	(r,s)				
Short-term rate	14.7	5.6	14.8				
Long-term rate	6.5	2.7	7.0				
Earnings-by-price	9.2	-6.5	9.0				
Inflation	-6.4	0.9	-5.9				
	(l,π)	(l,s)					
Short-term rate	-1.4	13.4					
Long-term rate	-0.5	5.8					
Earnings-by-price	-1.8	9.2					
Inflation	7.5	-5.8					
	(π,s)						
Short-term rate	15.4						
Long-term rate	8.8						
Earnings-by-price	11.0						
Inflation	8.5						

put together with the term spread, the resulting combination (π, s) under $B^{(C)}$ is the clear winner reaching up to $R_V^2 = 15.4\%$. This is probably good news in an actuarial context where the inflation benchmark can be seen as an important one in pension product applications. In the five-year case, π still does quite well in the range 6.8–10.8% (with an exception of 0.9% for (r, π)) and remains generally the best predictor under $B^{(C)}$, nevertheless it is no longer the globally best one.

3.4 Full benchmarking approach

The next step now is to analyze whether transforming the explanatory variables can improve predictions. Recall that fully nonparametric models suffer in general by the curse of dimensionality, as in our framework where we confront sparsely distributed annual observations in higher dimensions. Importing more structure in the estimation process can help reduce or circumvent such problems.

Here, we extend the study in Section 3.3 using economic structure in the sense that we consider adjusting both the independent and dependent variables according to the same benchmark. To this end, in our full (double) benchmarking approach, the prediction problems are reformulated as

$$Y_t^{(A)} = m(X_{t-1}^{(A)}) + \xi_t, \tag{11}$$

$$Z_t^{(A)} = m(X_{t-1}^{(A)}) + \nu_t, \tag{12}$$

where we use transformed predictive variables

$$X_{t-1}^{(A)} = \begin{cases} \frac{1+X_{t-1}}{B_{t-1}^{(A)}}, \quad X \in \{d, e, r, l, \pi\} \\ \frac{s_{t-1}}{B_{t-1}^{(A)}} = \frac{l_{t-1} - r_{t-1}}{B_{t-1}^{(A)}} \\ Y_{t-1}^{(A)} \end{cases}, \quad A \in \{R, L, E, C\}. \tag{13}$$

This approach can be interpreted as a way of reducing dimensionality of the estimation procedure as $X_{t-1}^{(A)}$ encompasses an additional predictive variable.

Results of this empirical study are presented for the one-year horizon in Table 4 and for the five-year horizon in Table 5.

We find that, in the majority of cases, the full outruns the single benchmarking approach, even more when we consider a longer horizon, and the number of models with nonnegative R_V^2 (i.e., cases of beating the historical average excess return) increases: 68 out of 82 models (full benchmarking, five-year); 55 out of 82 (full benchmarking, one-year); 64 out of 112 (single benchmarking, five-year); and 52 out of 112 (single benchmarking, one-year).

The pair $(d^{(R)}, s^{(R)})$ in the full benchmarking approach for a five-year horizon yields $R_V^2 = 21.9\%$ against 26.2% in the single benchmarking under $B^{(R)}$, whereas $(e^{(C)}, s^{(C)})$ in the full benchmarking approach for a one-year horizon yields $R_V^2 = 17.8\%$ against 15.4% using the predictor (π, s) in the single benchmarking under $B^{(C)}$. It therefore seems that s is an important predictor, whose power is mostly unveiled when combining with another predictor depending on the benchmark choice and the forecast horizon. In addition, although under $B^{(R)}$ and $B^{(L)}$, full benchmarking does not improve predictability, it does under $B^{(E)}$ and, especially, $B^{(C)}$ which is important if we aim to identify a likely common well-performing benchmark and predictor, that is, $(e^{(C)}, s^{(C)})$, independently of the horizon length. For $B^{(C)}$ full benchmarking, R_V^2 lies in the range 14.7–10.1% (five-year) and 17.8–11.5% (one-year), which are both an improvement from $B^{(C)}$ single benchmarking yielding R_V^2 in 13.7–0.9% (five-year) and 15.4–9.5% (one-year), that is, a maximum width reduction by a factor of almost 3 for the five-year horizon.

Overall, we conclude that term spread is a good predictor; if we aim to homogenize our choice of predictor and benchmark with respect to the horizon length, then earnings-by-price and term spread under the inflation benchmark would be an ideal compromise, even if not the winning one. This is welcoming, as, for example, in pension research or other long-term saving strategies, for example, it is sensible to look at real value and employ such a model with all returns and covariates net-of-inflation.

3.5 Real-income long-term pension prediction

In long-term pension planning, real-income protection is often an important aspect, see Merton (2014) and Gerrard et al. (2019a, 2019b). When optimizing investment asset allocation for the long term, one therefore needs a good econometric model in real terms. Based on the research in this paper and in Kyriakou et al. (2019), we are able to conclude that, in the natural double benchmark setting for real-income econometrics, it looks like that earnings divided by price is the natural covariate to consider. In Table 2 of Kyriakou et al. (2019) and in Table 4 of this paper, it is concluded that earnings divided by price is the best single covariate to use in the double inflation benchmark case and, in this paper's Table 5, it is concluded that this is also the case in the five-year view. On balance, we therefore conclude that the intuitively appealing earnings divided by price is a good long-term predictor for real-income forecasting. In the one-year view of Kyriakou et al. (2019), the nonparametric smoother estimated for the relationship between earnings divided by price and return in the inflation double benchmarking case has the exact functional form of a simple line. So, even though we consider a nonparametric estimator that can pick any functional form, the resulting functional form is a simple line. This provides us with a strong argument for using the simple line in this case. The

Table 4: Predictive power for one-year excess stock returns $Y_t^{(A)}$: the double benchmarking approach. The prediction problem is defined in (11). The predictive power (%) is measured by R_V^2 as defined in (10). The benchmarks $B^{(A)}$ considered are based on the short-term interest rate ($A \equiv R$), long-term interest rate ($A \equiv L$), earnings-by-price ratio ($A \equiv E$), and consumer price index ($A \equiv C$). The predictive variables used are $X_{t-1}^{(A)}$ using the indicated benchmark $B_{t-1}^{(A)}$ as shown in (13). X_{t-1} are given by the dividend-by-price ratio d_{t-1}, earnings-by-price ratio e_{t-1}, short-term interest rate r_{t-1}, long-term interest rate l_{t-1}, inflation π_{t-1}, term spread s_{t-1}, excess stock return $Y_{t-1}^{(A)}$, or the possible different pairwise combinations as indicated. "–" are not applicable cases of matched covariate with benchmark. Note: $s^{(R)}$ and $l^{(R)}$ (and their combinations with Y, d, e, π) have the same R_V^2 by construction of the transformed spread according to (13). For example, $s_{t-1}^{(R)} = (l_{t-1} - r_{t-1})/B_{t-1}^{(R)} = (1 + l_{t-1})/(1 + r_{t-1}) - 1$ and $l_{t-1}^{(R)} = (1 + l_{t-1})/(1 + r_{t-1})$. Similar is the case of $s^{(L)}$ and $r^{(L)}$.

Benchmark $B^{(A)}$	Explanatory variable(s) X_{t-1}						
	$Y^{(A)}$	$d^{(A)}$	$e^{(A)}$	$r^{(A)}$	$l^{(A)}$	$\pi^{(A)}$	$s^{(A)}$
Short-term rate	-1.6	3.1	5.2	–	9.5	-1.3	9.5
Long-term rate	-1.8	-0.2	0.7	6.1	–	-1.5	6.1
Earnings-by-price	-1.7	-2.3	–	-0.2	-1.0	-0.7	7.4
Inflation	-1.4	10.4	12.2	7.2	10.5	–	6.5
	$(Y^{(A)}, d^{(A)})$	$(Y^{(A)}, e^{(A)})$	$(Y^{(A)}, r^{(A)})$	$(Y^{(A)}, l^{(A)})$	$(Y^{(A)}, \pi^{(A)})$	$(Y^{(A)}, s^{(A)})$	
Short-term rate	1.8	3.2	–	6.0	-2.9	6.0	
Long-term rate	-1.9	-1.1	2.6	–	-3.1	2.6	
Earnings-by-price	-4.1	–	-2.3	-3.2	-2.7	4.3	
Inflation	10.8	11.5	6.2	9.6	–	4.1	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	2.4	–	9.8	1.5	9.8		
Long-term rate	-1.6	6.3	–	-1.8	6.3		
Earnings-by-price	–	-3.2	-3.6	-3.5	4.0		
Inflation	10.3	9.5	10.0	–	15.7		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	10.7	3.3	10.7			
Long-term rate	7.1	–	-0.5	7.1			
Earnings-by-price	–	–	–	–			
Inflation	11.4	11.3	–	17.8			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	3.6	–				
Earnings-by-price	4.9	-2.1	5.7				
Inflation	13.9	–	14.8				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	7.2	–					
Long-term rate	–	–					
Earnings-by-price	-2.4	5.4					
Inflation	–	14.7					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	7.2						
Long-term rate	3.6						
Earnings-by-price	5.1						
Inflation	–						

functional form of a line has been picked via a validation procedure against all functional forms. The linear expression is

$$\text{Real one-year stock return} = 0.004875 + 1.119 \times \text{real earnings-by-price}.$$

Table 5: Predictive power for five-year excess stock returns $Z_t^{(A)}$: the double benchmarking approach. The prediction problem is defined in (12). Additional notes: see Table 4.

Benchmark $B^{(A)}$	Explanatory variable(s) X_{t-1}						
	$Y^{(A)}$	$d^{(A)}$	$e^{(A)}$	$r^{(A)}$	$l^{(A)}$	$\pi^{(A)}$	$s^{(A)}$
Short-term rate	0.9	12.1	10.4	–	15.5	-2.5	15.5
Long-term rate	1.1	8.5	0.8	8.0	–	-1.6	8.0
Earnings-by-price	1.4	8.4	–	-4.9	-3.7	-0.7	11.4
Inflation	1.3	10.9	12.4	5.7	8.7	–	0.8
	$(Y^{(A)}, d^{(A)})$	$(Y^{(A)}, e^{(A)})$	$(Y^{(A)}, r^{(A)})$	$(Y^{(A)}, l^{(A)})$	$(Y^{(A)}, \pi^{(A)})$	$(Y^{(A)}, s^{(A)})$	
Short-term rate	10.8	9.9	–	16.5	-1.5	16.5	
Long-term rate	3.2	-0.1	9.2	–	-1.1	9.2	
Earnings-by-price	5.3	–	-5.3	-5.1	-0.2	14.0	
Inflation	11.1	12.4	5.8	9.0	–	2.2	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	8.5	–	21.9	7.3	21.9		
Long-term rate	1.4	13.8	–	5.0	13.8		
Earnings-by-price	–	-1.6	4.1	5.4	15.5		
Inflation	9.5	4.1	1.7	–	13.0		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	21.4	8.2	21.4			
Long-term rate	16.4	–	2.5	16.4			
Earnings-by-price	–	–	–	–			
Inflation	8.6	4.9	–	14.7			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	9.5	–				
Earnings-by-price	5.9	-6.2	6.0				
Inflation	10.8	–	10.0				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	16.0	–					
Long-term rate	–	–					
Earnings-by-price	-6.4	6.0					
Inflation	–	10.1					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	16.0						
Long-term rate	9.5						
Earnings-by-price	11.9						
Inflation	–						

Notice that a very good long-term predictor of real income can, therefore, be expressed as a simple linear relationship, where the expected return adds first 12% to the earnings divided by price and then another 0.5%. This is a very simple relationship that is easy to remember for long-term investors. Similarly to the one-year view, our validation procedure exactly picks a line against all other functional alternatives in the five-year view. The linear form for the five-year view is

$$\text{Real five-year stock return} = 0.2068 + 2.264 \times \text{real earnings-by-price}.$$

The top panel of Figure 5 shows the estimated nonparametric function $\hat{m}$ (red solid line) for the one-year horizon (left) and five-year horizon (right) under the double inflation benchmark for the earnings-by price covariate together with the corresponding linear model (dash-dotted green line), and a 45-degree line (dashed black line).[5] Note that the linear relationship discovered for the earnings-by-price predictor must not hold true for other covariates or their combinations. In those cases, the full benefit of our approach comes

[5] We present in Figure 5 the three single covariates with the largest R_V^2 ($T = 1$ and $T = 5$) for the double inflation benchmark case: earnings-by-price ratio (12.2% and 12.4%), dividend-by-price ratio (10.4% and 10.9%), and long-term interest rate (10.5% and 8.7%).

to its own. For example, the bottom panel of Figure 5 clearly shows nonlinearities for the five-year case when the long-term interest rate is considered. For a suitable statistical test (nonparametric versus linear model), see, for example, the test based on wild bootstrap proposed in Scholz et al. (2015).

4 Conclusions

In this paper, we extend the original working framework of Kyriakou et al. (2019) to forecasting stock returns from a one-year to a five-year horizon in excess of different benchmarks, including the short-term rate, long-term rate, earnings-by-price ratio and inflation. We use predictors such as the dividend-by-price ratio, earnings-by-price ratio, short interest rate, long interest rate, term spread, inflation, as well as the lagged excess stock return, in one- and two-dimensional settings, with the returns benchmarked or also the covariates used to predict them.

We find that for both one-year and five-year returns, the term spread is, overall, the most powerful predictive variable for excess stock returns. Combining this with the dividend-by-price in the five-year case boosts the predictive power to a maximum. In the one-year case, quite remarkable is the inflation predictor under the inflation benchmark either in itself, or combined with other covariates such as the term spread to achieve a best-performing pair for the given horizon. Notice that earnings seem to be the better overall predictor when working net-of-inflation. The double benchmarking approach has earnings as the best individual predictor net with the inflation benchmark, where it is almost as strong a predictor as when using earnings and spread combined. Based on the results of this paper and also of Kyriakou et al. (2019), we therefore conclude that modelling earnings-by-price is a good and relatively simple starting point when constructing forecasting models for real-value pension prognoses.

Finally, aiming for the right balance between the two different horizons, we can say that this is given by the earnings-by-price and term spread set of predictors under the inflation benchmark. This is an important observation for this paper which expresses an interest in going beyond the standard short rate benchmark, such as when modelling returns in real terms (inflation benchmark) that is relevant for long-term saving strategies.

Funding: The third author acknowledges that this research was funded by Institute and Faculty of Actuaries in the UK through the grant "Minimizing Longevity and Investment Risk while Optimizing Future Pension Plans". The second author acknowledges financial support from the Cass Business School Doctoral Training Programme.

Acknowledgments: The usual disclaimer applies.

Conflicts of Interest: The authors declare no conflict of interest.

References

Athey, S. and Imbens, G. W. (2019) Machine learning methods economists should know about. Stanford Graduate School of Business. Working Paper No. 3776.

Bergmeir, C., Hyndman, R. J. and Koo, B. (2018) A note on the validity of cross-validation for evaluating autoregressive time series predictions. *Computational Statistics and Data Analysis*, **120**, 70–83.

Burman, P., Chow, E. and Nolan, D. (1994) A cross-validatory method for dependent data. *Biometrika*, **81**, 351–358.

Chen, Q. and Hong, Y. (2010) Predictability of equity returns over diferent time horizons: a nonparametric approach. Cornell University/Department of Economics. Working Paper.

Cheng, T., Gao, J. and Linton, O. (2019) Nonparametric predictive regressions for stock return predictions. Cambridge Working Papers in Economics: 1932.

Chu, C. K. and Marron, J. S. (1991) Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, **19**, 1906–1918.

Coulombe, P. G., Leroux, M., Stevanovic, D. and Surprenant, S. (2019) How is machine learning useful for macroeconomic forecasting? Working Paper.

De Brabanter, K., De Brabanter, J., Suykens, J. and De Moor, B. (2011) Kernel regression in the presence of correlated errors. *Journal of Machine Learning Research*, **12**, 1955–1976.

Doukhan, P. (1994) *Mixing*. New York: Springer.

Fama, E. F. and French, K. R. (2018) Long-horizon returns. *The Review of Asset Pricing Studies*, **8**, 232–252.

Geller, J. and Neumann, M. H. (2018) Improved local polynomial estimation in time series regression. *Journal of Nonparametric Statistics*, **30**, 1–27.

Gerrard, R., Hiabu, M., Kyriakou, I. and Nielsen, J. P. (2019a) Communication and personal selection of pension saver's financial risk. *European Journal of Operational Research*, **274**, 1102–1111.

— (2019b) Self-selection and risk sharing in a modern world of life-long annuities. *British Actuarial Journal*, **23**.

Gu, S., Kelly, B. and Xiu, D. (2018) Empirical asset pricing via machine learning. Chicago Booth Research Paper No. 18-04.

Hansen, B. (2006) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**, 726–748.

Hastie, T., Tibshirani, R. and Friedman, J. (2017) *The elements of statistical learning*. New York: Springer.

Kyriakou, I., Mousavi, P., Nielsen, J. P. and Scholz, M. (2019) Forecasting benchmarks of long-term stock returns via machine learning. URL: `http://link.springer.com/article/10.1007/s10479-019-03338-4`. Annals of Operations Research (available as Online First).

Lettau, M. and Van Nieuwerburgh, S. (2008) Reconciling the return predictability evidence. *Review of Financial Studies*, **21**, 1601–1652.

Li, Q. and Racine, J. (2007) *Nonparametric Econometrics*. Princeton and Oxford: Princeton University Press.

Linton, O. B. and Mammen, E. (2008) Nonparametric transformation to white noise. *Journal of Econometrics*, **142**, 241–264.

Masry, E. (1996) Multivariate regression estimation local polynomial fitting for time series. *Stochastic Processes and their Applications*, **65**, 81–101.

Merton, R. (2014) The crisis in retirement planning. *Harvard Business Review*, **92**, 43–50.

Munk, C. and Rangvid, J. (2018) New assumptions of a pension forecast model: Background, level and consequences for individuals forecasted pension. *Finans/Invest*, 6–14.

Nielsen, J. and Sperlich, S. (2003) Prediction of stock returns: A new way to look at it. *ASTIN Bulletin*, **33**, 399–417.

Opsomer, J., Wang, Y. and Yang, Y. (2001) Nonparametric regression with correlated errors. *Statistical Science*, **16**, 134–153.

Pesaran, H. and Timmermann, A. (2002) Market timing and return predictability under model instability. *Journal of Empirical Finance*, **8**, 495–510.

Rapach, D. and Wohar, M. (2004) Structural change and the predictability of stock returns. St. Louis University and University of Nebraska at Omaha Working Paper.

Rapach, D. and Zhou, G. (2013) Forecasting stock returns. In *Handbook of Economic Forecasting* (eds. G. Elliott and A. Timmerman), 328–383. Amsterdam: Elsevier, 2a edn.

Resnick, B. G. and Shoesmith, G. L. (2002) Using the yield curve to time the stock market. *Financial Analysts Journal*, **58**, 82–90.

Scholz, M., Nielsen, J. and Sperlich, S. (2015) Nonparametric prediction of stock returns based on yearly data: The long-term view. *Insurance: Mathematics and Economics*, **65**, 143–155.

Scholz, M., Sperlich, S. and Nielsen, J. (2016) Nonparametric long term prediction of stock returns with generated bond yields. *Insurance: Mathematics and Economics*, **69**, 82–96.

Shiller, R. (1989) *Market Volatility*. Cambridge: MIT Press.

Su, L. and Ullah, A. (2006) More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory*, **22**, 98–126.

Welch, I. and Goyal, A. (2008) A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**, 1455–1508.

Xiao, Z., Linton, O. B., Carroll, R. J. and Mammen, E. (2003) More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of the American Statistical Association*, **98**, 980–992.

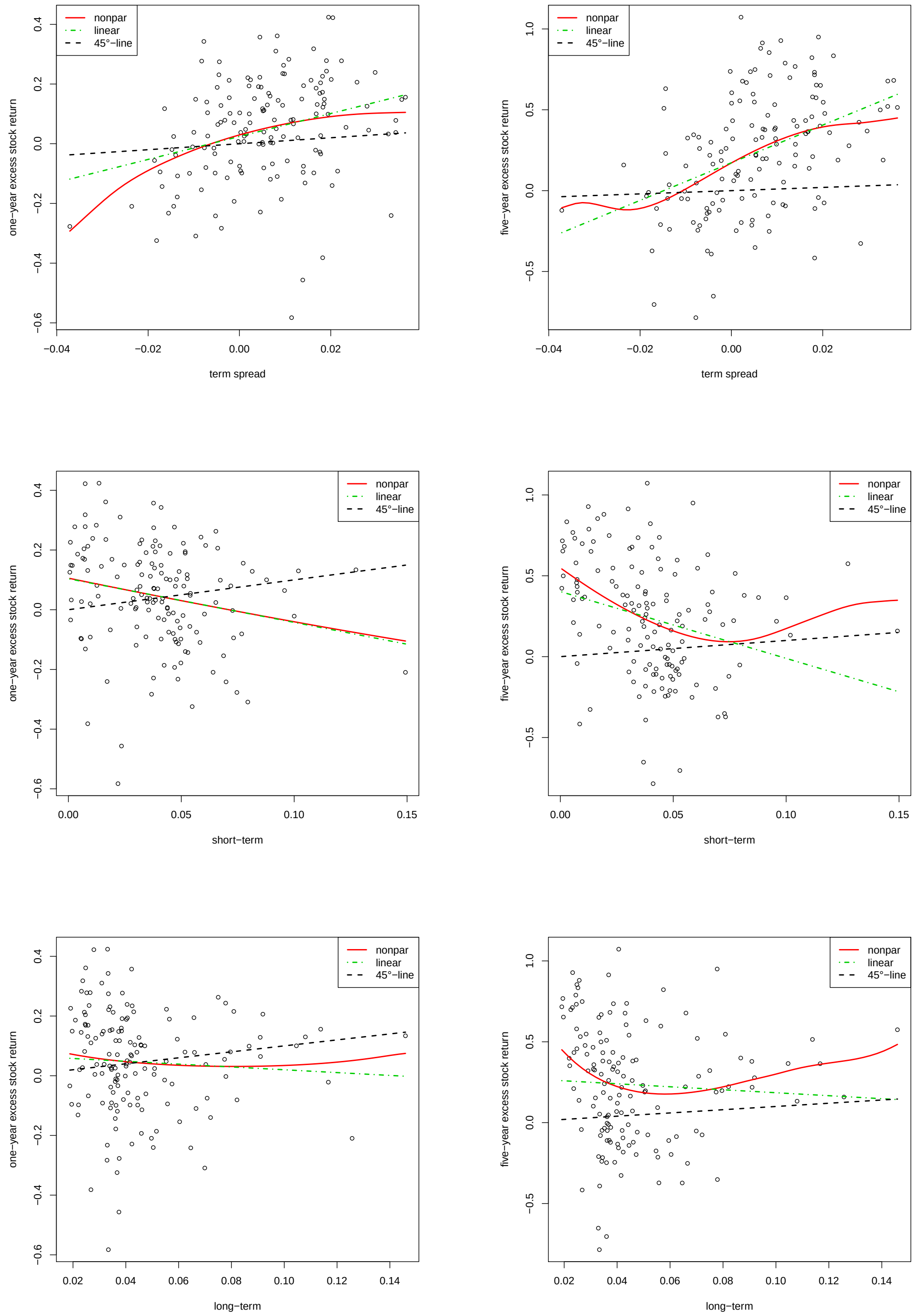

Figure 4: Single risk-free benchmark. Relation between excess stock returns and the spread (top), the short-term interest rate (middle), and the long-term interest rate (bottom). Estimated nonparametric function $\hat{m}$ (red solid line), linear model (dash-dotted green line), and 45-degree line (dashed black line). Left: one-year horizon. Right: five-year horizon. Period: 1872–2019. Data: annual S&P 500.

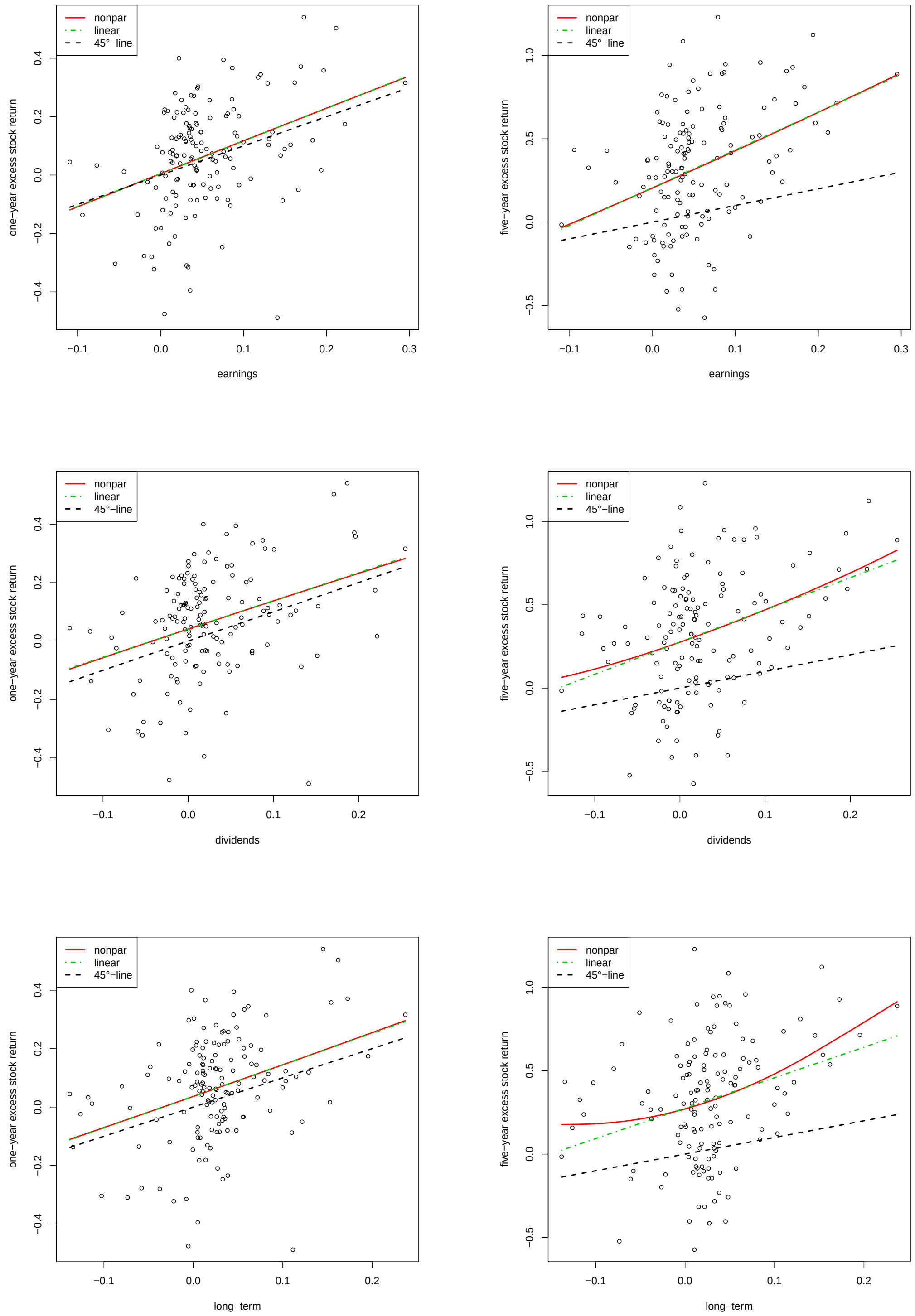

Figure 5: Double inflation benchmark. Relation between real stock returns and real earnings-by-price (top), real dividends-by-price (middle), and the real long-term interest rate (bottom). Estimated nonparametric function $\hat{m}$ (red solid line), linear model (dash-dotted green line), and 45-degree line (dashed black line). Left: one-year horizon. Right: five-year horizon. Period: 1872–2019. Data: annual S&P 500.

Graz Economics Papers

For full list see:
http://ideas.repec.org/s/grz/wpaper.html
Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

05–2019	**Ioannis Kyriakou, Parastoo Mousavi, Jens Perch Nielsen and Michael Scholz**: Machine Learning for Forecasting Excess Stock Returns The Five-Year-View
05–2019	**Robert J. Hill, Miriam Steurer and Sofie R. Waltl**: Owner Occupied Housing in the CPI and Its Impact On Monetary Policy During Housing Booms
04–2019	**Thomas Aronsson, Olof Johansson-Stenman and Ronald Wendner**: Charity, Status, and Optimal Taxation: Welfarist and Paternalist Approaches
03–2019	**Michael Greinecker and Christoph Kuzmics**: Limit Orders under Knightian Uncertainty
02–2019	**Miriam Steurer and Robert J. Hill**: Metrics for Measuring the Performance of Machine Learning Prediction Models: An Application to the Housing Market
01–2019	**Katja Kalkschmied and Jörn Kleinert**: (Mis)Matches of Institutions: The EU and Varieties of Capitalism
21–2018	**Nathalie Mathieu-Bolh and Ronald Wendner**: We Are What We Eat: Obesity, Income, and Social Comparisons
20–2018	**Nina Knittel, Martin W. Jury, Birgit Bednar-Friedl, Gabriel Bachner and Andrea Steiner**: The implications of climate change on Germanys foreign trade: A global analysis of heat-related labour productivity losses
19–2018	**Yadira Mori-Clement, Stefan Nabernegg and Birgit Bednar-Friedl**: Can preferential trade agreements enhance renewable electricity generation in emerging economies? A model-based policy analysis for Brazil and the European Union
18–2018	**Stefan Borsky and Katja Kalkschmied**: Corruption in Space: A closer look at the world's subnations

17–2018 **Gabriel Bachner, Birgit Bednar-Friedl and Nina Knittel**: How public adaptation to climate change affects the government budget: A model-based analysis for Austria in 2050

16–2018 **Michael Günther, Christoph Kuzmics and Antoine Salomon**: A Note on Renegotiation in Repeated Games [Games Econ. Behav. 1 (1989) 327360]

15–2018 **Meng-Wei Chen, Yu Chen, Zhen-Hua Wu and Ningru Zhao**: Government Intervention, Innovation, and Entrepreneurship

14–2018 **Yu Chen, Shaobin Shi and Yugang Tang**: Valuing the Urban Hukou in China: Evidence from a Regression Discontinuity Design in Housing Price

13–2018 **Stefan Borsky and Christian Unterberger**: Bad Weather and Flight Delays: The Impact of Sudden and Slow Onset Weather Events

12–2018 **David Rietzke and Yu Chen**: Push or Pull? Performance-Pay, Incentives, and Information

11–2018 **Xi Chen, Yu Chen and Xuhu Wan**: Delegated Project Search

10–2018 **Stefan Nabernegg, Birgit Bednar-Friedl, Pablo Muñoz, Michaela Titz and Johanna Vogel**: National policies for global emission reductions: Effectiveness of carbon emission reductions in international supply chains

09–2018 **Jonas Dovern and Hans Manner**: Order Invariant Tests for Proper Calibration of Multivariate Density Forecasts

08–2018 **Ioannis Kyriakou, Parastoo Mousavi, Jens Perch Nielsen and Michael Scholz**: Choice of Benchmark When Forecasting Long-term Stock Returns

07–2018 **Joern Kleinert**: Globalization Effects on the Distribution of Income

06–2018 **Nian Yang, Jun Yang and Yu Chen**: Contracting in a Continuous-Time Model with Three-Sided Moral Hazard and Cost Synergies

05–2018 **Christoph Kuzmics and Daniel Rodenburger**: A case of evolutionary stable attainable equilibrium in the lab

04–2018 **Robert J. Hill, Alicia N. Rambaldi, and Michael Scholz**: Higher Frequency Hedonic Property Price Indices: A State Space Approach

03–2018 **Reza Hajargasht, Robert J. Hill, D. S. Prasada Rao, and Sriram Shankar**: Spatial Chaining in International Comparisons of Prices and Real Incomes

02–2018 **Christoph Zwick**: On the origin of current account deficits in the Euro area periphery: A DSGE perspective

01–2018 **Michael Greinecker and Christopher Kah**: Pairwise stable matching in large economies

15–2017 **Florian Brugger and Jörn Kleinert**: The strong increase of Austrian government debt in the Kreisky era: Austro-Keynesianism or just stubborn forecast errors?

14–2017 **Jakob Mayer, Gabriel Bachner and Karl W. Steininger**: Macroeconomic implications of switching to process-emission-free iron and steel production in Europe

13–2017 **Andreas Darmann, Julia Grundner and Christian Klamler**: Consensus in the 2015 Provincial Parliament Election in Styria, Austria: Voting Rules,Outcomes, and the Condorcet Paradox

12–2017 **Robert J. Hill, Miriam Steurer and Sofie R. Waltl**: Owner Occupied Housing in the CPI and Its Impact On Monetary Policy During Housing Booms and Busts

11–2017 **Philipp Kohlgruber and Christoph Kuzmics**: The distribution of article quality and inefficiencies in the market for scientific journals

10–2017 **Maximilian Goedl**: The Sovereign-Bank Interaction in the Eurozone Crisis

09–2017 **Florian Herold and Christoph Kuzmics**: The evolution of taking roles

08–2017 **Evangelos V. Dioikitopoulos, Stephen J. Turnovsky and Ronald Wendner**: Dynamic Status Effects, Savings, and Income Inequality

07–2017 **Bo Chen, Yu Chen and David Rietzke**: Simple Contracts under Observable and Hidden Actions

06–2017 **Stefan Borsky, Andrea Leiter and Michael Paffermayr**: Product Quality and Sustainability: The Effect of International Environmental Agreements on Bilateral Trade

05–2017 **Yadira Mori Clement and Birgit Bednar-Friedl**: Do Clean Development Mechanism projects generate local employment? Testing for sectoral effects across Brazilian municipalities

04–2017 **Stefan Borsky, Alexej Parchomenko**: Identifying Phosphorus Hot Spots: A spatial analysis of the phosphorus balance as a result of manure application

03–2017 **Yu Chen, Yu Wang, Bonwoo Koo**: Open Source and Competition Strategy Under Network Effects

02–2017 **Florian Brugger**: The Effect of Foreign and Domestic Demand on U.S. Treasury Yields

01–2017 **Yu Chen**: On the Equivalence of Bilateral and Collective Mechanism Design

13–2016 **Jan de Haan, Rens Hendriks, Michael Scholz**: A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes

12–2016 **Katja Kalkschmied**: Complementary Institutional Elements and Economic Outcomes

11–2016 **Dieter Balkenborg, Josef Hofbauer, Christoph Kuzmics**: The refined best reply correspondence and backward induction

10–2016 **Karl W. Steininger, Wolf D. Grossmann, Iris Grossmann**: The economic value of dispatchable solar electricity: a Post-Paris evaluation

09–2016 **Sofie R. Waltl**: Estimating aggregate quantile-specific gross rental yields for residential housing in Sydney

08–2016 **Karl Farmer**: Financial integration and house price dynamics in equilibrium modeling of intra-EMU and global trade imbalances

07–2016 **Gabriel Bachner, Birgit Bednar-Friedl**: Counterbalancing the Effects of Climate Change Adaptation on Public Budgets: Factor Taxes, Transfers, or Foreign Lending?

06–2016 **Margareta Kreimer, Ricardo Mora**: Segregated Integration: Recent Trends in the Austrian Gender Division of Labor

05–2016 **Karl Farmer, Bogdan Mihaiescu**: Credit constraints and differential growth in equilibrium modeling of EMU and global trade imbalances

04–2016 **Daniel Eckert, Frederik Herzberg**: The birth of social choice theory from the spirit of mathematical logic: Arrow's theorem in the framework of model theory

03–2016 **Christoph Kuzmics, Jan-Henrik Steg**: On public good provision mechanisms with dominant strategies and balanced budget

02–2016 **Florian Gauer, Christoph Kuzmics**: Cognitive Empathy in Conflict Situations

01–2016 **Thomas Aronsson, Olof Johansson-Stenman, Ronald Wendner**:
Redistribution Through Charity, and Optimal Taxation when People are Concerned with Social Status