Functional prediction of intraday cumulative returns

Piotr Kokoszka\textsuperscript{1} and Xi Zhang\textsuperscript{2}
\textsuperscript{1}Department of Statistics, Colorado State University, USA
\textsuperscript{2}Department of Mathematics & Statistics, Utah State University, USA

Abstract: We define cumulative intraday returns and consider their prediction from such returns on a market index. We model these returns as curves in a function space. We propose several functional regression models which can be viewed as extensions of the capital asset pricing model to intraday returns defined as curves. After deriving parameter estimates and prediction functions for these models, we compare their prediction errors by application to cumulative intraday returns of large U.S. corporations. We find that complex functional regression models do not perform better than a simple model. In particular, we find that modelling error dependence does not improve forecasts.

Key words: functional linear prediction; intraday returns

Received November 2011; revised February 2012; accepted February 2012

1 Introduction

A well-known application of linear regression to financial data is the celebrated capital asset pricing model (CAPM), see, e.g., Chapter 5 of Campbell et al. (1997) or Chapter 16 of Ruppert (2011). In its simplest form, it is defined by the straight line regression

\[ r_n = \alpha + \beta r_{m,n} + \varepsilon_n, \]  

where

\[ r_n = 100(\ln P_n - \ln P_{n-1}) \approx 100 \frac{P_n - P_{n-1}}{P_{n-1}} \]  

is the return, in per cent, over a unit of time on a specific asset, e.g., a stock, and \( r_{m,n} \) is the analogously defined return on a relevant market index. The unit of time can be day, month or year. This simple model has been extended in many ways and extensively investigated over the decades. It has been used to test various hypotheses on the behaviour of investors and markets. But the underlying idea is to assess how strongly a return on an asset or a portfolio depends on a return on an index portfolio, though the estimation of the so-called betas, i.e., the regression slope coefficients.
In this paper, we focus on intraday price data, which have different properties than daily or monthly closing prices, see Chapter 5 of Tsay (2005), Guillaume et al. (1997) and Andersen and Bollerslev (1997a, 1997b). Before we explain the objectives of our research, we introduce the following definition.

Definition 1.1 Suppose $P_n(t_j), n = 1, \ldots, N, j = 1, \ldots, m$ is the price of a financial asset at time $t_j$ on day $n$. The functions

$$r_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)],$$

are defined as the *intraday cumulative returns*.

The above definition implicitly assumes that $t_{j+1} > t_j$.

Graphs of intraday cumulative returns are shown in Figure 1. They should be contrasted with high frequency returns defined as $\ln P(t_j) - \ln P(t_{j-1})$ and displayed in

![Graph of intraday cumulative returns](image1.png)

**Figure 1** Intraday cumulative returns on 10 consecutive days for the Exxon Mobil corporation (XOM) and the Standard & Poor’s 100 index (S&P)
Figure 2. To be more specific, we work with one-minute averages, so \( t_{j+1} - t_j = 1 \) min, and \( P(t_j) \) is the average of the maximum and minimum price within the \( j \)th minute.

The intraday cumulative returns have an obvious interpretation as curves describing the cumulative return during a given day. They are similar to the curves of the price \( P_n(t_j) \) for a trading day \( n \) which are routinely displayed on business news and investment companies’ websites. The curves \( P_n(t_j), j = 1, \ldots, m \), with \( P_n(t_1) \) corresponding to the asset price in the first and \( P_n(t_m) \) to the price in the last minute of trading day \( n \), are followed throughout the day by both longer term investors who seek to predict a closing price for the day, and by intraday investors who wish to trade at the most opportune time of the day. In the recent economic turmoil, on some days, intraday index price movements are followed even by non-business news casts. The intraday cumulative return curves give more relevant information because they show how the return evolves during a trading day, but since \( r_n(t_j) \approx 100(P_n(t_j) - P_n(t_1))/P(t_1) \), with \( P_n(t_1) \) being a constant for a given day \( n \), both curves look very similar.

The previous paragraph emphasizes that a natural way to look at the values \( r_n(t_j) \) is to treat them as continuous curves, one curve per day, and this is, in fact, how they are displayed at the relevant websites, and how they are viewed by investors. For every fixed day, these curves exhibit a specific pattern, typically with some upward or downward momentum, disturbed by some noise. It is therefore tempting to study
their statistical behaviour using the framework of functional data analysis (FDA), which has grown over the last 20 years into an important and expanding branch of statistics. At its core is the idea that curves should be treated as individual and complete statistical objects, rather than as collections of individual observations. Statistical tools of FDA typically rely on some form of smoothing to transform high dimensional or incomplete data building up a curve into a smoother curve that can be described by a smaller number of parameters. The tools of FDA are appropriate for the intraday cumulative returns but not for high frequency returns like those shown in Figure 2; the latter are very noisy and any reasonable smoothing destroys the information they contain (the smoothed curves are typically very close to a constant zero curve). In the following sections of the paper, we describe the tools of FDA we need, but obviously we cannot go into much detail in this short paper; Horváth and Kokoszka (2012) and Hörmann and Kokoszka (2012) contain the most relevant broader background. We also recommend the comprehensive introductory works of Ramsay and Silverman (2002, 2005) and Ramsay et al. (2009), and more theoretical expositions by Bosq (2000), Ferraty and Vieu (2006), Bosq and Blanke (2007) and Ferraty and Romain (2011).

The goal of this paper is to model the relationship between the intraday cumulative return curves for a single asset and those for a market index, and to evaluate their relevance by comparing their predictive power. The functional linear models we consider can be viewed as expansions of the CAPM model (1.1). We study models in which the linear relationship is quantified by means of a bivariate kernel \( \psi(\cdot, \cdot) \), with \( \psi(t, s) \) describing the impact of the value of \( r_{m,n}(s) \) on \( r_n(t) \). We also consider a much simpler model with the kernel \( \psi(\cdot, \cdot) \) replaced by a single coefficient \( \psi \). We study models with and without intercept, and with independent and with correlated errors. We seek to answer several questions: Can a simpler model with a scalar coefficient give predictions as good as a model with a kernel coefficient? Does including an intercept improve predictions, or does this extra parameter actually make them worse? Does modelling error correlation lead to improved predictions? As we will see in the following, the intercept is very important. In the classical scalar CAPM, the intercept is related to the risk-free rate: if the risk-free rate is zero, the intercept is zero (at least theoretically). Since the intraday cumulative returns focus only on the changes within a day (they exclude \( P_n(0) - P_{n-1}(1) \)), the impact of the risk-free rate is not obvious. Error correlation is very common in regression for economic data, such data actually motivated the introduction of linear regression models with correlated errors by Cochrane and Orcutt (1949). One might therefore expect that taking into account error correlation leads to more useful models also in the context of intraday curves. Surprisingly, we found no evidence for it if the predictive power is used as a criterion.

We conclude this introduction by noting that while modelling scalar returns, (1.2) have been the subject of research for over a hundred years, which lead to the development of several areas of statistics and probability, almost nothing is known about the statistical properties of the intraday cumulative returns, and no modelling framework has been studied in any depth. This paper is an attempt to investigate
for the intraday returns curves a question which has been of central importance for the scalar returns (1.2), namely the connection between single risky asset returns and returns on a market portfolio. The modelling approach we adopt focuses on the instantaneous dependence between the shapes of curves of two assets. It should be contrasted with extensive research on temporal predictability of intraday returns which has received a fair amount of attention lately: Chordia et al. (2005) connect a pattern of intraday dependence lasting less than one hour to order imbalances within a trading day. Matías and Reboredo (2012) and Roboredo et al. (2011) study the temporal predictability of intraday returns taken over intervals ranging from five minutes to one hour and find that relatively simple nonlinear models perform better than linear models or more complex nonlinear models, especially if economic criteria are used. Wang and Yang (2010) study temporal predictability of returns on such scales based on energy market assets, and find predictability only in ‘bubble’ periods.

The paper is organized as follows. In Section 2, we postulate several regression models for intraday cumulative returns, while Section 3 focuses on their estimation. Prediction of intraday cumulative return is addressed in Section 4. Section 5 compares the predictions for selected U.S. stocks. We conclude with Section 6 which summarizes our findings and states broad conclusions.

2 Regression models for intraday cumulative returns

We now describe the models we propose. In our application, we work with stocks traded at NYSE. In the formulas that follow, the trading period (9:30AM to 4:00PM EST) is re-scaled onto the interval [0, 1]. All integrals refer to integration over this unit interval. To further ease the notation, in the following, we denote the regressors \(r_m, n(t)\) by \(X_n(t)\) and the responses \(r_n(t)\) by \(Y_n(t)\). Randomness is introduced to all models via error functions \(\varepsilon_n(t)\), which are assumed to be identically distributed and independent of the regressors \(X_n(t)\) (but not necessarily serially independent).

Simple Functional CAPM (SF). A simple functional CAPM is defined as

\[ Y_n(t) = \alpha + \psi X_n(t) + \varepsilon_n(t), \quad t \in [0, 1]. \]  (2.1)

A model without the intercept \((\alpha \equiv 0)\), denoted SF*, is also considered.

Fully Functional CAPM (FF). This model is defined by the relation

\[ Y_n(t) = \alpha(t) + \int \psi(t, s) X_n(s) ds + \varepsilon_n(t), \quad t \in [0, 1]. \]  (2.2)

If \(\alpha \equiv 0\), this model is denoted FF*.

In the above models, the error functions \(\varepsilon_n(\cdot)\) are assumed to be independent. If the latter assumption is dropped, we arrive at the following models.

Simple Functional CAPM with dependent errors (SFDE). This model is defined by (2.1), but the errors are assumed to follow a functional autoregressive process of
order 1, FAR(1) process:

\[ \epsilon_n(t) = \int \varphi(t, s) \epsilon_{n-1}(s) \, ds + w_n(t), \quad (2.3) \]

where the \( w_n \) are iid mean zero random functions.

**Fully Functional CAPM with dependent errors (FFDE).** This model is defined by (2.2) with errors which follow the FAR(1) process.

As will be seen in the following, models without the intercept perform poorly, so to conserve space, we do not consider models with dependent errors but without the intercept.

To complete the description of these models, we must specify assumptions on the distribution of error and regressor functions and on the parameters. All statements listed below are verified in Horváth and Kokoszka (2012) (they are easy to verify). The errors \( \epsilon_n \) and the regressors \( X_n \) are assumed to be random elements of \( L^2 = L^2([0, 1]) \), the Hilbert space of square integrable functions with the norm and the inner product defined by

\[ \|x\|^2 = \int x^2(t) \, dt, \quad \langle x, y \rangle = \int x(t)y(t) \, dt, \quad x, y \in L^2. \]

We assume that the \( X_n \), just like the \( \epsilon_n \) are identically distributed, and that the following moment conditions hold:

\[ E\|X_n\|^2 < \infty, \quad E\|\epsilon_n\|^2 < \infty, \quad E\epsilon_n(t) = 0. \]

Under these conditions, and the conditions on the parameters stated below, the responses \( Y_n \) are also random elements of \( L^2 \) which satisfy \( E\|Y_n\|^2 < \infty \). In models SF and SF*, the parameters \( \alpha \) and \( \psi \) can be any real numbers. In models FF and FF*, the regression kernel is assumed to satisfy

\[ \int \int \psi^2(t, s) \, dt \, ds < \infty. \quad (2.4) \]

The intercept function \( \alpha(\cdot) \) is an element of \( L^2 \). In models SFDE and FFDE, the autoregressive kernel \( \varphi(t, s) \) must satisfy

\[ \int \int \varphi^2(t, s) \, dt \, ds < 1. \quad (2.5) \]

Condition (2.5) implies that the \( \epsilon_n \) in (2.3) form a strictly stationary sequence of functions in \( L^2 \).

One could impose a more general dependence structure on the error functions, like the approximability condition introduced by Hörmann and Kokoszka (2010), but for the purpose of prediction, we need an estimable structure. That is why we follow a fairly standard practice, and restrict ourself to the FAR(1) model, which is relatively easy to estimate.
The assumption of stationarity of the return curves is difficult to verify in a fully satisfactory manner, as we are not aware of such tests in the functional setting. Application of several standard univariate tests to the integrated curves $\int r_n(t) dt$ indicates that these integrals form a stationary time series for the data we use.

3 Estimation of the regression models

In this section, we explain how the parameters in the models of Section 2 are estimated. Even though the derivation of the estimators is not difficult, we present it in some detail because the framework and tools of FDA may be unfamiliar to some readers, and the closed-form formulas we display may be useful in other applications.

All calculations have been performed in the R package fda, see Ramsay et al. (2009) for a solid introduction to main computational techniques and packages used in FDA. The first step is to convert the cumulative returns in one-minute resolution to functional objects. This is done by expanding the vectors containing the 390 daily values with respect to an orthonormal basis. We used the Fourier basis with 99 basis functions. In this first step, the return curves are replaced by slightly smoother curves constructed using the first 99 Fourier coefficients. (This step is needed only to create functional objects, so that further calculations are possible. The results are the same if B-splines are used.) These coefficients are stored and used for all further computations. In particular, they allow us to compute the empirical functional principal components (EFPCs) of the data. For example, using the regressor functions $X_1, X_2, \ldots, X_N$ (represented through their Fourier coefficients), we can calculate, for any $p \geq 1$, orthonormal functions $\hat{v}_1, \ldots, \hat{v}_p$ such that the $X_k$ can be optimally represented as $X_k - \mu_X \approx \sum_{i=1}^{p} \xi_{ik} \hat{v}_i$, $(\mu_X = EX_k)$, or more explicitly as

$$X_k(t) - \mu_X(t) \approx \sum_{i=1}^{p} \hat{\xi}_{ik} \hat{v}_i(t), \quad \hat{\xi}_{ik} = \int (X_k(t) - \mu_X(t)) \hat{v}_i(t) dt. \quad (3.1)$$

The optimality and the meaning of the approximation are discussed in Ramsay and Silverman (2005) or Horváth and Kokoszka (2012). The EFPCs $\hat{v}_i$ are also known as the ‘optimal empirical orthonormal basis’ or ‘natural orthonormal components’. The coefficients $\hat{\xi}_{ik}$ are called the (empirical) scores. The idea behind expansion (3.1) is that the data can be represented using only a few optimal basis functions, $p$ is typically a single digit number. There are several methods of determining the optimal $p$. A popular one uses the fact that $\sum_{i=1}^{N} \hat{\lambda}_i$, where $\hat{\lambda}_i$ is the eigenvalue associated with $\hat{v}_i$, is equal to the sample variance of the $X_k$. A recommendation that is often given is to use $p$ such that $\sum_{i=1}^{p} \hat{\lambda}_i$ is equal to between 85 and 90% of the total variance.
Piotr Kokoszka and Xi Zhang

\[ \sum_{i=1}^{N} \hat{\lambda}_i \]. We will work with an analogous expansion for the responses \( Y_k \):

\[ Y_k(t) - \mu_Y(t) \approx \sum_{j=1}^{q} \hat{\xi}_{jk} \hat{u}_j(t), \quad \hat{\xi}_{jk} = \int (Y_k(t) - \mu_Y(t)) \hat{u}_j(t) dt. \] (3.2)

We will use the usual sample means

\[ \bar{Y}_N(t) = \frac{1}{N} \sum_{n=1}^{N} Y_n(t), \quad \bar{X}_N(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t) \] (3.3)

and the centered cumulative returns

\[ Y^c_n(t) = Y_n(t) - \bar{Y}_N(t), \quad X^c_n(t) = X_n(t) - \bar{X}_N(t) \] (3.4)

In the remainder of this section, we derive the parameter estimators in each model. To facilitate the exposition, we progress from the simplest to the most complex models.

**SF* (SFCAPM \( \alpha \equiv 0 \)). This model is \( Y_n(t) = \psi X_n(t) + \varepsilon_n(t) \). The optimal \( \psi \) minimizes the expected integrated square error

\[ E || Y_n - \psi X_n ||^2 = E \int [Y_n(t) - \psi X_n(t)]^2 dt. \]

Observing

\[ E || Y_n - \psi X_n ||^2 = E ||Y_n||^2 - 2 \psi E \langle Y_n, X_n \rangle + \psi^2 E \| X_n \|^2. \]

Differentiating with respect to \( \psi \), we obtain

\[ \hat{\psi} = \frac{\sum_{n=1}^{N} \langle Y_n, X_n \rangle}{\sum_{n=1}^{N} \| X_n \|^2}. \] (3.5)

**SF (SFCAPM \( \alpha \neq 0 \)). Taking expectation of both sides in (2.1), we get \( EY_n(t) = \alpha + \psi EX_n(t) \), i.e., \( \alpha = \mu_Y(t) - \psi \mu_X(t) \). Inserting into (2.1), we get

\[ Y_n(t) - \mu_Y(t) = \psi(X_n(t) - \mu_X(t)) + \varepsilon_n(t). \] (3.6)

Using the centered returns (3.4), we can rewrite (3.6) as

\[ Y^c_n(t) = \psi X^c_n(t) + \varepsilon^*_n(t), \] (3.7)

where the errors \( \varepsilon^* \) contain the errors in the mean estimation. By analogy to (3.5), the estimator of \( \psi \) is

\[ \hat{\psi} = \frac{\sum_{n=1}^{N} \langle Y^c_n(t), X^c_n(t) \rangle}{\sum_{n=1}^{N} \| X^c_n(t) \|^2}. \] (3.8)
The estimator of $\alpha$ in model SF is

$$\hat{\alpha} = \bar{Y}_N(t) - \hat{\psi} \bar{X}_N(t).$$  (3.9)

**FF* (FCAPM $\alpha \equiv 0$). This model is**

$$Y_n(t) = \int \psi(t, s) X_n(s) ds + \varepsilon_n(t).$$  (3.10)

The estimation of the kernel $\psi(\cdot, \cdot)$ based on the EFPCs is discussed in Horváth and Kokoszka (2012). It can be shown that the method of moments estimator based on the minimization of an appropriate Hilbert–Schmidt norm is given by

$$\hat{\psi}(t, s) = \sum_{k=1}^{q} \sum_{\ell=1}^{p} \hat{\lambda}_k^{\ell-1} \hat{\sigma}_{k \ell} \hat{u}_k(t) \hat{v}_\ell(s).$$  (3.11)

where $\hat{\lambda}_k$ is the eigenvalue corresponding to $\hat{v}_\ell$ and

$$\hat{\sigma}_{k \ell} = \frac{1}{N} \sum_{i=1}^{N} \langle X_i, \hat{v}_\ell \rangle \langle Y_i, \hat{u}_k \rangle.$$  (3.12)

**FF (FCAPM $\alpha \neq 0$). Taking the expectations in (2.2), we obtain**

$$Y_n(t) - \mu_Y(t) = \int \psi(t, s)(X_n(s) - \mu_X(s)) ds + \varepsilon_n(t);$$  (3.13)

$$\alpha(t) = \mu_Y(t) - \int \psi(t, s) \mu_X(s) ds.$$

Recall the definition of the centered observations $Y^c_n$ and $X^c_n$ given in (3.4) and denote by $\hat{u}^c_k$ and $\hat{v}^c_\ell$ their EFPCs. Following the argument for the model without the intercept, an estimator for $\psi(t, s)$ is

$$\hat{\psi}(t, s) = \sum_{k=1}^{q} \sum_{\ell=1}^{p} \hat{\lambda}_k^{\ell-1} \hat{\sigma}_{k \ell}^c \hat{u}^c_k(t) \hat{v}^c_\ell(s),$$  (3.14)

where

$$\hat{\sigma}_{k \ell}^c = \frac{1}{N} \sum_{i=1}^{N} \langle X^c_i, \hat{v}^c_\ell \rangle \langle Y^c_i, \hat{u}^c_k \rangle,$$  (3.15)

and where $\hat{\lambda}_k^c$ is the eigenvalue corresponding to $\hat{v}^c_\ell$. The estimator of $\alpha(t)$ is

$$\hat{\alpha}(t) = Y_N(t) - \int \hat{\psi}(t, s) \bar{X}_N(s) ds.$$  (3.16)
SFDE and FFDE. Recall that these models are defined, respectively, by equations (2.1) and (2.2) in which the errors \( \varepsilon_n \) satisfy autoregression (2.3). The parameters are \( \psi, \alpha \) and the autoregressive kernel \( \phi(\cdot, \cdot) \). The idea of estimation in these models is as follows. In the first step, we ignore the dependence of the \( \varepsilon_n \) and estimate \( \psi \) and \( \alpha \), respectively, by (3.8) and (3.9) for model SF, and by (3.14) and (3.16) for model FF. We use these estimates to construct residuals \( \hat{\varepsilon}_n \). Assuming that these residuals (rather than the unobservable errors \( \varepsilon_n \)) follow autoregression (2.3), we estimate the kernel \( \phi(\cdot, \cdot) \).

Scalar linear regression with dependent errors has been extensively studied beginning with the pioneering work of Cochrane and Orcutt (1949), and many estimation methods have been proposed, see Durbin (1960) and Rao and Griliches (1969). We extend to the functional setting the original method of Cochrane and Orcutt (1949), which has been found to be competitive in the scalar case. We now explain the above steps via corresponding formulas for model FFDE, the formulas for model SFDE are analogous, and simpler.

The residuals are

\[
\hat{\varepsilon}_n(t) = Y_n(t) - \hat{\alpha}(t) - \int \hat{\psi}(t, s) X_n(s) ds
\]

To estimate the autoregressive kernel \( \phi(\cdot, \cdot) \) in (2.3) (but with the \( \varepsilon_n \) replaced by the \( \hat{\varepsilon}_n \)), we use the method proposed by Bosq (2000), which was shown by Didericksen et al. (2011) to produce the best predictions of \( \varepsilon_n \) based on \( \varepsilon_{n-1}, \varepsilon_{n-2}, \ldots \). Denote by \( \hat{m}_k, k = 1, 2, \ldots, p \), the EFPCs of \( \hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_N \). The estimator of \( \phi(\cdot, \cdot) \) is given by a formula similar to (3.11), namely

\[
\hat{\phi}_p(t, s) = \sum_{k, \ell=1}^{p} \hat{\phi}_{k\ell} \hat{m}_k(t) \hat{m}_\ell(s),
\]

where

\[
\hat{\phi}_{k\ell} = \hat{\gamma}_\ell^{-1}(N - 1)^{-1} \sum_{n=1}^{N-1} \langle \hat{\varepsilon}_n, \hat{m}_\ell \rangle \langle \hat{\varepsilon}_{n+1}, \hat{m}_k \rangle.
\]

and where \( \hat{\gamma}_\ell \) is the eigenvalue corresponding to \( \hat{m}_\ell \).

4 Prediction of intraday return curves

The prediction function for models SF* and SF is

\[
\hat{Y}_n(t) = \hat{\alpha} + \hat{\psi} X_n(t).
\]
Functional prediction of intraday cumulative returns

with \( \hat{\alpha} \equiv 0 \) for model SF*. In model SF, by inserting the estimator of \( \alpha \), i.e., (3.9), into prediction function (4.1), we get a more explicit formula

\[
\hat{Y}_n(t) = \hat{Y}_N(t) + \hat{\psi} (X_n(t) - \bar{X}_n(t)).
\]

(4.2)

The general prediction function for models FF* and FF is

\[
\hat{Y}_n(t) = \hat{\alpha}(t) + \int \hat{\psi}(t, s) X_n(s) ds,
\]

(4.3)

where \( \hat{\alpha}(t) \equiv 0 \) in model FF*. In model FF, by inserting the estimator of \( \alpha(t) \), i.e., (3.16), into the prediction function, we get

\[
\hat{Y}_n(t) = \bar{Y}_N(t) + \int \hat{\psi}(t, s) (X_n(s) - \bar{X}_n(s)) ds.
\]

(4.4)

The remainder of this section is devoted to prediction using the models with dependent errors. To understand the idea, it is useful to focus first on a simple scalar model

\[
Y_n = \psi X_n + \varepsilon_n, \quad \varepsilon_n = \varphi \varepsilon_{n-1} + w_n.
\]

All quantities in the above equation, including \( Y_n, X_n, \varepsilon_n, w_n \) are scalars. Consider the random variables

\[
\tilde{Y}_n = Y_n - \varphi Y_{n-1}, \quad \tilde{X}_n = X_n - \varphi X_{n-1}
\]

and observe that

\[
\psi \tilde{X}_n = \psi (X_n - \varphi X_{n-1})
= \psi X_n - \varphi \psi X_{n-1}
= Y_n - \varepsilon_n - \varphi [Y_{n-1} - \varepsilon_{n-1}]
= Y_n - [\varphi \varepsilon_{n-1} + w_n] - \varphi Y_{n-1} + \varphi \varepsilon_{n-1}
= [Y_n - \varphi Y_{n-1}] - w_n = \tilde{Y}_n - w_n.
\]

(4.5)

(4.6)

The above calculation, proposed by Cochrane and Orcutt (1949), shows that the transformed variables satisfy the regression \( \tilde{Y}_n = \psi \tilde{X}_n + w_n \), in which the errors \( w_n \) are independent. The estimate of \( \psi \) in this new regression will therefore be unbiased. The key to obtaining this regression, is the cancellation of the terms \( \varphi \varepsilon_{n-1} \) in (4.6). It is achieved by changing the order of \( \psi \) and \( \varphi \) in (4.5). This operation is not possible in model FFDE because kernel operators do not commute. It is, however, possible in model SFDE, as we now describe.

Recall the SFDE model equation

\[
Y_n(t) - \mu_Y(t) = \psi (X_n(t) - \mu_X(t)) + \varepsilon_n(t)
\]

together with (2.3). Introduce the transformed functions

\[
\tilde{X}_n(t) = (X_n(t) - \mu_X(t)) - \int \varphi(t, s) (X_{n-1}(t) - \mu_X(t));
\]
Piotr Kokoszka and Xi Zhang

\[ \tilde{Y}_n(t) = (Y_n(t) - \mu_Y(t)) - \int \varphi(t, s) (Y_{n-1}(t) - \mu_Y(t)) ds \]

and observe that
\[ \psi \tilde{X}_n(t) = \psi (X_n(t) - \mu_X(t)) \]
\[ = (Y_n(t) - \mu_Y(t)) - \int \varphi(t, s) \psi (X_{n-1}(s) - \mu_X(s)) ds \]
\[ = (Y_n(t) - \mu_Y(t)) - \int \varphi(t, s) \{ [Y_{n-1}(s) - \mu_Y(s)] - \varepsilon_{n-1}(s) \} ds \]
\[ = (Y_n(t) - \mu_Y(t)) - \int \varphi(t, s) \varepsilon_{n-1}(s) ds - w_n(t) \]
\[ = \tilde{Y}_n(t) - w_n(t). \]

We have verified that the transformed functions satisfy the regression
\[ \tilde{Y}_n(t) = \psi \tilde{X}_n(t) + w_n(t) \]
with independent error functions \( w_n(t) \). The transformed functions are unobservable, so we replace them by

\[ X^*_n(t) = X_n(t) - \int \hat{\varphi}(t, s) X_{n-1}(s) ds; \]
\[ Y^*_n(t) = Y_n(t) - \int \hat{\varphi}(t, s) Y_{n-1}(s) ds. \]

The above calculation shows that the pairs \((X^*_n, Y^*_n)\) approximately follow model SF*. In replacing the functions \( \tilde{X}_n, \tilde{Y}_n \) by \( X^*_n, Y^*_n \), we have, however, neglected the effect of mean estimation, and since, as we will have seen, models with intercept lead to better predictions, we postulate that the pairs \((X^*_n, Y^*_n)\) follow model SF. This leads to the prediction function
\[ \hat{Y}^*_n(t) = \alpha^\wedge + \psi^\wedge X^*_n(t) \]
(4.7)
with \( \psi^\wedge \) and \( \alpha^\wedge \) given, respectively, by (3.8) and (3.9), but with \((X_n, Y_n)\) replaced by \((X^*_n, Y^*_n)\). Then the prediction function for SFDE model is derived recursively as
\[ \hat{Y}_n(t) = \bar{Y}_N(t) + \hat{Y}^*_n(t) + \int \hat{\varphi}(t, s) Y^*_n(s) ds, \quad n \geq 2. \]
(4.8)
5 Application to U.S. stocks

In this section, we apply functional prediction to cumulative intraday returns on stocks of 10 large U.S. corporations representing five sectors. Details are presented in Table 1. Market returns are represented by the Standard & Poor’s 100 index which contains the largest U.S. corporations. We evaluate the quality of prediction by the integrated mean squared error defined as

\[ MSEP(N) = N^{-1} \sum_{n=1}^{N} \int (Y_n(t) - \hat{Y}_n(t))^2 dt. \] (5.1)

Our goal is to answer the questions raised in Section 1, and to identify any additional patterns.

Using the integrated mean squared error, (5.1) is motivated by the theory of Section 3 which fits the models by minimizing the integrated expected squared error. In the context of scalar point forecasts, Gneiting (2011) argues that their evaluation should be based on the measures used to derive them. No similar arguments are available for function-valued predictions, especially for nontemporal ones, but we believe that the measure (5.1) is natural and useful. We note that if one moves to interval and density forecasts, one finds more sophisticated approaches to the evaluation of predictive power, see, e.g., Clements and Taylor (2003) and Hong et al. (2007).

Before reporting the results of the whole empirical study, we note that the prediction methods have been derived under the assumptions of stationarity of the cumulative intraday returns and the finiteness of their second moments, cf. Section 2. If these assumptions are violated in a major way, none of the methods will perform well. In particular, our methods are sensitive to large outliers and to regime changes. This is also true for the classical CAPM, and is difficult to remedy. In the analysis that follows, we consider 1000-day long periods (possibly different for different stocks). All these periods are contained in the interval from 01.03.2000 to 02.22.2006.

### Table 1: Description of 10 Stocks representing five sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Stocks</th>
<th>Full Name</th>
<th>1000 days period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>XOM</td>
<td>Exxon Mobil Corporation</td>
<td>05/25/2000—05/19/2004</td>
</tr>
<tr>
<td></td>
<td>CVX</td>
<td>Chevron Corporation</td>
<td>10/10/2001—07/23/2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12/13/2004—02/22/2006</td>
</tr>
<tr>
<td>Information Technology</td>
<td>MSFT</td>
<td>Microsoft Corporation</td>
<td>05/25/2000—05/19/2004</td>
</tr>
<tr>
<td></td>
<td>IBM</td>
<td>IBM Corporation</td>
<td>01/03/2000—12/24/2003</td>
</tr>
<tr>
<td>Financial</td>
<td>CITI</td>
<td>Citi Bank</td>
<td>10/17/2000—03/07/2005</td>
</tr>
<tr>
<td></td>
<td>BOA</td>
<td>Bank of America Corporation</td>
<td>03/13/2001—12/19/2005</td>
</tr>
<tr>
<td>Consumer Staples</td>
<td>KO</td>
<td>Coca-Cola</td>
<td>05/25/2000—05/19/2004</td>
</tr>
<tr>
<td></td>
<td>WMT</td>
<td>Wal-Mart Stores</td>
<td>05/25/2000—05/19/2004</td>
</tr>
<tr>
<td>Consumer Discretionary</td>
<td>MCD</td>
<td>McDonald's Corporation</td>
<td>10/17/2000—03/07/2005</td>
</tr>
<tr>
<td></td>
<td>DIS</td>
<td>The Walt Disney Corporation</td>
<td>05/25/2000—05/19/2004</td>
</tr>
</tbody>
</table>
Figure 3  MSEPs for four independent error models. The y-axis represents the MSEP, the x-axis represents the 10 periods.
some stocks, the one thousand long period contains a gap. We separate the periods of 1000 days into 10 consecutive 100-day periods, that do not contain visible regime switches or outliers. We report the MSEP (Mean Squared Error of Prediction) for each 100-day long period separately. We obtain an overall measure of performance by reporting the averages of these 10 numbers. The fact that different periods are used for different stocks does not impact our conclusions because we compare models for many datasets, not the datasets themselves.

We first consider in-sample prediction. Figure 3 shows the average MSEPs for the 10 periods for the stocks representing the energy and information technology sectors. The pattern for the remaining six stock is similar, and is not shown to conserve space. Figure 4 shows the grand average of the 10 averages for all 10 stocks. Figures 3 and 4 focus on the methods that use the assumption of independent regression errors. It is immediately seen that models without an intercept perform much worse, and that model SF, despite its simplicity, produces forecasts as good as the more complex model FF. We therefore compare model SFDE only to model SF, to better see potential small differences. Figure 5 illustrates a general finding that for some stocks and some periods, model SFDE can be slightly better, but can sometimes give predictions with a larger MSEP. This is confirmed by the examination of Figure 6 which compares the grand averages for all 10 stocks.
Figure 5  MSEPs for SF and SFDE models on XOM and MCD stocks for the 10 periods

Figure 6  Comparison of MSEPs for SF and SFDE models for 10 stocks. The y-axis is the average MSEP over ten 100-day periods

Functional prediction of intraday cumulative returns

Figure 7 Out-of-sample MSEPs for four independent error models and four selected stocks in nine periods
We now turn to out-of-sample predictions. Each model is estimated on a 100-day period, and used to predict cumulative intraday returns in the next 100-day period. This leads to nine MSEP for each stock and each method. Again, to conserve space, we show these MSEP only for the same four stocks as in Figure 3, see Figure 7. Figures 8–10 are analogous to Figures 4, 5 and 6.

6 Summary and conclusions

The goal of this paper was to propose several functional regression models that might be useful for predicting cumulative intraday returns on individual stock from those on a market index, to compare their predictive power, and to arrive at practical recommendations. The conclusions are fairly clear from the empirical study described in Section 5. We can summarize them as follows:
Models with intercept, i.e., SF and FF, make better prediction than models without intercept, i.e., SF* and FF*. The latter should not be used.

2. Modelling error dependence with a functional AR(1) model does not improve MSEPs.

3. The two models with intercept, i.e., SF and FF, do NOT dominate each other. They have almost the same MSEPs.

4. Out-of-sample predictions are not as good as the in-sample predictions, but the ranking of the models remains the same. There is a hint that in the out-of-sample prediction model, FF might be slightly better than model SF, but the possible improvement is not worth the increased model complexity.
The most practical conclusion of the research reported in this paper is to recommend model SF if minimizing the MSEP is the only concern. This model is intuitive, cf. (2.1), its estimation is straightforward, cf. (3.8) and (3.9), and the prediction equation is very simple, cf. (4.2).

It is somewhat surprising that taking the correlation of the residuals into account does not improve the predictions. We applied the test proposed by Gabrys and Kokoszka (2007) to check if the residual curves can be assumed to form an iid functional sequence. We applied it to the residual curves of the SF model which we recommend. The test yielded the rejection of the null hypothesis that these curves are iid elements of $L^2$. The rejection was at the 5% level in all cases, and at the 1% level in most cases. This rejection may be due to either the violation of the assumption of identical distribution or the assumption of the lack of correlation.

**Acknowledgements**

This research was partially supported by NSF grant DMS-0931948. We thank two referees and the Associate Editor for raising several important issues.
Functional prediction of intraday cumulative returns

References


Piotr Kokoszka and Xi Zhang


