Graz Economics Papers – GEP

GEP 2012–10

# Nonparametric prediction of stock returns with generated bond yields

Michael Scholz, Stefan Sperlich, Jens Perch Nielsen

December 2012

Department of Economics
Department of Public Economics
University of Graz

# Nonparametric prediction of stock returns with generated bond yields

**Abstract**

The question of whether empirical models are able to forecast the equity premium more accurately than the simple historical mean is intensively debated in the financial literature. The low prediction power is disappointing, even when using nonparametric models that make use of typical predictor variables. Motivated by the co-movement of bond and stock returns, the inclusion of the current bond yield in a prediction model is proposed. This results in a notable improvement in the prediction of stock returns, as measured by the validated $R^2$. Since the current bond yield is unknown, it is predicted in a prior step. The essential point is that the inclusion of the generated bond can be seen as a kind of dimension and complexity reduction that imposes more structure in an appropriate way that circumvents the curse of dimensionality and complexity.

**Keywords and Phrases:** Prediction, Stock returns, Bond yield, Cross validation, Generated regressors

**JEL Classification Numbers:** C58, G17, C14

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

Michael.Scholz@uni-graz.at

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

Jens.Nielsen.1@city.ac.uk

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

Stefan.Sperlich@unige.ch

# 1 Introduction and Motivation

For a long time predicting asset returns has been a main objective in the empirical finance literature. It started with simple regressions of independent predictor variables on stock market returns. Typically, valuation ratios are used that primarily characterise the stock, for example the dividend price ratio, dividend yield, earnings price ratio or the book-to-market ratio. Other variables related to the interest rate like treasury-bill rates and long-term bond yield, or macroeconomic indicators like inflation, are often incorporated to improve prediction. For a detailed overview we refer to the examples in Rapach et al. (2005) or Campbell and Thompson (2008).

The apparent predictability found by many authors was controversially discussed. As Lettau and Nieuwerburgh (2008) note, correct inference is problematic due to the high persistence of financial ratios, which have poor out-of sample forecasting power that moreover shows significant instability over time. Therefore, the question of whether empirical models are really able to forecast the equity premium more accurately than the simple historical mean was intensively debated in the finance literature. Recently, Goyal and Welch (2008) fail to provide benefits of predictive variables compared to the historical mean. In contrast, Rapach et al. (2010) recommend a combination of individual forecasts. Their method includes the information provided from different variables and reduces this way the forecast volatility.

A classical approach to valuate the price of a stock is the well-known Gordon growth or dividend discount model which expresses the dividend price ratio, $d_t$, in terms of the long-term discount rate, $R$, and the long-term growth rate of dividends, $G$, both hold constant, $d_t = R - G$. Allowing for time-varying discount rates, for example Campbell and Shiller (1988) introduce a dividend ratio model, where the price of a stock today is seen as the discounted present value of future cash flows to the investor. Additionally, it is well-known that one can find a high correlation of any relevant discount rate to inflation and interest rate. Many authors conclude that a decrease in discount rate is related to an increase in the stock return, and point to the high correlation with an increase in the bond yield.

A direct comparison of stocks and bonds, mostly used by practitioners, makes the so-called FED model. It relates yields on stocks, as ratios of dividends or earnings to stock prices, to yields on bonds. Asness (2003) shows the empirical descriptive power of the model, but notes also that it fails in predicting stock returns. One of his criticisms is the comparison of real numbers to nominal ones. Actually, most studies discuss separately the predictability in stock and bond markets. However, Shiller and Beltratti (1992) analyse the relation between stock prices and changes in long-term bond yields. Engsted and Tanggaard (2001) pose the interesting question of whether expected returns on stocks and bonds are driven by the same information, and to what extent they move together. In their empirical setting, they find that excess stock and bond returns are positively correlated. Engsted and Tanggaard (2001) also note that simple present value models cannot explain this finding. For additional literature on the relation between stock and bond returns see, for example, Guidolin and Timmermann (2006), Connolly et al. (2010), or Baele et al. (2010).

One overall idea of the this paper is to exploit the interrelationship of present values of stock returns and bond returns. They are after all both discounted cash flows. Our underlying assumption implies that expected returns are associated with variables related to longer-term aspects of business conditions, as mentioned in Campbell (1987). Consequently, we include in a nonparametric prediction model[1] of excess stock returns the bond yield of the same year. This way, the bond captures a most important part of the stock return, namely the part related to the change in long-term interest rate. We do this nonparametrically due to the promising findings of Nielsen and Sperlich (2003) which improved significantly the prediction power by nonlinearities and interactions. Local linear kernel regression is used to nest the linear model without bias. For the purpose of bandwidth selection and to measure the quality of prediction we apply the validated $R^2$ of Nielsen and Sperlich (2003) which compares our cross validated model with the cross validated mean.

An obvious problem is that the current bond yield is unknown. Thus, we have to predict it in a first step. This raises the question why it is necessary to use a two-step procedure. One could directly include the variables used for the bond prediction when forecasting stock returns. The problem is that such a model would suffer from the curse of dimensionality and complexity in several aspects: The dimension of the covariates, possible over-fitting, and the interpretability. In nonparametrics it is well known that the import of structure is an appropriate way to circumvent these problems[2]. This article is based on the structure that is inherent in the economic process that generates the data, resulting in the inclusion of bond yields when predicting stock returns. In other words, one may think of the inclusion of predicted bond yields as a kind of dimension (or say, complexity) reduction. Additionally, Park et al. (1997) showed that an appropriate transformation of the predictors can significantly improve nonparametric prediction. Here, we use the additional knowledge about structure to improve the prediction of stock returns. To our knowledge we are the first including nonparametrically generated regressors for nonparametric prediction of time series data. Therefore we also have to develop the theoretical justification for the use of constructed variables in nonparametric regression when the data are dependent.

For the empirical part we use the annual[3] Danish stock market data from Lund and Engsted (1996). This has been done for reasons of comparability and reproducibility as these were also the data used in the above mentioned papers of Engsted and Tanggaard (2001) and Nielsen and Sperlich (2003). It will be shown that the inclusion of predicted bond yields improves greatly

[1] Nonlinear forecasting methods are a growing area of empirical research, see for example McMillan (2007) or Guidolin et al. (2009).

[2] An other possibility could be the optimal choice of regressors, see Vieu (1994).

[3] We focus on the long term actuarial view.

the prediction quality of stock returns in terms of the validated $R^2$. With our best prediction model for one-year stock returns we not only beat the simple historical mean but we also obtain an impressive validated $R^2$ of 28.3 compared to 5.9 from the best model without constructed bonds. We also include in our empirical analysis the prediction of the ratio of stock returns and dividend yields getting similar results.

The paper proceeds as follows. The next section describes the prediction framework: the measure used for quantifying the quality of prediction as well as the model in mathematical terms. In Section 3, the mathematical justification is introduced. Section 4 presents our findings from the empirical study. We also evaluate our proposed method in a small simulation study. While Section 5 concludes.

## 2 The prediction framework

The most used prediction quality measures in the financial and actuarial literature are traditional in-sample approaches like the classic $R^2$, the adjusted $R^2$, goodness-of-fit or testing methods. In various articles the authors construct tests to check whether apparent prediction power is only due to high persistence. In our study, we use the *validated* $R^2(R^2_V)$ which was invented as a measure for out-of sample prediction power. It measures how well the model predicts in the future compared to the sample mean. The classical $R^2$ is often used in empirical finance, because it is easy to calculate and has a straight forward interpretation. But it should be known that it can hardly be used for prediction nor for comparison issues as it always prefers here the most complex model. See also Valkanov (2003) or Dell'Aquila and Ronchetti (2006) for more relevant arguments for disregarding the classical $R^2$ measure when selecting a model. For comparison often the adjusted $R^2$ is applied, which penalises complexity via a degree of freedom adjustment. It is well known that this correction does not work in our case, see for example Sperlich et al. (1999).

When the aim of the analyses is to select a model with predictive power, one is not interested in the in-sample fit of the model that $R^2$ or $R^2_{adj}$ are directed towards. One is interested in the predictive power that $R^2_V$ is designed to catch. The idea of the $R^2_V$ is to replace total variation and not explained variation by their cross validated analogs. Note that cross validation (cv) is a quite common in the nonparametric time series context, see Györfi et al. (1990). More formally, consider the two models

$$Y_t = \mu + \varepsilon_t \qquad \text{and} \qquad Y_t = g(X_t) + \zeta_t,$$

where $\mu$ is estimated by the sample mean $\bar{Y}$ and the unknown function $g$ by local linear kernel regression. We suppress a subscription for the chosen smoothing parameter $h$, since we always apply the bandwidth $h$ that maximises the $R_V^2$. It is defined as

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

where cross validated values $\hat{g}_{-t}$ and $\bar{Y}_{-t}$ are used, i. e. the function $g$ and the mean $\bar{Y}$ are predicted at $t$ without the information contained in this point in time. In fact, it is an out-of-sample measure. We apply the "leave-(2s+1)-out" version of cv, with $s$ depending on the forecast horizon ($= s + 1$). For annual time series this means that for the prediction of stock returns, for example, over the next four years, one has to exclude the $t$-th observation and the three years before and after year $t$, i. e. $s = 3$. The reason is that the information of one year is contained in seven consecutive stock returns, cf. Chu and Marron (1991) or Györfi et al. (1990). A possible generalisation of our approach would be the use of the Do-validation principle to obtain more robust validated $R^2$-values. See Mammen et al. (2011) and Gámiz-Pérez et al. (2012) for some recent improvement of the cross-validation approach. Since maximizing the $R_V^2$ is equivalent to minimize the cross-validation criterion, we use the validated $R^2$ also to find the optimal prediction bandwidth. Notice some basic properties of the $R_V^2$: it is independent of the amount of parameters and takes it values inside $(-\infty, 1]$. As it measures how well a given model and estimation principle predicts compared to the cross validated mean, an $R_V^2 < 0$ indicates that one predicts worse than the cv-mean. In practical prediction it is well known, and confirmed in our empirical study, that it is hard to find models with important explanatory variables which beat even this cv-mean. Nielsen and Sperlich (2003) mentioned that complexity is one of the worst enemies of a good prediction. Therefore, the $R_V^2$ punishes not just under- but also overfitting (pretending a functional relationship that is not real) resulting in $R_V^2 < 0$.

We analyse excess stock returns defined as

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

where $D_t$ denotes the (nominal) dividends paid during year $t$, $P_t$ the (nominal) stock price at the end of year $t$, and $r_t$ the short-term interest rate, which is $r_t = \log(1 + R_t/100)$ using the discount rate $R_t$. For prediction, we consider $Y_t = \sum_{i=0}^{T-1} S_{t+i}$, the excess stock return at time $t$ over the next $T$ years. Based on the motivations from the introduction, we include the same years bond yield as a single regressor or together with further lagged covariates in the model equation, i.e. we consider the model

$$Y_t = g(\hat{b}_t, v_{t-1}) + \varepsilon_t, \tag{2}$$

with the unknown function $g$, the constructed bond yield $\hat{b}_t$, a vector of further regressors $v_{t-1}$ and error terms $\varepsilon_t$, i.e. mean zero variables given the past. As said, the problem which occurs is that the current bond yield is unknown. Therefore, we must predict them in a prior step, i. e. we construct the bond yield with the fully nonparametric model

$$b_t = p(w_{t-1}) + \zeta_t, \tag{3}$$

where $p$ is an unknown function, $w_{t-1}$ is a vector of explanatory variables as for example, lagged interest rates or bond yields, and $\zeta_t$ an error. Both, model (2) and (3), are estimated with a local linear kernel smoother using cross validation. For the choice of the bandwidth, we basically have two possibilities. Either we treat each model separately, determining first the best (in terms of $R_V^2$) bond model and using this in the second step, or we choose the bandwidth in both steps according to the best $R_V^2$ for the stock return prediction.

As discussed, not only economic intuition motivates the inclusion of the constructed bond yields but also statistical arguments. In Theorem 3.7 we will develop the mathematical justification for the use of constructed variables in the case of dependent data. In other words, when we estimate nonparametrically stock returns using a generated regressor, we asymptotically obtain the same function as if we had observed the real bond yield. Basically, the bias of the final estimate is enlarged by an additive factor which is proportional to the bias of the predicted variable from the first step. A similar relationship holds for the variance which is increased by an additive term proportional to the variance of the constructed regressor. This relates the bond to the stock prediction. For simplification ignore for a moment $v_{t-1}$ in (2), and call the function containing the real bonds $\tilde{g}$. A closer look to the prediction error $\varepsilon_t$ gives

$$Y_t - g(\hat{b}_t) = [Y_t - \tilde{g}(b_t)] + [\tilde{g}(b_t) - \tilde{g}(\hat{b}_t)] + [\tilde{g}(\hat{b}_t) - g(\hat{b}_t)] \tag{4}$$

$$\simeq \tilde{\varepsilon}_t + \tilde{g}'(\xi)(b_t - \hat{b}_t). \tag{5}$$

The last term in (4) vanishes as we will see in Theorem 3.7 and the second term can be easily approximated. The gain in our two-step procedure comes now from the fact that the second term in (5) is quite predictable, as we confirm in the empirical part 4.2, especially documented in Table 2. An other idea would be the following: first, estimate $g$ with the available bond data $b_{t-1}$, and second, evaluate $\hat{g}$ at the constructed $\hat{b}_t$. Since, however, this procedure did not improve the stock forecasts, we skip it from further considerations.

One could directly use the variables in the vector $w_{t-1}$ as regressors in model (2). But the model would suffer from complexity and dimensionality in several aspects: The dimension of the covariates as well as their interplay. In the nonparametric literature, typically two strategies are proposed to circumvent this; either semiparametric modeling or additivity, both to import structure. Nielsen and Sperlich (2003) showed that additive models fail to improve the prediction of stock returns due to a non-ignorable interaction between the predictors. We believe in getting better results by introducing additional structure which is inherit by the underlying financial process. We think of the same years bond yield as an important factor which captures some of the relevant features for the expected stock returns. Then, the inclusion of bond yields when predicting stock returns nonparametrically acts as a kind of complexity and dimension reduction due to the import of more structure. To see if it is possible to further improve the predictive power in our setting, we will also analyse the model (2) with a different dependent variable. We consider the ratio between current stock returns and dividend yield, i.e. $Y_t^* = Y_t/d_t$, and the ratio of current stock returns and long-term interest rate, $L_t$, $Y_t^{**} = Y_t/L_t$ or risk-free rate, $r_t$, $Y_t^{***} = Y_t/r_t$; see Section 4.

## 3 Mathematical justification

We prove the consistency of a function estimate which makes use of constructed variables and derive its asymptotic properties. For the prediction in the time series context, we follow the steps from Ferraty et al. (2001) and combine them with Sperlich (2009)[4]. Let us consider a sample of real random variables $\{(X_i, Y_i), i = 1, \ldots, n\}$ which are not necessarily independent and want to estimate the unknown function $m(x) = E(Y|X = x)$, $x \in \mathbb{R}$, that should always exist. Note that for time series $\{(Z_i), i \in \mathbb{N}\}$ a $k$-step ahead forecast is included in a natural way setting $Y_i = Z_{i+k}$ and $X_i = Z_i$. We concentrate only on the case of an auto-regression function of order one. Since we face constructed realisations for $X$, we assume a predictor[5] with an additive bias and a stochastic error:

$$\hat{x} = x + b(x) + u(x) \tag{6}$$

uniformly, where $u(x) := u \cdot \sigma_u(x)$. The independent random variables $u$ are normalised versions of $u(x)$. The conditional variance at point $x$ is $\sigma_u(x)$. This is rather general as it holds for almost all common predictors. For technical reasons, we further assume finite higher moments for $u$. Then, for example the Nadaraya-Watson estimator is

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

$$\text{with } \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right) \quad \text{and} \quad \hat{q}(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{\hat{X}_i - x}{h}\right)$$

where $K$ denotes some kernel function with bandwidth $h$.

[4]For more technical details of the proofs we refer to the appendix.

[5]We don't specify a particular one but we will need some assumptions on it. The used sample of size $N$, for instance, consists of some instruments $\boldsymbol{Z} \in \mathbb{R}^{\delta}$. In the following, we have N=n, since we use the same series in both the prediction and final step.

To measure the strength of dependence in the time series, we limit us to the *strong-* or *$\alpha$-mixing*[6] defined in Doukhan (1994) or Fan and Yao (2003) as $\lim_{n\to\infty} \alpha(n) = 0$, for the *mixing coefficient*

$$\alpha(n) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{n}^{\infty}} |P(A)P(B) - P(AB)|,$$

where $\mathcal{F}_i^j$ is the $\sigma$-algebra generated by $\{X_k, i \leq k \leq j\}$. We further assume that the sequence $\{(X_i, Y_i), i = 1, \ldots, n\}$ is *algebraic* $\alpha$-mixing, i.e. that for some real constants $a, c > 0$ we have $\alpha(n) \leq cn^{-a}$. To get the asymptotic properties in the context of strong mixing, we make use of an exponential inequality of the Fuk-Nagaev type, cf. Rio (2000).

**Lemma 3.1.** *For an algebraic $\alpha$-mixing sequence of random variables $\{(Z_i), i \in \mathbb{N}\}$, with $s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\mathrm{cov}(Z_i, Z_j)|$ and $||Z_i||_\infty < \infty$ for all $i$, holds for some $\varepsilon > 0$ and $r > 1$*

$$P\left(\left|\sum_{i=1}^n Z_i\right| > 4\varepsilon\right) \leq 4\left(1 + \frac{\varepsilon^2}{rs_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{a+1}.$$

Furthermore, we need the Billingsley inequality from Bosq (1998) to bound from above the covariance of two elements of a strong-mixing time series.

**Lemma 3.2.** *For an $\alpha$-mixing sequence of random variables $\{(Z_i), i \in \mathbb{N}\}$, with $||Z_i||_\infty < \infty$ for all $i \neq j$, holds $|\mathrm{cov}(Z_i, Z_j)| \leq 4||Z_i||_\infty ||Z_j||_\infty \alpha(|i - j|)$ .*

To prove the asymptotic behaviour of the kernel regression estimator (7), we make some common assumptions. As noted above, we analyse an algebraic $\alpha$-mixing sequence of real random variables $\{(X_i, Y_i), i = 1, \ldots, n\}$. We suppose that for all $i \neq j$ the joint density $f_{ij}$ for the pair $(X_i, X_j)$ exists and that $|Y| < C < \infty$ almost surely.

Also for the unobservables $X_i$, we assume a density function $f_X$ which is bounded and has a continuous second derivative. At the fix $x \in \mathbb{R}$ we suppose $f_X(x) > 0$.

Let the kernel $K$ be integrable, bounded, with compact support and continuous second derivative. It fulfils $\int K(s)ds = 1$ and $\int sK(s)ds = \int K'(s)ds = \int K''(s)ds = 0$.

For both, the deterministic and the stochastic part of the predicted realisations $\hat{x}$ in (6), we assume that $b(x)$ and $b'(x)$ are at least of order $O(h_0^2)$ uniformly, and $\sigma_u^2(x)$ of order $O((nh_0^\delta)^{-1})$. Here, $h_0$ is a smoothing parameter tending to zero when the sample size $n$ goes to infinity and $\delta$ refers to the dimension of the used instruments in the prediction step. Let further be $b(\cdot)$ and $\sigma_u(\cdot)$ Lipschitz-continuous.

To simplify our calculations, we further suppose that $h_0^2 h^{-1}$ and $(nh_0^\delta h)^{-1}$ go to zero, and use the usual assumption that $nh$ and $nh_0^\delta$ go to infinity as $n \to \infty$.

[6]The weakest of the usually defined mixing conditions.

Before we state the main result of the section, we collect some important facts. First, we define the following variables for $l \in \{0,1\}$

$$Z_i = Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right) - E\left[Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right] \tag{8}$$

and analyse the asymptotic behaviour of

$$s_n^{2*} = \sum_{i=1}^{n} \sum_{j \neq i} |\text{cov}(Z_i, Z_j)|.$$

**Proposition 3.3.** *Under the above assumptions holds $s_n^{2*} = o(nh) + O(n^2\alpha(\tilde{\Delta}))$, where $\tilde{\Delta}$ has the same order like the slowest from $\left\{\frac{1}{h\log n}, \frac{nh_0^\delta h}{\log n}, \frac{(nh_0^\delta h)^2 h}{\log n}\right\}$.*

When $(nh_0^\delta)^{-1} = O(h^2)$ the above proposition reduces to $s_n^{2*} = o(nh) + O(n^2\alpha((h\log n)^{-1}))$ as in the case without any prior prediction, i.e. $b(x) = \sigma_u(x) = 0$.

**Proposition 3.4.** *Under the given assumptions and if exists an $\varepsilon > 0$ such that $\Delta^{a-1} = O(n^{-1-\varepsilon})$, with $\Delta$ from $\{h, (nh_0^\delta h)^{-1}, ((nh_0^\delta)^2 h^3)^{-1}\}$, it holds that $s_n^{2*} = o(n\Delta)$.*

**Proposition 3.5.** *Under the assumptions of Proposition 3.4 we have* $\text{var}(Z_1) = O(\Delta)$ .

With $\Delta$ from Proposition 3.4 we can direct conclude

$$s_n^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |\text{cov}(Z_i, Z_j)| = n \cdot \text{var}(Z_1) + s_n^{2*} = O(n\Delta). \tag{9}$$

Before we specify the result about the convergence of estimator (7) we need

**Proposition 3.6.** *Under the assumptions and $\Delta$ from Proposition 3.4, verifying*

$$c_1 n^{\frac{3-a}{a+1}+\theta} \leq \Delta \leq c_2 n^{\frac{1}{1-a}-\theta}, \tag{10}$$

*with existing $c_1,\ c_2,\ \theta > 0$, it holds for $\nu$ and $\varepsilon > 0$ with $\psi = g$ or $\psi = f$*

$$P\left(\left|E\hat{\psi}(x) - \hat{\psi}(x)\right| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\right) = O(n^{-1-\nu}). \tag{11}$$

Now we can state the main theorem. For continuous (around $x$) functions $m$ and $f$ we get the quasi complete convergence[7] of Nadaraya-Watson estimators with constructed regressors.

[7]For a sequence of real random variables $X_n$ exists a real random variable $X$ such that for all $\varepsilon > 0$ holds $\sum_{i=1}^{\infty} P(|X_i - X| > \varepsilon) < \infty$, cf. Serfling (1980).

**Theorem 3.7.** *Under the above assumptions and* (10)*, it holds quasi completely that* $|\hat{m}_{NW}(x) - m(x)| \longrightarrow 0$.

The extension to the local linear estimator is almost straightforward. For $j = 0, 1, 2$

$$s_j(x) = \sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_i - x)^j \quad \text{and} \quad t_j(x) = \sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_i - x)^j Y_i,$$

we can define $\hat{m}_{LL}(x) := (t_0(x)s_2(x) - t_1(x)s_1(x))/(s_0(x)s_2(x) - s_1^2(x))$ what leads to

$$\hat{m}_{LL}(x) = \sum_{i=1}^{n} C\left(\frac{\hat{X}_i - x}{h}\right) Y_i \bigg/ \sum_{i=1}^{n} C\left(\frac{\hat{X}_i - x}{h}\right), \tag{12}$$

with $C\left(\frac{\hat{X}_i - x}{h}\right) = \sum_{j \neq i} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_j - \hat{X}_i) K\left(\frac{\hat{X}_j - x}{h}\right)(\hat{X}_j - x)$ as a discretisation of $C(u) = \int K(u-v)vK(u)udu$[8]. Since equation (12) is of the same form like (7) and the kernel $C$ fulfils the same conditions as $K$, the application of Theorem 3.7 yields

**Corollary 3.8.** *Under the assumptions of Theorem 3.7, it holds quasi completely that* $|\hat{m}_{LL}(x) - m(x)| \longrightarrow 0$.

For mean square convergence, asymptotic normality and higher order polynomials, one could directly extend the work of Masry and Fan (1997) to the case of predicted regressors.

# 4 Empirical evidence and simulation studies

We interpret our method presented as a two stage regression approach. Based on the idea that the bond of the same year captures an important part of the stock return we search in the first step the optimal prediction model for the bond. Afterwards, as we have seen in Theorem 3.7, we can consistently predict stock returns using the predicted bond yields.

## 4.1 Data description

Consider the annual Danish stock and bond market data for the period 1923 – 1996 from Lund and Engsted (1996). In the appendix of their work, a detailed description of the data can be found. We use a stock index based on a value weighted portfolio of individual stocks chosen to obtain maximum coverage of the market index of the Copenhagen Stock Exchange (CSE). Notice that the CSE was open during the second world war. When constructing the data, corrections were made for stock splits and new equity issues below market prices. Table 1 presents summary statistics of the available variables. In the following, we use the dividend price ratio, $d$, the stock

[8] Note that $C$ is a bimodal kernel. Since it puts more weight to points close to $x$, except if they are too close, than to points far from $x$, it is a natural choice in the case of strong mixing data, see Kim et al. (2009).

**Table 1:** Danish stock and bond market data (1923-1996)

| | Min | Max | Mean | Sd |
|---|---|---|---|---|
| CSE Stock Price Index | 64.78 | 3177.88 | 511.07 | 662.99 |
| Dividend Accruing to Index | 2.99 | 44.76 | 15.11 | 11.81 |
| Excess Stock Returns | -42.44 | 72.10 | 2.10 | 17.19 |
| Bond Yield | -13.70 | 60.30 | 8.61 | 12.07 |
| Dividend Yield | 0.01 | 0.08 | 0.04 | 0.01 |
| Long-term Interest Rate | 3.80 | 19.45 | 8.24 | 4.29 |
| Short-term Interest Rate | 2.50 | 17.86 | 6.96 | 3.46 |

**Table 2:** $R_V^2$-values (in percent) for bond model (3)

| $w_{t-1}$ | $S$ | $L$ | $r$ | $d, L$ | $d, r$ | $S, L$ | $S, r$ | $L, b$ | $r, b$ | $S, L, b$ | $S, r, b$ | $L, r, b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| par. | 11.6 | 24.0 | 22.3 | 21.9 | 19.4 | 31.9 | 33.1 | 29.2 | 35.2 | 31.9 | 37.4 | 30.9 |
| nonpar. | 16.3 | 24.0 | 26.8 | 23.2 | 19.4 | 31.9 | 33.1 | 29.2 | 35.5 | 31.9 | 37.4 | 31.8 |

return, $S$, the long-term interest rate, $L$, the short-term interest rate, $r$, and the bond yields, $b$, as explanatory variables.

### 4.2 The prior step: A simple bond yield predictor

We speak of a simple predictor as in the literature quite complex models can be found for this problem. Our main target, however, are the stock returns where bond yield prediction is just an auxiliary step in order to reduce complexity and dimension. Therefore, the model and bandwidth selection for (3) has to be based on the objective of maximising the $R_V^2$ of the stock return problem (2). Recognising that the model that maximises the $R_V^2$ for bond prediction is not necessarily the one that maximises the $R_V^2$ for stock returns, it becomes clear that it is worth to consider nonparametric alternatives for (3), even if parametric models seem to do a very good job for bond yield prediction alone. This is the reason why we need Theorem 3.7; for parametric predictors $\hat{x}$ the consistency of (7) follows trivially.

If we just look at the bond yield prediction, then we get positive $R_V^2$ for the models listed in Table 2. We observe that only in few cases a local linear predictor does a better job than a linear model as far as we look at the $R_V^2$ for bond yields. The interesting numbers, however, we will see only when looking at the $R_V^2$ for stock returns in Table 3, next section.

### 4.3 Stock prediction

For sake of brevity we concentrate on a forecast horizon of $T = 1$, i.e. $Y_t = S_t$. For larger horizons we got analog results but without further insight. First, we estimate model (2) without the constructed bond variable again with a simple parametric regression and a fully nonparametric kernel based method. The results are summarised in the first two lines of Table 3 and show that all parametric models produce negative validated $R_V^2$ values. It means that with a simple regression approach we cannot better forecast one-year stock returns than the simple mean. A more sophisticated technique is needed. In fact, our so far best nonparametric model[9] uses lagged bond yields, $b_{t-1}$, and gives an $R_V^2$ of 5.9.

**Table 3:** $R_V^2$-values (in percent) for stock model (2)

| $w_{t-1}$ | $d$ | $S$ | $L$ | $r$ | $b$ | $d,S$ | $d,L$ | $d,r$ | $d,b$ | $S,L$ | $S,r$ | $S,b$ | $L,r$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| nonpar | -1.4 | 1.8 | -4.2 | -3.6 | 5.9 | 5.5 | -6.0 | -7.4 | 3.1 | -3.5 | -7.1 | 4.6 | -9.4 |
| par | -1.3 | -1.8 | -4.2 | -5.7 | -4.0 | -3.5 | -5.8 | -7.2 | -6.2 | -6.8 | -7.9 | -6.6 | -9.3 |
| $\hat{b}_t$ | 10.0 | 1.3 | -3.5 | 1.4 | 10.6 | -1.5 | -3.8 | 2.9 | 10.1 | -1.1 | -3.1 | 3.9 | -0.6 |
| $\hat{b}_t$ par | 10.0 | -2.6 | -4.0 | -4.2 | -11.0 | -4.2 | -3.8 | -4.1 | -3.5 | -3.7 | -3.9 | 3.7 | -3.7 |
| $\hat{b}_t, v_{t-1}$ | 16.4 | 5.1 | 9.1 | 16.3 | 8.9 | -1.6 | 28.3 | 21.6 | 10.2 | 1.6 | 13.5 | -1.3 | 15.8 |
| $\hat{b}_t, v_{t-1}$par | 16.4 | -23.7 | 3.3 | -0.7 | 5.1 | -25.2 | 12.0 | 17.3 | 10.2 | -1.5 | 7.9 | -13.4 | -6.1 |

| $L,b$ | $r,b$ | $d,S,L$ | $d,S,r$ | $d,S,b$ | $d,L,r$ | $d,L,b$ | $d,r,b$ | $S,L,r$ | $S,L,b$ | $L,r,b$ | $S,r,b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.8 | 0.5 | -2.9 | -6.7 | 3.3 | -11.2 | -3.8 | 1.0 | -11.0 | 0.3 | -4.4 | -1.6 |
| -7.5 | -8.6 | -8.6 | -9.8 | -8.8 | -10.9 | -9.9 | -11.2 | -12.5 | -11.1 | -13.0 | -11.9 |
| -0.9 | -3.6 | 1.3 | -3.5 | 8.9 | -2.8 | -1.0 | -3.7 | -2.1 | 1.8 | -3.6 | 1.6 |
| -3.3 | -3.5 | -3.3 | -3.8 | -4.7 | -3.4 | -3.0 | -3.4 | -3.9 | -3.6 | -3.4 | -3.5 |
| 15.6 | 20.3 | 10.8 | 14.2 | 0.8 | 17.5 | 16.6 | 20.4 | 10.0 | 1.6 | 15.9 | 7.6 |
| 7.7 | 12.3 | -1.2 | 5.7 | -15.5 | -2.5 | 11.4 | 18.0 | -1.7 | -3.6 | -8.0 | 7.6 |

Nonparametric and parametric model without constructed bond (first and second line); with $\hat{b}_t$ as unique regressor (third and fourth line); with $\hat{b}_t$ and the same variables as in the first prediction step (fifth and sixth line). Bandwidth choice in the final step.

Second, we follow our in Section 2 proposed procedure and generate the current bond yield with model (3) to include this as a regressor in the final step (2). Let us do this first without any further regressor $v_{t-1}$. As discussed before, we have to choose the model and bandwidths along the largest $R_V^2$ value for predicting stock returns[10]. How much the predictive power has increased by this method can be seen when comparing line one with line three of Table 3. The best model again uses only bond yields in both steps and has an $R_V^2$ value of 10.6. In line four we added the results of the parametric counterpart. As one can clearly see, the nonparametric version produces better results, recall our discussion in the last section.

Third, we construct the current bond as before but accompany this regressor in model (2) by any combination of lagged variables from the predictor set $\{d, S, L, r, b\}$ as our vector $v_{t-1}$. Then, the two largest $R_V^2$ were achieved by $\hat{g}(\hat{b}_t, d_{t-1}, S_{t-1}, L_{t-1})$ where $\hat{b}_t = \hat{p}(d_{t-1})$ (yielding $R_V^2 = 30.3$) or $\hat{b}_t = \hat{p}(d_{t-1}, L_{t-1})$ (yielding $R_V^2 = 28.9$), respectively. Note that for an

[9] Nielsen and Sperlich (2003) report in an analog setting a $R_V^2$ value of 5.5 for a fully nonparametric two-dimensional model with dividend-price ratio, $d_{t-1}$, and lagged excess stock returns, $S_{t-1}$, as explanatory variables, but don't use bond yields in their analysis.

[10] If one chooses for example the bandwidth that predicts best bond yields in the prior step, then the values in Table 3 will shrink or remain the same. The best prediction model will still be $g(\hat{b}_t)$ but with an $R_V^2$ value of 9.0 instead of 10.6.

increasing set of regressor variables the corresponding multidimensional bandwidth grid on which we looked for the best predicting one had to be reduced for numerical reasons. Consequently, lower dimensional models have the tendency to be slightly favoured in our study. The full set of results for the 25 times 25 combinations of $\{d, S, L, r, b\}$ is not shown for the sake of presentation, but available on request. An interest finding is that for each variable set the 'diagonal' of all results, i.e. where $w_{t-1} = v_{t-1}$ and given in the fifth line of Table 3, seems to be among the best prediction models. We see now clearly that our proposal greatly improves the predictive power for stock returns. Line 6 shows the results of the parametric counterpart. We find again convincing evidence that the two nonparametric steps are better than the parametric counterpart. For the best model in Table 3 – we have $w_{t-1} = v_{t-1} = (d_{t-1}, L_{t-1})$ – we find an impressive $R_V^2$ value of 28.3, an about factor five improvement compared to the model without constructed bonds. This finding again indicates that the bond captures and provides a quite important part of the stock return which is related to the change in long term interest rate.

The last part of our empirical study concentrates on the change of the dependent variable. Up to now, we used the excess stock return but for the following we divide this value by the dividend yield, i.e. we use $Y_t^* = Y_t/d_t$. Table 4 summarises our findings for $Y^*$.

**Table 4:** $R_V^2$-values (in percent) for model (2) with $Y_t^* = Y_t/d_t$ as dependent variable.

| $w_{t-1}$ | $d, L$ | $d, r$ | $L, r$ | $L, b$ | $r, b$ | $d, L, r$ | $d, r, b$ | $L, r, b$ |
|---|---|---|---|---|---|---|---|---|
| par. | -8.1 | -10.4 | -11.2 | -4.5 | -6.8 | -15.0 | -11.2 | -12.7 |
| nonpar. | -8.3 | -10.5 | -11.3 | 1.3 | 11.4 | -15.2 | 2.4 | 12.3 |
| final with $v = w$ | 39.9 | 35.4 | 41.4 | 31.3 | 35.8 | 39.5 | 42.6 | 43.5 |

First line: fully parametric model, second line: nonparametric estimation without constructed bond, third line: nonparametric with constructed bond and bandwidth selection in the final step.

The first line refers again to the parametric version of model (2) and the second line to the fully nonparametric method, both without constructed bonds. Almost all of the parametric models have negative $R_V^2$ values and also only a small number of nonparametric models beat the simple mean. In contrast, when we include the constructed bond in the nonparametric prediction, a large increase of the validated $R_V^2$ can be observed. For example, the model which uses long- and short-term interest rate, and lagged bond yields for both the bond generation and following stock prediction, has a $R_V^2$ value (43.5%) that is over three and a half times larger than the value of the best model without constructed bonds (12.3%).

### 4.4 Simulation studies

A simulation study gives us the possibility to highlight the potential of our method. We first show the effects of a dimension reduction and afterwards of a pronounced curvature.

**Table 5:** $R_V^2$-values (in percent) for dimension reduction.

| $\sigma$ | 0.3 | 0.6 | 0.9 | 0.3 | 0.6 | 0.9 |
|---|---|---|---|---|---|---|
| $s_1$ | 57.3 (10.4) | 23.7 (12.7) | 10.5 (11.5) | 49.2 (6.2) | 19.2 (5.6) | 9.2 (4.4) |
| $s_2$ | 88.1 (3.5) | 66.4 (8.7) | 47.0 (11.8) | 85.2 (2.9) | 59.6 (5.9) | 39.7 (6.7) |
| 2 step method | 89.1 (3.3) | 68.1 (8.5) | 47.9 (11.7) | 86.7 (2.4) | 62.2 (5.7) | 42.0 (6.6) |
| fully nonpar. | 45.8 (11.7) | 35.0 (12.9) | 24.3 (13.5) | 47.6 (6.6) | 34.2 (6.5) | 23.0 (6.1) |

Note: For simplicity's sake we use $\sigma_{s_1} = \sigma_{s_2} = \sigma_m$ and refer to it as $\sigma$. Averaged values over 500 simulation runs with standard errors in brackets. Left panel: n=50, right panel: n=200.

Let us consider a four dimensional function that is separable into two terms: $m(x_1, \ldots, x_4) = \tilde{m}(s_1, s_2)$ with $s_1 = s_1(x_1, x_2)$ and $s_2 = s_2(x_3, x_4)$. We simulated data from the following models: $S_1 = x_1 + x_1 x_2 + \varepsilon_{\sigma_{s_1}}$, $S_2 = \exp(x_3 + x_4) + \varepsilon_{\sigma_{s_2}}$, and $Y = m(x_1, \ldots, x_4) + \varepsilon_{\sigma_m} = \tilde{m}(s_1, s_2) + \varepsilon_{\sigma_m} = s_1 + s_2 + \varepsilon_{\sigma_m}$. For each explanatory the support is $[0, 1]$. An autoregressive design with $\phi = 0.75, 0.2, 0.02$ for $x_1, \ldots, x_3$ was used; also a normal for $x_4$. Different parameter values $\sigma$ for the zero mean normal error distributions were investigated as well as different sample sizes $n$. The kernel used was the Gaussian. For computational reasons the bandwidths are chosen separately in each step of the simulation[11]. In step one we predict $s_1$ and $s_2$, used in step 2 to estimate function $\tilde{m}$.

Lines three and four of Table 5 present the results for the two-step approach for $\tilde{m}$ and the fully nonparametric method estimating $m$ in terms of $R_V^2$ values, averaged over 500 runs. The proposed two-step procedure succeed in improving on the fully nonparametric estimator in all cases by far. The effect of the dimension reduction is of course more pronounced for the smaller sample size and results in an almost factor 2 improvement.

For the second part we consider the function composition $m(x) = \tilde{m} \circ s(x)$, where the inner function $s$ has a pronounced curvature. We simulated data from the following models: $S = \sin(4\pi(x - 1/8)) + \cos(4/3 \cdot \pi(x - 1/2)) + 1.6 + \varepsilon_{\sigma_s}$ and $Y = m(x) + \varepsilon_{\sigma_m} = \sin \circ s(x) + \varepsilon_{\sigma_m}$, i. e. $\tilde{m}(x) = \sin(x)$. Note that $s$ is one of the example functions used in Park et al. (1997). A uniform design was used with the support $[0, 1]$. Different parameter values $\sigma$ for the zero mean normal error distributions were investigated for a sample size of $n = 50$. The kernel used was the Gaussian and the bandwidths are chosen separately in each step for the two-step part. Again we are aware of the suboptimality, i.e. we could even do better with respect to the $R_V^2$ but at the cost of computing time. Table 6 reports the results. We find that already in this very simple example the proposed two-step approach can help to obtain clearly better results, i. e. much larger $R_V^2$-values in all cases. Figure 1 shows the used inner function (left) and estimates of $m$ and $\tilde{m}$ (right). We see that our method can better estimate problematic regions, in particular

[11] Note that this is again suboptimal, i.e. a bandwidth choice along the final objective would give even better results for our method than those presented in Table 5.

**Table 6:** $R_V^2$-values (in percent) for pronounced curvature.

| $\sigma_m$ | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|
| $s$ | 95.7 | 95.6. | 95.7 | 85.8 | 85.5 | 85.9 | 69.8 | 70.8 | 70.2 |
| | (2.1) | (2.3) | (2.0) | (4.5) | (4.4) | (3.7) | (8.1) | (7.0) | (7.1) |
| 2 step method | 86.9 | 51.0 | 26.9 | 82.4 | 48.4 | 25.0 | 73.7 | 43.7 | 23.5 |
| | (5.1) | (10.8) | (11.6) | (6.9) | (10.8) | (11.6) | (11.4) | (11.6) | (12.1) |
| fully nonpar. | 72.6 | 38.3 | 16.1 | 72.6 | 39.3 | 16.2 | 72.4 | 39.2 | 17.0 |
| | (8.7) | (12.5) | (12.4) | (9.1) | (11.8) | (12.4) | (10.3) | (12.2) | (12.1) |

Note: Averaged values over 500 simulation runs with standard errors in brackets. Sample size: n=50. Left panel: $\sigma_s = 0.1$, middle panel: $\sigma_s = 0.3$, right panel: $\sigma_s = 0.5$.

by bias reduction.

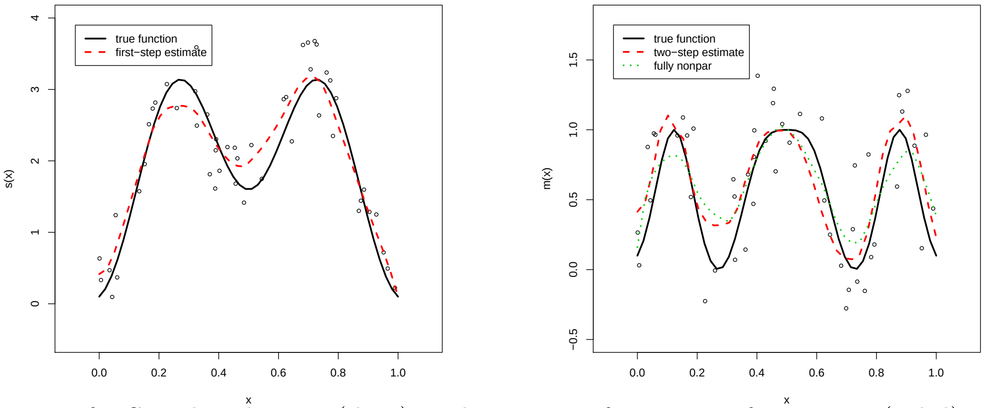


**Figure 1:** Left: Simulated Data (dots) with $\sigma_s = 0.3$ from inner function $s$ (solid), first-step estimate (dashed); Right: Simulated Data (dots) with $\sigma_m = 0.3$ from function composition $m$ (solid), two-step estimate (dashed), usual nonparametric estimate (dotted)

## 5 Concluding remarks and outlook

Motivated by economic theory and statistical arguments, we include the same years bond yield in the fully nonparametric prediction approach for excess stock returns. Since the current bond yield is unknown, we propose to construct it in a prior step using again nonparametric techniques. The bandwidths should be chosen in such a way that they maximise the $R_V^2$ of the final step. The empirical study demonstrates that this two-step approach can improve the stock return prediction enormously. We moreover prove the consistency of our method and derive the asymptotic behaviour of our final predictor. We illustrate the improvement due to our method using annual Danish stock and bond market data which were studied in detail in former articles by different authors. Our results confirm our motivation of including the same years bond yield, namely that it captures the most important part of the stock return, that one related to the change in long-term interest rate. This actually holds not only for stock returns but also for transformed variables, as for example returns divided by dividend yields.

The statistically insights are the following. It is clear that we face a regression model that exhibits high complexity and dimensionality. An obvious remedy would be the imposing of structure. Since it has been shown that additive separability is inappropriate because of unknown interactions, we make use of financial theory to exploit the inherit structure of stock returns. Alternatively, one could interpret the first stage as an optimal nonparametric transformation that maps, for example, the long-term interest rate to the current bond yield, $L_{t-1} \to \hat{b}_t$. The subsequent nonparametric smoother of the transformed variable is than characterised by less bias. Here, we present a practical example in the spirit of the somewhat theoretical method proposed by Park et al. (1997) which improves nonparametric regression with transformation techniques. Although we extend their method in several aspects, their paper provides some statistical intuition for the success of our approach. Our simulations additionally underpin the key idea of complexity and dimension reduction.

# A Proofs

*Proof of Proposition 3.3.* Since the variables $Z_i$ in (8) are centered, we calculate for $i \neq j$

$$\begin{aligned} |EZ_iZ_j| = \Big| & EY_i^l Y_j^l K\Big(\frac{\hat{X}_i - x}{h}\Big) K\Big(\frac{\hat{X}_j - x}{h}\Big) \\ & - EY_i^l K\Big(\frac{\hat{X}_i - x}{h}\Big) EY_j^l K\Big(\frac{\hat{X}_j - x}{h}\Big)\Big|. \end{aligned} \tag{13}$$

First, we analyse the second term in the last equation and use the assumption that all $Y_i$ are bounded.

$$EY_i^l K\Big(\frac{\hat{X}_i - x}{h}\Big) \leq C \int\int K\Big(\frac{u - x + b(u) + v\sigma(u)}{h}\Big) f(u,v)dudv.$$

A simple Taylor-expansion of the kernel leads to

$$\begin{aligned} C \int\int \{ & K\Big(\frac{u-x}{h}\Big) + K'\Big(\frac{u-x}{h}\Big)\Big(\frac{b(u)+v\sigma(u)}{h}\Big) + \\ K''\Big(\frac{u-x}{h} & + \kappa\frac{b(u)+v\sigma(u)}{h}\Big)\frac{(b(u)+v\sigma(u))^2}{2h^2}\Big\} f(u,v)dudv, \end{aligned}$$

where $\kappa \in (0,1)$. With the common substitution $s = (u-x)h^{-1}$ we get

$$EY_i^l K\Big(\frac{\hat{X}_i - x}{h}\Big) = O(h + (nh_0^\delta h)^{-1}).$$

Analog steps lead to

$$E\Big[Y_i^l Y_j^l K\Big(\frac{\hat{X}_i - x}{h}\Big) K\Big(\frac{\hat{X}_j - x}{h}\Big)\Big] = O(h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}),$$

and thus the covariance $|\text{cov}(Z_i, Z_j)|$ for $i \neq j$ is of the same rate.

On the other hand, we can directly make use of Lemma 3.2 because all $Y_i$ and $K$ are bounded so that $||Z_i||_\infty < \infty$. It follows

$$|\text{cov}(Z_i, Z_j)| \leq C\alpha(|i - j|).$$

The idea is now to combine both results. When the indices of the two variables $Z_i$ and $Z_j$ are close[12] to each other we use the first one, and when they are far from each other the second one. To control this, we introduce a sequence of integers $a_n$ and obtain

$$\begin{aligned} s_n^{2*} &= \sum_{i=1}^n \sum_{j \neq i} |\text{cov}(Z_i, Z_j)| \\ &\leq C\Big[\sum \sum_{0<|i-j|\leq a_n} \{h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}\} + \sum \sum_{|i-j|>a_n} \alpha(|i-j|)\Big]. \end{aligned}$$

Since $i \neq j$, the largest possible term for $a_n$ is $0 < |i - j| \leq a_n \cong n - 1$ and the smallest $|i - j| > a_n \cong 1$. Furthermore, the maximum number of elements in $s_n^{2*}$ is $n^2 - n$, and we obtain for $\Delta$ that is of the same order like the slowest term in $O(h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2})$

$$\frac{\sum \sum_{0<|i-j|\leq a_n} \Delta}{\Delta n a_n} \leq \frac{n^2 - n}{n(n-1)} \iff \sum \sum_{0<|i-j|\leq a_n} \Delta = O(\Delta n a_n)$$

and

$$\frac{\sum \sum_{|i-j|>a_n} \alpha(|i-j|)}{n^2 \alpha(a_n)} \leq \frac{n^2 - n}{n^2} \iff \sum \sum_{|i-j|>a_n} \alpha(|i-j|) = O(n^2 \alpha(a_n)).$$

This means that

$$s_n^{2*} = O(\{h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}\} n a_n + n^2 \alpha(a_n)).$$

Choosing $a_n$ from $\{\frac{1}{h \log n}, \frac{nh_0^\delta h}{\log n}, \frac{(nh_0^\delta)^2 h^3}{\log n}\}$ proves the Proposition. □

[12] Since we use time series data, this means that the two events are close in time.

*Proof of Proposition 3.4.* Using the algebraic mixing condition and Proposition 3.3

$$s_n^{2*} = o(nh) + O(n^2 \tilde{\Delta}^{-a}),$$

with $\tilde{\Delta}$ from

$$\left\{ \frac{1}{h \log n}, \frac{nh_0^\delta h}{\log n}, \frac{(nh_0^\delta h)^2 h}{\log n} \right\}.$$

Using the assumption (**??**) and noting that $\frac{(\log n)^a}{n^\varepsilon} \to 0$ for $n \to \infty$ closes the proof. □

*Proof of Proposition 3.5.* We use again that $Y_1$ is bounded so that remains to analyse

$$E\Big[K\Big(\frac{\hat{X}_1 - x}{h}\Big)\Big]^2 = \int\int K\Big(\frac{u - x + b(u) + v\sigma(u)}{h}\Big)^2 f(u,v) du dv.$$

With a Taylor-expansion and analog steps like in the proof of Proposition 3.3 we get

$$= \int\int \{K\Big(\frac{u-x}{h}\Big) + K'\Big(\frac{u-x}{h}\Big)\Big(\frac{b(u)+v\sigma(u)}{h}\Big) +$$
$$K''\Big(\frac{u-x}{h} + \kappa \frac{b(u)+v\sigma(u)}{h}\Big)\frac{(b(u)+v\sigma(u))^2}{2h^2}\Big\}^2 f(u,v) du dv,$$

where $\kappa \in (0,1)$, and find that

$$E\Big[K\Big(\frac{\hat{X}_1 - x}{h}\Big)\Big]^2 = O(h + (nh_0^\delta h)^{-1} + (nh_0^\delta)^2 h^3)^{-1})$$

what proves the Proposition. □

*Proof of Proposition 3.6.* Using $l = 0$ for $\psi = f$ and $l = 1$ for $\psi = q$, respectively, we directly get with (8)

$$\begin{aligned} |E\hat{\psi}(x) - \hat{\psi}(x)| &= \left| E\left(\frac{1}{nh}\sum_{i=1}^n Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right) - \frac{1}{nh}\sum_{i=1}^n Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right) \right| \\ &= \frac{1}{nh}\left| \sum_{i=1}^n \left\{ Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right) - E\left(Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right)\right\} \right| = \frac{1}{nh}\left|\sum_{i=1}^n Z_i\right|. \end{aligned}$$

Therefore, applying Lemma 3.1 we obtain

$$\begin{aligned} P\left(|E\hat{\psi}(x) - \hat{\psi}(x)| > \delta\right) &= P\left(\Big|\sum_{i=1}^n Z_i\Big| > nh\delta\right) \\ &\le 4\left(1 + \frac{\delta^2 n^2 h^2}{16 r s_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{8r}{nh\delta}\right)^{a+1}. \end{aligned}$$

Since we have seen in (9) that $s_n^2 = O(n\Delta)$, with $\Delta$ from Proposition 3.4, with $\delta = \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}$ we get

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \\ \le 4\left(1 + \frac{\varepsilon^2 \log n}{16r}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{8r}{\varepsilon}\right)^{a+1} (n\Delta \log n)^{-\frac{a+1}{2}}. \qquad (14)$$

Now, we can choose $r > 1$ such that $\log n = o(r)$, and use the limit definition

$$\exp(x) = \lim_{z\to\infty}\left(1 + \frac{x}{z}\right)^z,$$

with $z = -r/2$. For the first term of the right hand side of (14), we obtain for $z \to \infty$

$$\left(1 + \frac{\varepsilon^2 \log n}{16r}\right)^{-\frac{r}{2}} = \left(1 - \frac{\varepsilon^2 \log n}{32z}\right)^{z} \longrightarrow \exp\left(-\frac{\varepsilon^2 \log n}{32}\right) = n^{-\frac{\varepsilon^2}{32}}.$$

Noting that $C(\log n)^{-(a+1)/2} \le C$ for $n > 2$ and a constant $C$, (14) can be expressed as

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \qquad \le Cn^{-\frac{\varepsilon^2}{32}} + C\varepsilon^{-(a+1)} n^{1-\frac{a+1}{2}} r^a \Delta^{-\frac{a+1}{2}}.$$

With $r = n^b$ for $b > 0$, i.e. $\log n = o(r)$, and the left hand side of the assumption (10),

$$n^{1+ab-\frac{a+1}{2}}\Delta^{-\frac{a+1}{2}} \le n^{1+ab-\frac{a+1}{2}-\frac{3-a}{2}-\theta\frac{a+1}{2}} = n^{-1-\theta\frac{a+1}{2}+ab} = n^{-1-\nu}.$$

Thus, we obtain for a sufficiently small $b$ that

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \le Cn^{-\frac{\varepsilon^2}{32}} + C\varepsilon^{-(a+1)} n^{-1-\nu}.$$

Finally, for a sufficiently large $\varepsilon$, we get that exist $\nu, \varepsilon > 0$ such that

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \le Cn^{-1-\nu},$$

what proves the assertion. □

*Proof of Theorem 3.7.* From Proposition 3.6 follows directly

$$E\hat{q}(x) - \hat{q}(x) \longrightarrow 0, \quad \text{and} \quad E\hat{f}(x) - \hat{f}(x) \longrightarrow 0, \qquad (15)$$

both quasi completely. With the first part of the proof of Proposition 3.3 we obtain[13]

$$E\hat{f}(x) = \frac{1}{h} EK\left(\frac{\hat{X}-x}{h}\right) = f(x) + B_f(x) + o(h_0^2 + h),$$

with $B_f(x) = h^2/2f''(x)\mu_2(K) + \{b(x)f'(x) + b'(x)f(x)\}\mu_1(K')$, and thus

$$E\hat{f}(x) - f(x) \;\longrightarrow\; 0. \tag{16}$$

The analog can be shown for $E\hat{q}(x)$. With

$$E\hat{q}(x) = \frac{1}{h} EYK\left(\frac{\hat{X}-x}{h}\right),$$

and taking the conditional expectation for $X = x$, we get

$$E\hat{q}(x) = \frac{1}{h}\int\int m(u)K\left(\frac{u - x + b(u) + v\sigma(u)}{h}\right) f(u,v)dudv.$$

Repeating the same steps as in the first part of the proof of Proposition 3.3, using $q = m \cdot f$ as well as that the function $q$ is continuous over the compact support of the kernel $K$, i.e. that $q(x + hs) \longrightarrow q(x)$ uniformly in $s$, we obtain

$$E\hat{q}(x) - q(x) \;\longrightarrow\; 0. \tag{17}$$

Furthermore, from (16) and (11) follows the quasi complete convergence of $\hat{f}(x)$ to $f(x)$, i.e. for all $\varepsilon > 0$, it holds that

$$\sum_{n=1}^{\infty} P(|\hat{f}(x) - f(x)| > \varepsilon) < \infty.$$

Since $f(x) > 0$, we can define $\delta = \varepsilon = f(x)/2$ and get for $\delta > 0$

$$\sum_{n=1}^{\infty} P(\hat{f}(x) \leq \delta) < \infty. \tag{18}$$

Note, that with (7) and $q = f \cdot m$ we can state

$$\hat{m}_{NW}(x) - m(x) = \frac{\hat{q}(x) - q(x)}{\hat{f}(x)} + (f(x) - \hat{f}(x))\,\frac{m(x)}{\hat{f}(x)}, \tag{19}$$

and thus with (15) – (19) follows the assertion. □

[13] A similar result can be found in Theorem 2.1 (i) in Sperlich (2009).

# References


Asness, C. (2003). Fight the Fed Model: The Relationship Between Future Returns and Stock and Bond Market Yields. *Journal of Portfolio Management*, 30(1), 11–24.

Baele, L., Bekaert, G., Inghelbrecht, K. (2010). The Determinants of Stock and Bond Return Comovements. *The Review of Financial Studies*, 23(6), 2374–2428.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd ed. Springer, NY.

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

Campbell, J. Y., Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies*, 1(3), 195–228.

Campbell, J. Y., Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies*, 21(4), 1455–1508.

Chu, C. K., Marron, J. S. (1991). Comparison of Two Bandwidth Selectors With Dependent Errors. *The Annals of Statistics*, 19(4), 1906–1918.

Connolly, R., Stivers, C., Sun, L. (2010). Stock Market Uncertainty and the Stock-Bond Return Relation. *Journal of Financial and Quantitative Analysis*, 40(1), 161–194.

Dell'Aquila, R., Ronchetti, E. (2006). Stock and Bond Return Predictability: The Discrimination Power of Model Selection Criteria. *Computational Statistics and Data Analysis*, 50, 1478–1495.

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

Engsted, T., Tanggaard, C. (2001). The Danish Stock and Bond Markets: Comovement, Return Predictability and Variance Decomposition. *Journal of Empirical Finance*, 8, 243–271.

Fan, J., Yao, Q. (2003). *Nonlinear Time Series*. Springer, New-York.

Ferraty, F., Núñez-Antón, V., Vieu, P. (2001). *Regresión No Paramétrica: Desde la Dimensión Uno Hasta la Dimensión Infinita*. Servicio Editorial da la Universidad del Pais Vasco, Bilbao.

Gámiz-Pérez, M. L., Martínez-Miranda, M. D., Nielsen, J. P. (2012). Smoothing Survival Densities in Practice. *Computational Statistics and Data Analysis*, to appear.

Goyal, A., Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21(4), 1455–1508.

Guidolin, M., Hyde, St., McMillan, D., Ono, S. (2009). Non-linear Predictability in Stock and Bond Returns: When and Where is it Exploitable? *International Journal of Forecasting*, 25(2), 373–399.

Guidolin, M., Timmermann, A. (2006). An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns. *Journal of Applied Econometrics*, 21(1), 1–22.

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

Kim, T. Y., Park, B. U., Moon, M. S., Kim, K. (2009). Using Bimodal Kernel for Inference in Nonparametric Regression with Correlated Errors. *Journal of Multivariate Analysis*, 100, 1487–1497.

Lettau, M., Nieuwerburgh, S. V. (2008). Reconciling the Return Predictability Evidence. *The Review of Financial Studies*, 21(4), 1607–1652.

Lund, J., Engsted, T. (1996). GMM and Present Value Tests of the C_CAPM: Evidence from Danish, German, Swedish, and UK Stock Markets. *Journal of International Money and Finance*, 15, 497–521.

Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P., Sperlich, S. (2011). Do-validation for Kernel Density Estimation. *Journal of the American Statistical Association*, 106(494), p. 651–660.

Masry, E., Fan, J. (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian Journal of Statistics*, 24, 165–179.

McMillan, D. G. (2007). Non-linear Forecasting of Stock Returns: Does Volume Help? *International Journal of Forecasting*, 23(1), 115–126.

Nielsen, J. P., Sperlich, S. (2003). Prediction of Stock Returns: A New Way to Look at it. *Astin Bulletin*, 33, 399–417.

Park, B. U., Kim, W. C., Ruppert, D., Jones, M. C., Signorini, D. F., Kohn, R. (1997). Simple Transformation Techniques for Improved Non-parametric Regression. *Scandinavian Journal of Statistics*, 24(2), 145–163.

Rapach, D. E., Strauss, J. K., Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, 23(2), 821–862.

Rapach, D. E., Wohar, M. E., Rangvid, J. (2005). Macro Variables and International Stock Return Predictability. *International Journal of Forecasting*, 21(1), 137–166.

Rio, E. (2000). Théorie Asymptotique pour des Processus Aléatoires Faiblement Dépendants. In *SMAI, Mathématiques et Applications* 31. Springer, Berlin.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics* . Wiley, NY.

Shiller, R. J., Beltratti, A. E. (1992). Stock Prices and Bond Yields. *Journal of Monetary Economics*, 30, 25–46.

Sperlich, S. (2009). A Note on Non-parametric Estimation with Predicted Variables. *The Econometrics Journal*, 12, 382–395.

Sperlich, S., Linton, O., Härdle, W. (1999). Integration and Backfitting Methods in Additive Models – Finite Sample Properties and Comparison. *Test*, 8, 419–458.

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

Vieu, Ph. (1994). Choice of Regressors in Nonparametric Estimation. *Computational Statistics and Data Analysis*, 17, 575–594.

## Graz Economics Papers

For full list see:
http://www.uni-graz.at/vwlwww/vwlwww_forschung/vwlwww_gep.htm
Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

---

10–2012 **Michael Scholz, Stefan Sperlich, Jens Perch Nielsen**: Nonparametric prediction of stock returns with generated bond yields

09–2012 **Joern Kleinert, Nico Zorell**: The export-magnification effect of offshoring

08–2012 **Robert J. Hill, Iqbal A. Syed**: Hedonic Price-Rent Ratios, User Cost, and Departures from Equilibrium in the Housing Market

07–2012 **Robert J. Hill, Iqbal A. Syed**: Accounting for Unrepresentative Products and Urban-Rural Price Differences in International Comparisons of Real Income: An Application to the Asia-Pacific Region

06–2012 **Karl Steininger, Christian Lininger, Susanne Droege, Dominic Roser, Luke Tomlinson**: Towards a Just and Cost-Effective Climate Policy: On the relevance and implications of deciding between a Production versus Consumption Based Approach

05–2012 **Miriam Steurer, Robert J. Hill, Markus Zahrnhofer, Christian Hartmann**: Modelling the Emergence of New Technologies using S-Curve Diffusion Models

04–2012 **Christian Groth, Ronald Wendner**: Embodied learning by investing and speed of convergence

03–2012 **Bettina Brüggemann, Joern Kleinert, Esteban Prieto**: The Ideal Loan and the Patterns of Cross-Border Bank Lending

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

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