Graz Economics Papers – GEP

GEP 2019–08

Conditional variance forecasts for long-term stock returns

Enno Mammen, Jens Perch Nielsen, Michael Scholz and Stefan Sperlich

August 2019

Department of Economics
Department of Public Economics
University of Graz

Conditional variance forecasts for long-term stock returns

A Preprint

Enno Mammen
Institute for Applied Mathematics
University of Heidelberg
Heidelberg, Germany
mammen@math.uni-heidelberg.de

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

Stefan Sperlich
Institute of Economics and Econometrics
University of Geneva
Geneva, Switzerland
stefan.sperlich@unige.ch

August 21, 2019

Abstract

In this paper, we apply machine learning to forecast the conditional variance of long-term stock returns measured in excess of different benchmarks, including the short-term interest rate, long-term interest rate, earnings-by-price ratio, and inflation. In particular, we apply and implement in a two-step procedure a fully nonparametric smoother with the covariates and the smoothing parameters chosen via cross-validation. We find that volatility forecastability is much less important at longer horizons regardless of the chosen model and that the homoscedastic historical average of the squared return prediction errors gives an adequate approximation of the unobserved realised conditional variance for both the one-year and five-year horizon.

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

JEL: C14; C53; C58; G17; G22

1 Introduction

The volatility of financial assets has important implications for the theory and practice of asset pricing, portfolio selection, risk management, and market-timing strategies. Therefore, it is of fundamental interest to measure *ex-ante*, or forecast successfully, the conditional variance of returns. Of cause, the evaluation of the latter and the forecasting itself have been complicated by the unobservability of the realised conditional variance (Galbraith and Kisinbay, 2005). An extensive amount of research is engaged in analysing the distributional and dynamic properties of stock market volatility, see, for example, Andersen et al. (2001) and citations therein. The standard approaches applied include estimation of parametric (G)ARCH-type or stochastic volatility models for the underlying returns based upon specific distributional assumptions. Alternatives, especially for data of higher frequency, are based on constructing model-free estimates of

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

ex-post *realized* volatilities by summing squares and cross-products of intraday high-frequency returns (Andersen et al., 2001).

The present paper instead uses annual U.S. stock market data to construct excess stock returns at the one-year and five-year horizon and to examine their model-based variance forecasts. Note that the risk depends on the horizon considered, and different horizons are relevant for different applications (Christoffersen and Diebold, 2000). Little is known about the forecastability of variance at horizons beyond a year. Here we take the long-term actuarial view and extend the work of Kyriakou et al. (2019a, 2019b). In a two-step procedure, we first apply machine learning (ML) to forecast stock returns in excess of different benchmarks, including the short-term interest rate, long-term interest rate, earnings-by-price ratio, and inflation. Second, the squared residuals are used to analyse model-based volatility forecastability. Here, we compare these forecasts with the forecast implicit in the unconditional residual variance, as proposed, for example, by Galbraith and Kisinbay (2005). We find that volatility forecastability is much less important at longer horizons regardless of the chosen model and that the homoscedastic historical average of the squared return prediction errors gives an adequate approximation of the unobserved realised conditional variance for both the one-year and five-year horizon.

Our preferred ML technique applied in this paper is local-linear smoothing in combination with a leave-*k*-out cross-validation for the following reasons. First, we are interested in longer-horizon stock returns based on annual observations and their volatility. Thus, we are not in the high-frequency 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). 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 as documented, for example, in Lettau and Van Nieuwerburgh (2008), Chen and Hong (2010), or Cheng et al. (2019). Thus, the local-linear smoother 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 glasshouse, not in a black box – in contrast to other fancier but less clear ML methods.

Note further that longer horizons are important to long-term investors, such as pension funds or market participants saving for distant payoffs. These investors are generally willing to take on more risk for higher rewards and, thus, volatility forecastability is for them of fundamental interest. Rapach and Zhou (2013) show that longer horizons tend to produce better estimates than shorter horizons. While 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 exemplarily concentrate on the one-year and five-year view. However, shorter horizons based on monthly, weekly, or even daily data do not seem to provide the pension saver with good information about 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 her risk appetite.

The remaining of this paper is organized as follows. In Section 2, we present our framework for the purpose of conditional variance prediction. We define the underlying financial model, introduce our two-step procedure, and present our validation criterion for model selection. In addition, we review different ways of estimating the conditional variance and discuss bootstrap-tests for the null hypothesis of no predictability. In Section 3, we provide a description of our data set and of 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 conditional variance prediction

In this section, we focus on nonlinear predictive relationships between squared residuals of model-based predicted 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 volatility predictability over return 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 $B_{t-1}^{(A)}$:

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

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 the location-scale model

$$Y_t^{(A)} = m(X_{t-1}) + \nu(X_{t-1})^{1/2}\zeta_t, \tag{2}$$

where

$$m(x) = \mathbb{E}(Y^{(A)}|X = x) \text{ and } \nu(x) = Var(Y^{(A)}|X = x), \ x \in \mathbb{R}^q, \tag{3}$$

are the unknown smooth conditional mean- and variance-function, resp., and ζ_t are serially uncorrelated zero-conditional-mean random error terms, given the past, with the conditional variance of one.

Our aim is to forecast the conditional variance of 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 $Y_{t-1}^{(A)}$. Note that the set of explanatory variables X_{t-1} in (2) could be different or overlapping for the mean and variance function.

Based on (2), in a two-step procedure we first estimate $\hat{Y}_t^{(A)} = \hat{m}(X_{t-1})$ as in Kyriakou et al. (2019b), and in a second step, we estimate $\hat{\nu}(X_{t-1})$ from

$$\nu(x) = \mathbb{E}((Y^{(A)} - m(X))^2|X = x), \ x \in \mathbb{R}^q, \tag{4}$$

using the squared residuals $\hat{\varepsilon}_t^2 := (Y_t^{(A)} - \hat{m}(X_{t-1}))^2$ as the dependent variable and a local-linear smoother in both steps.[2]

2.2 Longer-horizon predictions

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 (2) between $Y_t^{(A)}$ and X_{t-1}, as well as the 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 $\xi_t := \nu_\theta(X_{t-1})^{1/2}\zeta_t$ similar to the error term in (2) and a parametric specification for the conditional variance $\nu_\theta(\cdot)$, and η_t being white noise. The T-year regression problem that is implied by this pair of one-year regressions is now

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

[2] For a description and statistical properties of the local-linear smoother, see, for example, Section 2.3 in Kyriakou et al. (2019b). Note further that the smoothing parameters are separately chosen in each step.

i.e., the excess stock return for the 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 ψ_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 (2)

$$Z_t^{(A)} = m(X_{t-1}) + \nu(X_{t-1})^{1/2}\omega_t, \tag{5}$$

where

$$m(x) = \mathbb{E}(Z^{(A)}|X = x) \text{ and } \nu(x) = Var(Z^{(A)}|X = x),\ x \in \mathbb{R}^q, \tag{6}$$

are the unknown smooth conditional mean- and variance-function. The important difference between (2) and (5) is now that the error process $\psi_t := \nu(X_{t-1})^{1/2}\omega_t$ in (5) will be serially correlated by construction.[3] The predictive variables X under consideration are again 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 $Y^{(A)}$.

Based on (5), our two-step procedure consists now of, first, estimating $\hat{Z}_t^{(A)} = \hat{m}(X_{t-1})$, and second, estimating $\hat{\nu}(X_{t-1})$ from

$$\nu(x) = \mathbb{E}((Z^{(A)} - m(X))^2|X = x),\ x \in \mathbb{R}^q, \tag{7}$$

using the squared residuals $\hat{\varepsilon}_t^2 := (Z_t^{(A)} - \hat{m}(X_{t-1}))^2$ as the dependent variable and a local-linear smoother again in both steps.

2.3 Alternative ways in estimating the conditional variance function

For the estimation of the conditional variance or volatility function of a response variable Y in a location-scale model similar to (2) or (5), mainly four different approaches are proposed in the literature: the direct, the residual-based, the likelihood-based, and the difference-sequence method.

The direct method, see, for example, Härdle and Tsybakov (1997), uses the variance expressed as the difference of the first two conditional moments:

$$Var(Y|X = x) = \mathbb{E}(Y^2|X = x) - \mathbb{E}(Y|X = x)^2. \tag{8}$$

Both parts of the right-hand side in (8) are separately estimated and, thus, the result is not necessarily nonnegative and also not fully adaptive to the mean function.[4]

As described above, the residual-based method consists of two stages: First, estimating the conditional mean function $m(\cdot)$ and calculating $\hat{\varepsilon}^2 = (Y - \hat{m}(X))^2$, and second, estimating the conditional variance function $\nu(\cdot)$ by regressing the squared residuals $\hat{\varepsilon}^2$ on a set of explanatory variables X. There exist different variants of residual based methods for the second step: (i) Fan and Yao (1998) apply a local-linear kernel smoother in both stages. The result is again not necessarily nonnegative but asymptotically fully adaptive to the unknown mean function. (ii) Ziegelmann (2002) proposes the local exponential estimator to ensure nonnegativity:

$$\sum_{t=1}^{T} \left[\hat{\varepsilon}_t^2 - \Psi\{\alpha + \beta(X_{t-1} - x)\}\right]^2 K_h(X_{t-1} - x) \quad \Rightarrow \quad Min_{\alpha,\beta},$$

where $\Psi(x) \equiv \exp(x)$, K_h denotes some kernel function which depends on a set of bandwidths h. (iii) Mishra et al. (2010) apply the combined estimator proposed first by Glad (1998) in the context of mean-regression and based on the identity

$$\nu = \nu_\theta \cdot \frac{\nu}{\nu_\theta}, \tag{9}$$

where a parametric guide ν_θ captures some roughness features of ν and the nonparametric correction factor (the ratio in the product of (9)) will be easier to estimate. Note that Mishra et al. (2010) apply their method to

[3] For possible solutions to the problem of autocorrelation, see, for example, Xiao et al. (2003), Su and Ullah (2006), Linton and Mammen (2008), or more recently Geller and Neumann (2018). The implementation and analysis of these techniques remain for future research. In our approach, we account for autocorrelation in the validation criterion with a leave-k-out strategy, where $k = 2T - 1$; see Section 2.4.

[4] It does not estimate the volatility function as efficiently as if the true mean function were known.

daily data and ignore the first step, the estimation of the mean function which is set to zero, and, thus, the possibility of bias reduction with the same method applied to excess stock returns as shown, for example, in Scholz et al. (2015). (iv) Xu and Phillips (2011) instead use a re-weighted local constant estimator maximising the empirical likelihood such that it becomes a bias-reducing moment restriction.

The preferred estimators of Yu and Jones (2004) are based on a localised normal likelihood, using a standard local-linear form for estimating the mean m and a local log-linear form for estimating the variance ν by maximising

$$-\sum_{t=1}^{T}\left\{\frac{(Y_t - m(X_{t-1}))^2}{\nu^{1/2}(X_{t-1})} + \log(\nu(X_{t-1}))\right\} K_h(X_{t-1}-x),$$

and allowing for separate bandwidths for mean and variance estimation.

Finally, examples for the difference-sequence method in a fixed design can be found for the homoscedastic case in Wang and Yu (2017) and citations therein. Wang et al. (2008) analyse for the heteroscedastic case the effect of the unknown mean on the estimation of the variance function. They also compare the performance of the residual-based estimators to a first-order-difference-based estimator. Their results indicate that, contrary to the common practice, it is not desirable to base the estimator of the variance function on the residuals from an optimal estimator of the mean when the mean function is not smooth. Instead, it is more desirable to use estimators of the mean with minimal bias.

In the empirical part of this paper in Section 3, we show the results of the residual-based method applying a local-linear kernel smoother in both stages. As a robustness check, we have implemented in the second step the local-exponential estimator (Ziegelmann, 2002) and the combined estimator (Mishra et al., 2010) getting almost always very similar results.[5] We do not consider: (i) the direct method, since it is not fully adaptive to the mean function, (ii) the re-weighted local constant estimator (Xu and Phillips, 2011) due to its asymptotic similarity to the local-linear method, (iii) the method based on the assumption of normal error terms (Yu and Jones, 2004), since skewness and excess kurtosis are common properties of stock returns, and (iv) the difference-sequence method, since it was not convincingly performing in a small sample study, the mean functions are rather smooth in our problem, and bias reduction is key due to sparsity.[6]

2.4 The principle of validation: model selection and the choice of smoothing parameters

As we use a nonparametric technique, we require an adequate measure of predictive power. Classical in-sample measures, such as the R^2 or adjusted R^2, are not appropriate because they either prefer the most complex model or need a degrees of freedom adjustment which is an unclear concept in nonparametric estimation. Furthermore, 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 and optimal bandwidth selection, we use an extended version of the validated R^2_V introduced in the actuarial literature by Nielsen and Sperlich (2003) and based on a leave-k-out cross-validation. Note that this criterion is very similar to the forecast content function of Galbraith (2003) and Galbraith and Kisinbay (2005) defined as the proportionate reduction in the mean squared forecast error available, relative to the unconditional mean forecast.

Our validation criteria for the first and second step are defined as

$$R^2_{V,m} = 1 - \frac{\sum_t (Z_t^{(A)} - \hat{m}_{-t})^2}{\sum_t (Z_t^{(A)} - \bar{Z}_{-t}^{(A)})^2} \quad \text{and} \quad R^2_{V,\nu} = 1 - \frac{\sum_t (\hat{\varepsilon}_t^2 - \hat{\nu}_{-t})^2}{\sum_t (\hat{\varepsilon}_t^2 - \overline{\hat{\varepsilon}^2_{-t}})^2}, \tag{10}$$

where leave-k-out estimators are used: $\hat{m}_{-t}$ and $\hat{\nu}_{-t}$ for the nonparametric functions m and ν, resp., $\bar{Z}_{-t}^{(A)}$ and $\overline{\hat{\varepsilon}^2_{-t}}$ for the unconditional mean of $Z_t^{(A)}$ and $\hat{\varepsilon}_t^2$, resp. All 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

[5] Those results are available upon request by the authors.

[6] There is also a lack of studies using the difference-sequence method in a random design and in multivariate problems as in our case.

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. Note that the validated R_V^2 measures the predictive power of a given model compared to the cross-validated historical mean. Thus, positive values imply that the predictor-based regression model outperforms the corresponding historical average over T years. Negative values in the first step of our approach suggest, that the historical mean return should be preferred over a model-based approach. While negative values in the second step indicate a constant homoscedastic conditional variance forecast.

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., 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 under-smoothed. 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 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 that this amounts to a leave-nine-out set of $Z_{t-4}^{(A)}, \ldots, Z_{t+4}^{(A)}$ (see, for example, Kyriakou et al., 2019b, Fig. 1).

2.5 A bootstrap-test: no predictability vs. predictability of the conditional variance

We test the null of no predictability of the conditional variance. Formally, this is equivalent to say that, under the null, ν is a constant function, which essentially corresponds to the historical average of the squared residuals, i.e. constant volatility. In particular, let $\nu(\cdot)$ be the true volatility function as in (2) or (5) for some specified set of regressors X_t, i.e. (4) or (7) holds. Let $\overline{\hat{\varepsilon}^2}$ be the sample mean of the squared residuals from step one in our approach. The distance

$$\int \left| \nu(x) - \overline{\hat{\varepsilon}^2} \right|^2 w(x)\, dx \tag{11}$$

for some weighting function w has been studied by several authors and statistics have been derived from the above, for example, in Härdle and Mammen (1993, HM) or Kreiss et al. (2008, KNY). We use the statistic derived in equation 2.3 of Kreiss et al. (2008)

$$h^{q/2} T \int \left| \frac{1}{T} \sum_{t=1}^{T} K_h(x - X_t) \left(\hat{\varepsilon}_t^2 - \overline{\hat{\varepsilon}^2} \right) \right|^2 w(x)\, dx, \tag{12}$$

where $K_h(x)$ is a symmetric kernel smoother with bandwidth h. The bandwidth is selected using R_V^2 for the Nadaraya-Watson kernel estimator rather than a local-linear one. We choose w to be proportional to the uniform density with support in the range of the sample data and replace integration by the mean over uniform independent observations $X_1', X_2', \ldots, X_N'$ in the range of the data:

$$\tau := \frac{h^{q/2} T}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} K_h(X_i' - X_t) \left(\hat{\varepsilon}_t^2 - \overline{\hat{\varepsilon}^2} \right) \right|^2. \tag{13}$$

Then the error in the integral is $O\left(N^{-1/2}\right)$ (Geweke, 1996). Under the null, the above test statistic τ is small. This choice could lead to a statistic whose power is lower than the one in HM due to some implicit over-smoothing resulting in the weight function w (see comment in Kreiss et al., 2008, just after their equation

2.5). Power may also improve by using a local-linear smoother in the test. However, the theory for this has not been developed, yet, so we refrain from such extension.

Critical values for τ are best derived via wild bootstrap (Härdle and Mammen, 1993). For the bootstrap critical values to be consistent, the procedure needs to be independent of whether the null is true or not. Hence, in correspondence to equation 2.10 in Kreiss et al. (2008), for $b = 1, \ldots, B$,

$$\tau^b := \frac{h^{q/2}T}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} K_h \left(X_i' - X_t \right) \left[u_t^b \left(\hat{\varepsilon}_t^2 - \hat{v}\left(X_t \right) \right) \right] \right|^2, \tag{14}$$

where the u_t^b's are independent and identically distributed random variables with mean zero and variance one, for example, $u_t^b \sim N(0, 1)$. To decide, if we reject or not, we use critical values from corresponding quantiles of the empirical distribution[7],

$$F^*(\tau) = \frac{1}{B} \sum_b \mathbb{1}_{\{\tau^b \le \tau\}}. \tag{15}$$

The consistency of the procedure for stationary sequences is given in Kreiss et al. (2008).

An alternative version for a wild bootstrap test is used in Scholz et al. (2015, SNS). There the B bootstrap samples are constructed using the residuals under the null, $\iota_t^0 := \hat{\varepsilon}_t^2 - \overline{\hat{\varepsilon}^2}$, and u_t^b's as above, such that

$$\hat{\varepsilon}_t^{2,b} = \overline{\hat{\varepsilon}^2} + \iota_t^0 \cdot u_t^b.$$

Then, in each bootstrap iteration b, the cross-validated mean is calculated of the $\hat{\varepsilon}_t^{2,b}$, $t = 1, \ldots, T$, as well the estimates of the alternative model $\hat{v}_{-t}^b$ in order to get $R_{V,v}^{2,b}$ like in (10). Critical values are chosen from corresponding quantiles of the empirical distribution function similar to (15).

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

3.1 The data

In this paper, we extend the analysis of Kyriakou et al. (2019b), who considered the forecasting of long-term stock returns, to conditional variance predictions. Thus, we base our predictions on the same 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 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%. Table 1 displays standard descriptive statistics for one-year and five-year returns as well as the available covariates.

7 The symbol $\mathbb{1}_A$ denotes the indicator function of an appropriate condition A, i.e. it is one when A is true and zero otherwise.

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
One-year excess stock returns $Y^{(C)}$	54.04	-48.81	6.41	18.05	-0.40	0.64
Five-year excess stock returns $Z^{(R)}$	107.27	-78.54	23.49	36.69	-0.14	-0.37
Five-year excess stock returns $Z^{(C)}$	122.96	-57.34	32.34	36.42	-0.05	-0.40
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

3.2 Single benchmarking approach

In this section, we consider a single benchmarking approach as in Kyriakou et al. (2019a,b) where only the variable S_t is adjusted according to some benchmark $B_{t-1}^{(A)}$, as shown in (1), while the independent variable(s) is (are) measured on the original nominal scale. The models (2) and (5) are estimated in both steps with a local-linear kernel smoother using the quartic kernel and the optimal bandwidths are chosen by cross-validation, i.e., by maximizing the corresponding validation measure 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 $R^2_{V,\nu}$ values based on different validated scenarios shown for the one-year horizon in Table 2 and the five-year horizon in Table 3. Here, the same predictive variables X_{t-1} are used in both steps of our approach.

Overall, we find for the one-year horizon that only a few variables have small positive validated $R^2_{V,\nu}$'s and thus, possibly some low explanatory power. For example, for the benchmarks $B^{(R)}$, $B^{(L)}$, and $B^{(E)}$, the excess stock return $Y_{t-1}^{(A)}$ has the largest validated $R^2_{V,\nu}$ values for one-dimensional models (2.2%, 2.4%, and 1.5%). This finding would support an ARCH-type variance structure. For the inflation benchmark $B^{(C)}$, the model with the long-term interest rate produces the largest validated $R^2_{V,\nu}$ of 0.5%. When we apply the bootstrap tests introduced in Section 2.5, the KNY-test does not reject the null of no predictability for all cases at the 5%-level. The SNS-test rejects the null only for the $Y_{t-1}^{(A)}$ covariate under the benchmarks $B^{(R)}$, $B^{(L)}$ and $B^{(E)}$ at the 5%-level.[8] Note that the two-dimensional models do not add predictive power as the validated $R^2_{V,\nu}$ values remain in the same low range.

Contrary to the mean prediction, where Kyriakou et al. (2019b) find that five-year predictability improves over the one-year case, we observe that the majority of predictor based volatility models do not surpass the constant volatility alternative for the five-year horizon. Even though some models produce small positive $R^2_{V,\nu}$ values, this time both the SNS- and the KNY-test do not reject the null of no predictability. Note that our results are in line with Christoffersen and Diebold (2000) who conclude that volatility forecastability may be much less important at longer horizons.

[8] The tests were conducted with 1000 repetitions at the 5% significance level. We do not present the p-values of the tests to save space. For the models which are not discussed in the main text, the p-values exceed the 5%-level. The results are available upon request by the authors.

Table 2: Predictive power for the variance of one-year excess stock returns $Y_t^{(A)}$: the single benchmarking approach. The prediction problem is defined in (2). The same predictive variables X_{t-1} are used in the predictions for the conditional mean and variance function. The predictive power (%) is measured by $R^2_{V,\nu}$ 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	2.2	-1.1	-0.6	-0.3	0.3	-1.2	-0.1
Long-term rate	2.4	-1.2	-0.6	0.3	0.6	-1.4	-0.1
Earnings-by-price	1.5	-1.3	-0.7	-0.1	0.5	-1.4	0.1
Inflation	0.2	0.1	-1.3	-0.4	0.5	-1.2	-0.6
	$(Y^{(A)},d)$	$(Y^{(A)},e)$	$(Y^{(A)},r)$	$(Y^{(A)},l)$	$(Y^{(A)},\pi)$	$(Y^{(A)},s)$	
Short-term rate	2.4	1.9	1.1	2.2	0.1	0.3	
Long-term rate	1.5	1.4	1.1	2.1	-0.2	0.1	
Earnings-by-price	1.6	1.4	0.9	2.0	-0.2	0.1	
Inflation	-1.0	-1.1	-0.6	0.6	-2.1	-1.0	
	(d,e)	(d,r)	(d,l)	(d,π)	(d,s)		
Short-term rate	-2.1	-1.5	-0.8	-2.4	-1.5		
Long-term rate	-2.0	-1.1	-0.6	-2.2	-1.5		
Earnings-by-price	-1.9	-1.4	-0.7	-2.3	-1.5		
Inflation	-0.4	-1.0	-0.2	-2.3	-1.3		
	(e,r)	(e,l)	(e,π)	(e,s)			
Short-term rate	-1.0	-0.4	-2.3	-0.8			
Long-term rate	-0.6	-0.2	-2.2	-0.8			
Earnings-by-price	-1.0	-0.2	-2.2	-0.8			
Inflation	-1.7	-0.9	-2.2	-1.6			
	(r,l)	(r,π)	(r,s)				
Short-term rate	1.3	-1.5	1.4				
Long-term rate	1.3	-1.0	1.4				
Earnings-by-price	1.4	-1.5	1.6				
Inflation	1.3	-1.5	1.2				
	(l,π)	(l,s)					
Short-term rate	-1.2	1.4					
Long-term rate	-0.9	1.4					
Earnings-by-price	-1.0	1.6					
Inflation	-0.9	1.3					
	(π,s)						
Short-term rate	0.2						
Long-term rate	0.2						
Earnings-by-price	-0.6						
Inflation	-0.1						

3.3 Full benchmarking approach

In the next step, we consider the double benchmarking approach of Kyriakou et al. (2019a,b) to analyze now whether transforming the explanatory variables can improve the predictions for the volatility function. Recall that fully nonparametric models suffer in general by the curse of dimensionality, as in our framework

Table 3: Predictive power for the variance of five-year excess stock returns $Z_t^{(A)}$: the single benchmarking approach. The prediction problem is defined in (5). The same predictive variables X_{t-1} are used in the predictions for the conditional mean and variance function. 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	0.6	-1.7	-1.7	-1.2	-1.0	-2.0	-3.0
Long-term rate	0.0	-1.5	-1.3	-1.2	-1.1	-1.2	-2.7
Earnings-by-price	0.8	-1.8	-1.1	-1.8	-2.7	-0.3	-3.8
Inflation	-1.0	-3.8	-4.7	-0.7	-1.5	1.4	0.5
	$(Y^{(A)},d)$	$(Y^{(A)},e)$	$(Y^{(A)},r)$	$(Y^{(A)},l)$	$(Y^{(A)},\pi)$	$(Y^{(A)},s)$	
Short-term rate	-2.8	-2.5	-1.7	-1.7	-1.7	-3.9	
Long-term rate	-2.5	-2.1	-1.6	-1.8	-1.2	-3.4	
Earnings-by-price	-2.3	-2.1	-1.2	-4.1	0.4	-3.4	
Inflation	-5.1	-4.7	-1.5	-2.6	0.4	-0.9	
	(d,e)	(d,r)	(d,l)	(d,π)	(d,s)		
Short-term rate	-3.6	-3.1	-2.2	-2.8	-4.1		
Long-term rate	-3.1	-3.2	-2.7	-2.3	-4.3		
Earnings-by-price	-4.1	-4.0	-5.3	-2.3	-4.9		
Inflation	-5.2	-5.0	-8.9	-2.5	-3.2		
	(e,r)	(e,l)	(e,π)	(e,s)			
Short-term rate	-3.3	-3.3	-3.5	-4.9			
Long-term rate	-2.8	-3.3	-2.9	-4.9			
Earnings-by-price	-4.5	-5.5	-2.7	-6.5			
Inflation	-8.5	-7.8	-4.9	-6.4			
	(r,l)	(r,π)	(r,s)				
Short-term rate	-3.8	-1.7	-3.9				
Long-term rate	-4.1	-1.3	-4.2				
Earnings-by-price	-5.3	-1.9	-5.4				
Inflation	-3.9	0.3	-1.9				
	(l,π)	(l,s)					
Short-term rate	-1.7	-3.9					
Long-term rate	-1.3	-4.2					
Earnings-by-price	-2.6	-5.4					
Inflation	-1.2	-1.8					
	(π,s)						
Short-term rate	-4.4						
Long-term rate	-3.5						
Earnings-by-price	-4.8						
Inflation	-0.1						

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 presented in Section 3.2 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

$$Y_t^{(A)} = m(X_{t-1}^{(A)}) + \nu(X_{t-1}^{(A)})^{1/2}\zeta_t, \tag{16}$$

$$Z_t^{(A)} = m(X_{t-1}^{(A)}) + \nu(X_{t-1}^{(A)})^{1/2}\omega_t, \tag{17}$$

where we use transformed predictive variables

$$X_{t-1}^{(A)} = \begin{cases} \frac{1+X_{t-1}}{B_{t-1}^{(A)}}, \quad X \in \{d,e,r,l,\pi\} \\ \frac{s_{t-1}}{B_{t-1}^{(A)}} = \frac{l_{t-1}-r_{t-1}}{B_{t-1}^{(A)}} \\ Y_{t-1}^{(A)} \end{cases}, \quad A \in \{R,L,E,C\}. \tag{18}$$

This approach can be interpreted as a way of reducing the 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 comparison to the single-benchmarking approach in the one-year case the double benchmarking improves in 15 out of 82 models (in the sense of producing a positive and higher $R^2_{V,\nu}$ as before). But predictability is still questionable. The best model under the long-term interest rate benchmark $B^{(L)}$ uses the pair $(Y_{t-1}^{(L)}, e_{t-1}^{(L)})$ and yields $R^2_{V,\nu} = 3.0$, while the best model under $B^{(E)}$ uses the pair $(Y_{t-1}^{(E)}, l_{t-1}^{(E)})$ and yields $R^2_{V,\nu} = 2.5$. The SNS-test rejects for both the null of no predictability, while the KNY-test does not. For the rest of the new combinations of predictive variables in all benchmarks, both tests again do not reject.

For the five-year case, we find that in comparison to the single-benchmarking the double benchmarking improves in 11 out of 82 models. The best model under $B^{(E)}$ uses $d_{t-1}^{(E)}$ and yields $R^2_{V,\nu} = 1.8$, while under $B^{(C)}$ the covariates $d_{t-1}^{(C)}$ and $l_{t-1}^{(C)}$ both yield $R^2_{V,\nu} = 1.6$. Nevertheless, we do not find any combination of covariates with statistically significant predictive power.

3.4 Real-income long-term pension prediction

In long-term pension planning or other asset allocation problems optimized with regard to real-income protection (Merton, 2014; Gerrard et al., 2019a,b), the econometric models should reflect those needs and use covariates net-of-inflation. Therefore, we take the inflation benchmark $B^{(C)}$ and analyse in more detail the best model found by Kyriakou et al. (2019b), which uses the earnings-by-price variable for the mean prediction and produced a $R^2_{V,m} = 12.4$ for the one-year horizon and $R^2_{V,m} = 12.2$ for the five-year horizon (see Kyriakou et al., 2019b, Tables 4 and 5) in the double benchmarking case. For this specific model, we are now interested in finding the set of covariates that best predicts the conditional variance.[9,10] The empirical findings in terms of $R^2_{V,\nu}$ are shown for the one-year horizon in Table 6 and the five-year horizon in Table 7. For the one-year horizon we find in the double benchmarking approach when inflation is the benchmark, $B^{(C)}$, that the dividend-by-price $d^{(C)}$ together with the short-term interest-rate $r^{(C)}$ or the long-term interest-rate $l^{(C)}$ are chosen as best predictive variables in terms of $R^2_{V,\nu}$ (2.9% and 2.0%). Note that these values are rather low and that the SNS-test does reject the null of no predictability for both models, while the KNY-test does not reject. For all other combinations and also the five-year case we do not find evidence for statistical significant predictability of the conditional variance. Therefore, we conclude that the constant volatility model is appropriate for practical purposes.

Note further that the ratio in our validation criterion for the mean prediction, $R^2_{V,m}$, in (10) compares the sample variance of the estimated residuals from our model based on earnings-by-price (the numerator) with the sample variance of the benchmarked stock returns (the denominator). For the one-year case, we find from Table 1 the latter to be equal to $0.1805^2 = 0.03258$. A simple calculation using the corresponding $R_{V,m} = 12.4\%$ leads then to $0.03258(1 - 0.124) = 0.02854$ or a standard deviation of 16.89% for returns based on the earnings-model. This means that the linear expression of real stock returns in terms of real earnings-by-price presented in Kyriakou et al. (2019b) as

$$\text{Real one-year stock return} = 0.004875 + 1.119 \times \text{real earnings-by-price} \tag{19}$$

gives on average 2.5% higher returns at the same risk as the historical mean $\bar{Y}^{(C)}$.[11] Similarly, for the five-year case we get from Table 1 that $0.3642^2 = 0.1326$. From the $R_{V,m} = 12.2\%$ we obtain then $0.1326(1 - 0.122) = 0.1164$ or a standard deviation of 34.12% for returns based on the earnings-model. Thus, the linear expression

9 Note that until now we have used the same set of covariates in both steps of our analysis to reduce the overwhelming number of models. It is also clear that not all combinations of variables are practically relevant. Now we relax this restriction for the model with the highest predictive power for the returns.

10 Tables 6 and 7 also present the results for the short- and long-term interest benchmarks $B^{(R)}$ and $B^{(L)}$. But it is again hard to find predictability at all in these cases. Note that the benchmark using the earnings-by-price variable $B^{(E)}$ is not applicable since it matches the covariate and the benchmark in the first step.

11 Here, we use the Sharpe-ratio for the comparison. From Table 1, we get $\bar{Y}^{(C)} = 6.41\%$ and divide it either by 18.05% or by 16.89%. We obtain 0.355 and 0.380, which corresponds to a difference of 2.5% points.

Table 4: Predictive power for the variance of one-year excess stock returns $Y_t^{(A)}$: the double benchmarking approach. The prediction problem is defined in (16). The same predictive variables $X_{t-1}^{(A)}$ are used in the predictions for the conditional mean and variance. The predictive power (%) is measured by $R^2_{V,\nu}$ 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 (18). 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^2_V by construction of the transformed spread according to (18). 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	2.2	-0.3	0.7	–	-0.2	0.1	-0.2
Long-term rate	2.4	0.2	-0.5	-0.1	–	-0.2	-0.1
Earnings-by-price	1.5	-0.2	–	0.6	-0.2	-0.7	0.0
Inflation	0.2	-0.9	-1.2	-0.3	-0.2	–	-0.7
	$(Y^{(A)}, d^{(A)})$	$(Y^{(A)}, e^{(A)})$	$(Y^{(A)}, r^{(A)})$	$(Y^{(A)}, l^{(A)})$	$(Y^{(A)}, \pi^{(A)})$	$(Y^{(A)}, s^{(A)})$	
Short-term rate	0.8	0.7	–	0.2	0.1	0.2	
Long-term rate	1.3	3.0	0.1	–	-0.3	0.1	
Earnings-by-price	0.2	–	0.7	2.5	0.0	0.1	
Inflation	-3.1	-1.4	-1.5	-1.9	–	-1.0	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	-1.3	–	0.9	0.0	0.9		
Long-term rate	-1.0	0.9	–	-0.7	0.9		
Earnings-by-price	–	-0.3	-0.8	-1.8	0.4		
Inflation	-1.9	0.7	1.6	–	-0.7		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	-0.4	-2.6	-0.4			
Long-term rate	-0.6	–	-2.5	-0.6			
Earnings-by-price	–	–	–	–			
Inflation	-1.6	-1.5	–	-1.6			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	-1.2	–				
Earnings-by-price	-0.5	-2.1	-0.3				
Inflation	-1.9	–	-1.6				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	-1.4	–					
Long-term rate	–	–					
Earnings-by-price	-2.5	-0.5					
Inflation	–	-1.7					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	-1.4						
Long-term rate	-1.2						
Earnings-by-price	-1.6						
Inflation	–						

of real stock returns in terms of real earnings-by-price presented in Kyriakou et al. (2019b) as

$$\text{Real five-year stock return} = 0.2068 + 2.264 \times \text{real earnings-by-price} \tag{20}$$

Table 5: Predictive power for the variance of five-year excess stock returns $Z_t^{(A)}$: the double benchmarking approach. The prediction problem is defined in (17). The same predictive variables $X_{t-1}^{(A)}$ are used in the predictions for the conditional mean and variance. 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	0.6	-2.2	-3.2	–	-3.1	-3.2	-3.1
Long-term rate	0.0	-3.4	-2.8	-2.8	–	-1.3	-2.8
Earnings-by-price	0.8	1.8	–	-2.3	-3.2	0.6	-3.8
Inflation	-1.0	1.6	0.3	0.6	1.6	–	0.3
	$(Y^{(A)}, d^{(A)})$	$(Y^{(A)}, e^{(A)})$	$(Y^{(A)}, r^{(A)})$	$(Y^{(A)}, l^{(A)})$	$(Y^{(A)}, \pi^{(A)})$	$(Y^{(A)}, s^{(A)})$	
Short-term rate	-2.1	-4.3	–	-4.0	-1.2	-4.0	
Long-term rate	-3.8	-3.2	-3.6	–	-1.1	-3.6	
Earnings-by-price	1.1	–	-2.8	-3.8	-0.5	-3.4	
Inflation	0.3	-0.8	-0.3	0.4	–	-1.0	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	-3.7	–	-5.4	-2.1	-5.4		
Long-term rate	-4.2	-5.8	–	-3.3	-5.8		
Earnings-by-price	–	-0.4	-2.6	0.3	-3.3		
Inflation	-4.3	-0.2	-0.8	–	-0.8		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	-5.9	-4.9	-5.9			
Long-term rate	-6.1	–	-4.1	-6.1			
Earnings-by-price	–	–	–	–			
Inflation	-4.8	-4.1	–	-2.1			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	-2.3	–				
Earnings-by-price	-6.3	-3.2	-6.1				
Inflation	-1.0	–	0.5				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	-3.4	–					
Long-term rate	–	–					
Earnings-by-price	-3.6	-6.2					
Inflation	–	0.5					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	-3.4						
Long-term rate	-2.3						
Earnings-by-price	-4.6						
Inflation	–						

gives on average 6.0% higher returns at the same risk as the historical mean $\bar{Y}^{(C)}$.[12] Figure 1 shows the estimated nonparametric function $\hat{m}$ (red solid line) for the one-year horizon (left) and the five-year horizon (right) under the double inflation benchmark for the earnings-by-price covariate together with the corresponding historical mean (dashed green line). Figure 2 depicts histograms and a kernel density estimate (red solid line) of the standardized predicted returns for the one-year horizon (left) and the five-year horizon (right). The similarity for both horizons is striking and driven by the fact that the ratio of the slope of the regression lines in (19) and (20) with the corresponding standard deviation given above yields almost the same value of 6.63.

[12] Here, we use again the Sharpe-ratio for the comparison. From Table 1, we get $\bar{Y}^{(C)} = 32.34\%$ and divide it either by 36.42% or by 34.12%. We obtain 0.888 and 0.948, which corresponds to a difference of 6.0% points.

Table 6: Predictive power for the variance of one-year excess stock returns $Y_t^{(A)}$: the double benchmarking approach for the conditional mean model with earnings-by price as single covariate. The prediction problem is defined in (16). The predictive power (%) is measured by $R^2_{V,\nu}$ as defined in (10). The benchmarks $B^{(A)}$ considered are based on the short-term interest rate ($A \equiv R$), long-term interest rate ($A \equiv L$), 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 (18). 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 $R^2_{V,\nu}$ by construction of the transformed spread according to (18). 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	1.0	0.3	0.7	–	0.1	-0.4	0.1
Long-term rate	1.4	0.1	-0.5	0.9	–	-0.1	0.9
Inflation	0.4	-0.6	-1.2	-0.4	-0.1	–	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	0.6	0.7	–	0.2	-0.5	0.2	
Long-term rate	0.7	2.0	0.7	–	-0.6	0.7	
Inflation	-1.7	-1.6	-1.5	-1.7	–	-0.4	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	0.0	–	-0.5	-0.4	-0.5		
Long-term rate	-1.0	0.3	–	-1.4	0.3		
Inflation	-1.9	2.9	2.0	–	1.5		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	0.5	-2.2	0.5			
Long-term rate	-0.7	–	-2.5	-0.7			
Inflation	-0.9	-1.7	–	-0.3			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	-0.4	–				
Inflation	-0.5	–	0.7				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	0.1	–					
Long-term rate	–	–					
Inflation	–	-0.2					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	0.1						
Long-term rate	-0.4						
Inflation	–						

4 Conclusions

In this paper, we extend the original working framework of Kyriakou et al. (2019a,b) of forecasting stock returns to modelling their conditional variance and test for predictability in this context. We consider returns of one-year and five-year horizons 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.

In our analysis, we find only little to no evidence of model-based volatility predictability for the one-year and five-year horizon. We get only for a few of the models considered under different benchmarks validation

Table 7: Predictive power for the variance of five-year excess stock returns $Z_t^{(A)}$: the double benchmarking approach for the conditional mean model with earnings-by price as single covariate. The prediction problem is defined in (17). Additional notes: see Table 6.

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	0.1	-1.8	-3.2	–	-4.5	-2.5	-4.5
Long-term rate	0.6	-3.9	-2.8	-4.2	–	-1.1	-4.2
Inflation	0.0	-0.1	0.3	-0.4	-0.1	–	-2.6
	$(Y^{(A)}, d^{(A)})$	$(Y^{(A)}, e^{(A)})$	$(Y^{(A)}, r^{(A)})$	$(Y^{(A)}, l^{(A)})$	$(Y^{(A)}, \pi^{(A)})$	$(Y^{(A)}, s^{(A)})$	
Short-term rate	-1.7	-4.6	–	-5.7	-3.7	-5.7	
Long-term rate	-4.5	-4.5	-4.2	–	-2.5	-4.2	
Inflation	-1.9	-1.8	-1.9	-1.7	–	-3.9	
	$(d^{(A)}, e^{(A)})$	$(d^{(A)}, r^{(A)})$	$(d^{(A)}, l^{(A)})$	$(d^{(A)}, \pi^{(A)})$	$(d^{(A)}, s^{(A)})$		
Short-term rate	-6.2	–	-7.1	-4.3	-7.1		
Long-term rate	-4.5	-7.9	–	-5.2	-7.9		
Inflation	-3.9	-2.1	-3.2	–	-2.8		
	$(e^{(A)}, r^{(A)})$	$(e^{(A)}, l^{(A)})$	$(e^{(A)}, \pi^{(A)})$	$(e^{(A)}, s^{(A)})$			
Short-term rate	–	-8.1	-5.8	-8.1			
Long-term rate	-6.6	–	-4.9	-6.6			
Inflation	-2.8	-3.4	–	-2.6			
	$(r^{(A)}, l^{(A)})$	$(r^{(A)}, \pi^{(A)})$	$(r^{(A)}, s^{(A)})$				
Short-term rate	–	–	–				
Long-term rate	–	-5.7	–				
Inflation	-3.0	–	-3.1				
	$(l^{(A)}, \pi^{(A)})$	$(l^{(A)}, s^{(A)})$					
Short-term rate	-6.5	–					
Long-term rate	–	–					
Inflation	–	-3.0					
	$(\pi^{(A)}, s^{(A)})$						
Short-term rate	-6.5						
Long-term rate	-5.7						
Inflation	–						

measures that are positive but of a rather small magnitude. The bootstrap test of KNY does not reject the null hypothesis of no predictability at all, while the SNS-test does reject for some variable combinations.

In the practically important double inflation benchmarking case, we conclude that the best model is not only a linear model based on real earnings-by-price but also of constant variance for both horizons. This is an important observation because modelling earnings-by-price in this way is a good and relatively simple starting point when constructing forecasting models for real-value pension prognoses for long-term saving strategies.

Funding: The second 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 fourth author acknowledges financial support from the Research Project APIE 1/2015-17: "New methods for the empirical analysis of financial markets" of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander.

Acknowledgments: The usual disclaimer applies.

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

References

Andersen, T., Bollerslev, T., Diebold, F. and Ebens, H. (2001) The distribution of realized stock return volatility. *Journal of Finacial Economics*, **61**, 43–76.

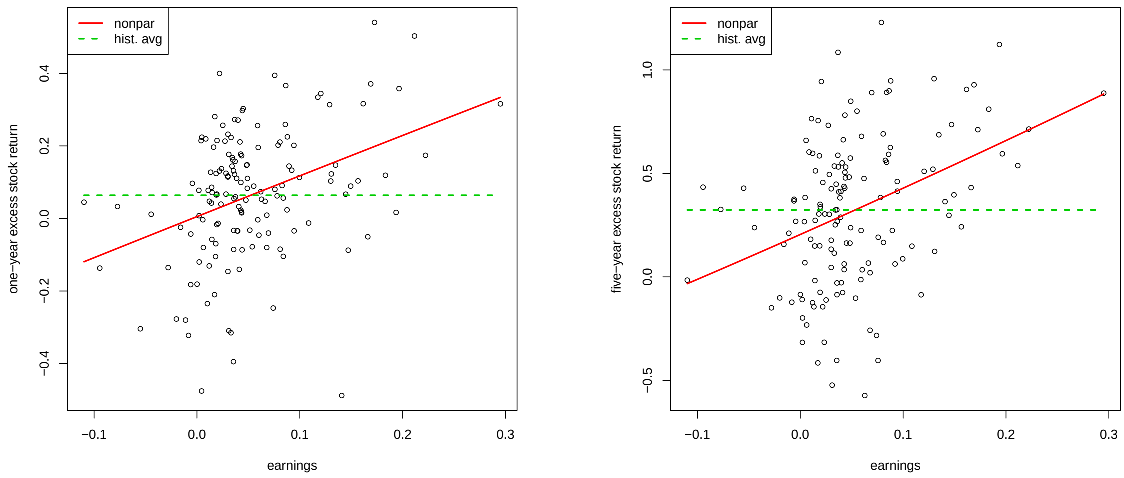

Figure 1: Double inflation benchmark. Relation between real stock returns and real earnings-by-price. Estimated nonparametric function $\hat{m}$ (red solid line) and historical average (dashed green line). Left: one-year horizon. Right: five-year horizon. Period: 1872–2019. Data: annual S&P 500.

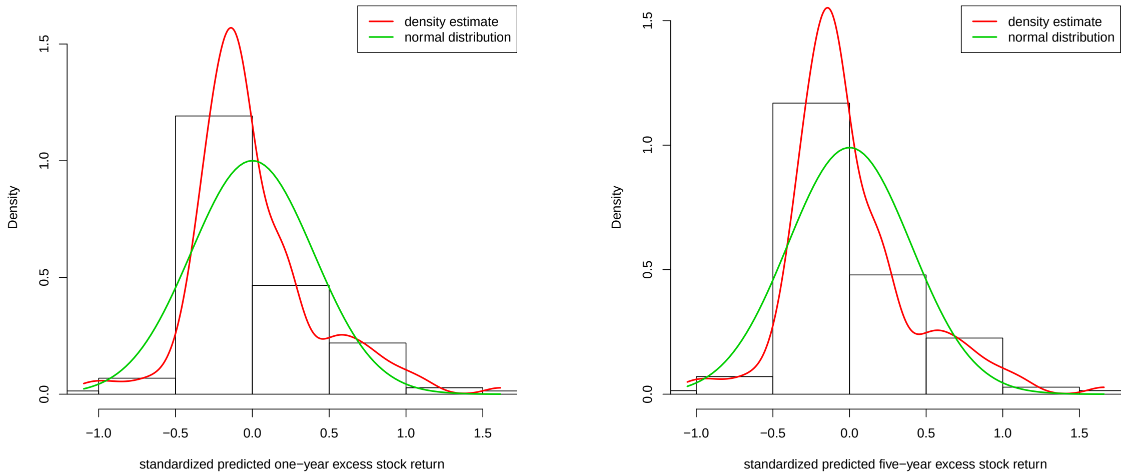

Figure 2: Standardized predicted stock returns in excess of the inflation benchmark (based on the model using earnings-by-price as covariate for mean-prediction; double benchmarking). Histogram, kernel density estimate (red), and fitted normal distribution (green). Left: one-year horizon. Right: five-year horizon. Period: 1872–2019. Data: annual S&P 500.

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 different 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.

Christoffersen, P. and Diebold, F. (2000) How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics*, **82**, 12–22.

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

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.

Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645–660.

Galbraith, J. (2003) Content horizons for univariate time-series forecasts. *International Journal of Forecasting*, **19**, 43–55.

Galbraith, J. and Kisinbay, T. (2005) Content horizons for conditional variance forecasts. *International Journal of Forecasting*, **21**, 249–260.

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**.

Geweke, J. (1996) Monte carlo simulation and numerical integration. In *Handbook of Computational Economics* (eds. H. Amman, D. Kendrick and J. Rust), vol. I, 731–800. Amsterdam: Elsevier.

Glad, I. (1998) Parametrically guided non-parametric regression. *Scandinavian Journal of Statistics*, **25**, 649–668.

Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926–1947.

Härdle, W. and Tsybakov, A. (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, **81**, 223–242.

Kreiss, J., Neumann, M. and Yao, Q. (2008) Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and Its Interfaces*, **1**, 367–380.

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

— (2019b) Machine learning for forecasting excess stock returns – the five-year view. Graz Economics Papers 2019-06, University of Graz, Departmemt of Economics.

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

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

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

Mishra, S., Su, L. and Ullah, A. (2010) Semiparametric estimator of time series conditional variance. *Journal of Business & Economic Statistics*, **28**, 256–274.

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.

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.

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.

Wang, L., Brown, L., Cai, T. and Levine, M. (2008) Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, **36**, 646–664.

Wang, W. and Yu, P. (2017) Asymptotically optimal differenced estimators of error variance in nonparametric regression. *Computational Statistics and Data Analysis*, **105**, 125–143.

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.

Xu, K. and Phillips, P. (2011) Tilted nonparametric estimation of volatility functions with empirical applications. *Journal of Business & Economic Statistics*, **29**, 518–528.

Yu, K. and Jones, M. (2004) Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association*, **99**, 139–144.

Ziegelmann, F. (2002) Nonparametric estimation of volatility functions: The local exponential estimator. *Econometric Theory*, **18**, 985–991.

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

08–2019 **Enno Mammen, Jens Perch Nielsen, Michael Scholz and Stefan Sperlich**: Conditional variance forecasts for long-term stock returns

07–2019 **Christoph Kuzmics, Brian W. Rogers and Xiannong Zhang**: Is Ellsberg behavior evidence of ambiguity aversion?

06–2019 **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