Dynamic functional regression with application to the cross–section of returns

Piotr Kokoszka* Hong Miao† Matthew Reimherr ‡ Bahaeddine Taoufik §

Abstract

Motivated by testing the significance of risk factors for a cross–section of returns, we develop an inferential framework which involves function-on-scalar regression. Asymptotic theory is developed assuming the factors form a weakly dependent vector–valued time series, and the regression errors are weakly dependent functions. To accommodate the empirical behavior of the cross–section of returns and of the factors, we allow both the factors and the error functions can exhibit mild departures from stationarity. This requires new new asymptotic theory which leads to several tests for the significance of function–valued regression coefficients. The new approach to the study of the significance of risk factors for a cross–section of returns complements the established Fama–French approach based on portfolio construction. It is more suitable if the statistical significance of the risk factors is to be rigorously controlled.

JEL Classification: C32, G14, C53, C12,

Key words: Cross–section of returns, Functional regression, Hilbert Space, Weak dependence, Time series.

1 Introduction

Over the last forty years, the cross–section of returns has been one of the most extensively studied aspects of finance. To a large extent, the work of Fama and MacBeth (1973) set a paradigm which has been elaborated on in hundreds of papers, Harvey *et al.* (2014) provide a comprehensive account. Fama and MacBeth (1973) were motivated by testing the two–parameter portfolio model of Markowitz (1959) and its extensions. They introduced

*Department of Statistics, Colorado State University, Fort Collins, CO 80523
†Department of Finance and Real Estate, Colorado State University, Fort Collins, CO 80523
‡Department of Statistics, Pennsylvania State University, University Park, PA 16802
§Gabelli School of Business, Fordham University, New York, NY 10023

a technique based on recursive construction of portfolios and least squares testing for resulting time series regressions, a summary of this approach is provided in Appendix C.

The theory and methodology presented in this paper are motivated by the problem of testing the significance of risk factors which is at the core of the research on the cross–section of returns. Before presenting our model and theory in greater generality, we outline the motivation and the specific application to this testing problem. Denote by $R_n(i)$ the return on security i in period n. The Arbitrage Pricing Theory, see e.g. Campbell *et al.* (1997), postulates that

$$R_n(i) = \alpha(i) + \beta_m(i)M_n + \sum_{k=1}^{K} \beta_k(i)F_n(k) + \varepsilon_n(i), \tag{1.1}$$

where $R_n(i)$ and M_n are excess returns, respectively, on asset i and on a well diversified market portfolio. The temporal dynamics are modeled through M_n and other potential common factors $F_n(k)$. The errors $\varepsilon_n(i)$ are typically assumed to be martingale differences with nontrivial cross–sectional covariances. The interest is in testing which factors $F_n(k)$ are significant. Throughout the paper, the factors are considered known, they are either directly observed or constructed.

We now explain the essence of our approach. For each fixed i, one can obtain an estimate $\hat{\beta}_k(i)$. These estimates are extremely noisy, no matter what ordering of securities is used, as illustrated in Figure 1. This is one of the reasons why tests have been performed using suitably constructed portfolios, which involve averaging the $\hat{\beta}_k(i)$ over a handful of ranges of i, Appendix C. The resulting testing procedures contain many levels of uncertainty, which makes the assessment of the actual significance levels difficult. If the objective is merely to test $H_0 : \beta_k = 0$, i.e. that the function β_k is identically zero, an effective approach with a well controlled significance level can be developed as follows. In any time period n, the returns can be ordered from the lowest to the largest leading to smooth functions $R_n(t)$, where the set of values t is the same as the set of the securities i. The difference is that the curves $R_n(t)$ are nondecreasing, smooth functions of t, as shown in Figure 2. This way of looking at the cross–section of returns provides another approach to testing the significance of the risk factors. We replace model (1.1) with the model

$$R_n(t) = \alpha(t) + \beta_m(t)M_n + \sum_{k=1}^{K} \beta_k(t)F_n(k) + \varepsilon_n(t). \tag{1.2}$$

in which the functions $R_n(\cdot)$ are ordered returns. As we will see, in such a formulation, the estimates of the functions α, β_m and β_k are smooth, and it is possible to test their significance within a setting akin to linear regression. In model (1.1), for every security i, the coefficients are the same for every period. This is no longer the case in model (1.2), however the functions $\beta_k(\cdot)$ contain information about the significance of the risk

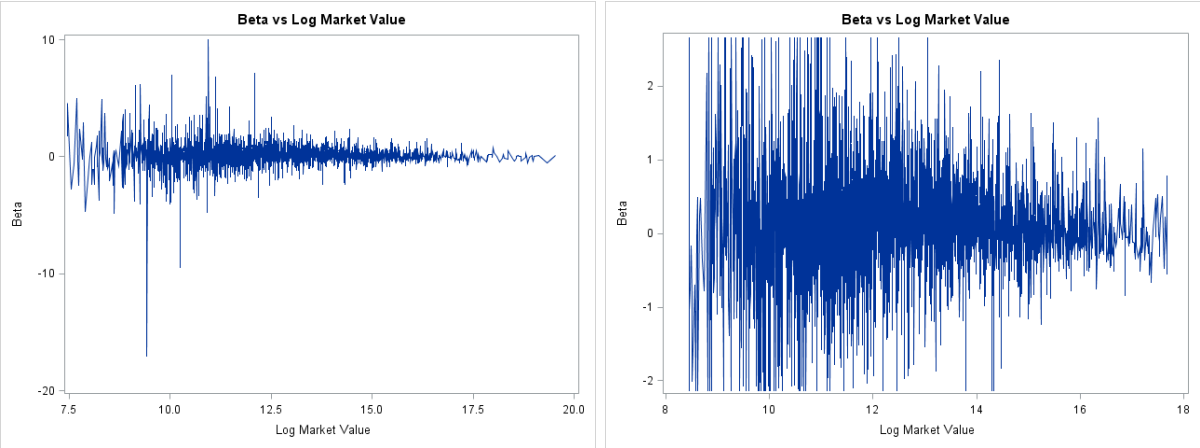

Figure 1: Estimated $\hat{\beta}_k(i)$ for all SP500 stocks i and for $F_n(k)$ equal to a volatility proxy. The stocks are ordered by log Market value for January, 2010. The left panel shows raw $\hat{\beta}_k(i)$, the values in the right panel are winsorized at 1% and 99%.

factors $F(k)$, which do not depend on the ordering of the returns. Compared to the portfolio approach described in Appendix C, this approach reverses the roles of returns and the betas; instead of sorting the assets by their estimated betas every month, they are sorted by the observed returns. The estimated smooth function $\hat{\beta}_k$ then tells us how best, mid- and worst performing stocks in a given month respond (over the long run) to the movements of the scalar factor $F_n(k)$ (if the function β_k is statistically different from zero).

The portfolio based approach to cross–sectional returns has many well–known advantages; ease of application and interpretation being perhaps the two most important ones. It does however have disadvantages. Conclusions may be sensitive to how the portfolios are constructed - they often do not hold for equal weighted portfolios. Second, most of the papers draw conclusions based on the statistical differences between the parameters of the top and bottom portfolios. Third, the relationship may be different within the portfolios. Our approach deals with the entire universe of the stock in the sample. At least theoretically, it reveals more information than a typical Fama–French approach. It must be emphasized that methods based on portfolio construction potentially provide insights beyond establishing the statistical significance of scalar factors. For example, the analysis in Ang *et al.* (2006) shows that a portfolio constructed from stocks with low sensitivities (large negative) to daily change in volatility in month n can be expected to yield a higher monthly return at the end of month $n+1$. This may imply a trading strategy of buying stocks with low sensitivities to market aggregate volatility and short selling stocks with high sensitivities to realize arbitrage profit. However, this kind of portfolio contains too many stocks (including many very illiquid ones) making such a trading strategy difficult

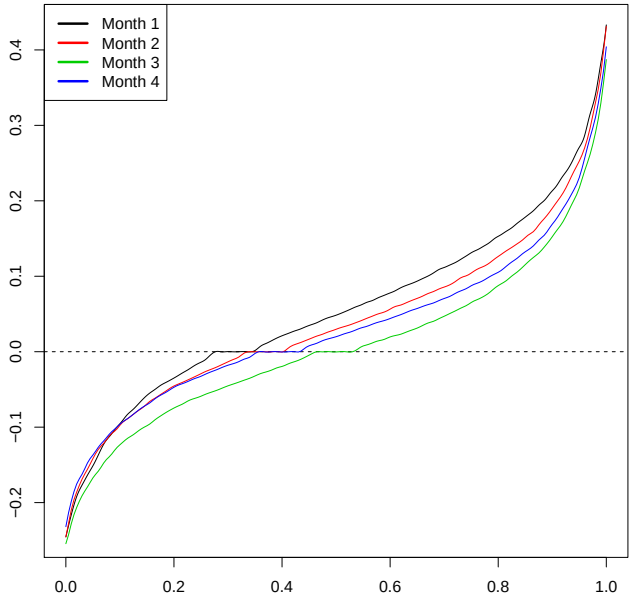

Figure 2: Ordered cross–sectional returns $R_n(t)$ for four consecutive months.

to realize. In conclusion, our approach provides at least a new angle on an important problem of finance research, and may have some advantages relative to well–established approaches.

This paper contributes to econometrics by developing estimation and testing theory in the context of a general regression which includes model (1.2). The factors $M_n, F_n(k)$ and the error functions ε_n do not have to be stationary; previous research assumes stationarity in models similar to (1.2), like the general model (2.1) introduced in Section 2. We treat smooth response curves as whole statistical entities, functions whose temporal dynamics are driven by the dynamics of the factors $F_n(k)$ which contribute to the level and shape of the response curves. The statistical methodology thus falls into the field of functional data analysis (FDA). The paper of Ramsay and Ramsey (2001) introduced FDA methodology to econometrics, but recent years have seen renewed interest in application of FDA to the econometric analysis of financial data, see Kargin and Onatski (2008), Müller *et al.* (2011), Horváth *et al.* (2014), Horváth and Rice (2015), among others. Several other applications of FDA to the econometric analysis of financial data are given or referenced in Horváth and Kokoszka (2012).

Recent contributions most closely related to our work include Hays *et al.* (2012) and Kokoszka *et al.* (2015), Gagliardini *et al.* (2016) and Renault *et al.* (2016). Hays *et al.*

(2012) consider the model

$$X_n(t) = \sum_{k=1}^{K} \gamma_{nk} F_k(t) + \varepsilon_n(t). \tag{1.3}$$

The factors F_k are *unknown functions* that do not depend on the period n. The dynamics are in the coefficients γ_{nk} which are assumed to follow Gaussian autoregressive processes. Model (1.3) is motivated by and applied to forecasting yield curves. Kokoszka *et al.* (2015) consider the regression

$$R_n(t) = \beta_0(t) + \sum_{j=1}^{p} \beta_j F_{nj}(t) + \varepsilon_n(t),$$

in which the factors $F_{nj}(t)$ are *known functions* which depend on n and the coefficients β_j are scalars. In their application, $R_n(t), F_{nj}(t)$ are intraday returns on assets on day n. In contrast to both (1.3) and (1), in (1.2) the factors are scalars and the coefficients are functions. Gagliardini *et al.* (2016) consider the regression

$$R_n(t) = \alpha_n(t) + \sum_{k=1}^{K} \beta_{nk}(t) F_n(k) + \varepsilon_n(t), \tag{1.4}$$

in which $t \in [0,1]$ represents a continuum of assets, so mathematically (1.2) is a special case of (1.4). Compared to (1.2), the coefficients of the factors $F_n(k)$ depend on the period n. Their objective is not to test the significance of these (sequences of) coefficients, but to estimate the temporal evolution of risk premia. Renault *et al.* (2016) consider a model similar to (1.4), $t \in [0,1]$ is the fraction t of assets. Their primary interest is in pricing idiosyncratic variance factors. They assume that the functional coefficients are cumulative integrals of "densities" defined on the interval $[0,1]$. We also note the work of Romano and Wolf (2013). While not concerned with functional regressions, it clarifies the issue of testing for monotonicity of functions, which has bearing on the problem of testing the significance of scalar risk factors.

The paper is organized as follows. Section 2 provides the formulation of the model and explains the relevant inference. In Section 3, we state assumptions under which our method is valid and provide the requisite asymptotic justification. Following a small simulation study in Section 4, the application to testing for significance of risk factors is presented in Section 5. Three appendices provide computational and mathematical details, as well as a summary of the portfolio formation methodology. Asymptotic theory for the GLS estimates introduced in Section 2 is presented in an on–line supplement.

2 Statistical model, estimation and testing

The general statistical model we consider is

$$Y_n(t) = x_{n1}\beta_1(t) + x_{n2}\beta_2(t) + \dots x_{np}\beta_p(t) + \varepsilon_n(t), \quad n = 1, 2, \dots, N. \tag{2.1}$$

The index n may denote day, week, month, quarter or year. The $\{x_{nk}, 1 \le n \le N\}$, $k = 1, 2, \dots, p$, are realizations of scalar time series which can exhibit temporal dependence, and even some mild departures from stationarity. In the applications in Section 5, the x_{nk} are the risk factors $F_n(k)$ in (1.2). Our objective in this section is to develop an estimation and testing methodology for the coefficient functions β_k.

Model (2.1) formally resembles the functional response model, see Faraway (1997), Chiou *et al.* (2004), Chapter 13 of Ramsay and Silverman (2005), Reimherr and Nicolae (2014), among many others. The currently available methodology and theory assumes either that the regressors x_{nk} are deterministic or are independent replications across n. Such assumptions are suitable for inference based on designed experiments. For time series regressors, like the risk factors, new methodology and asymptotic theory are called for.

The term functional regression also refers to a class of models in which regression coefficients depend on exogenous and/or lagged variables, see Fan and Zhang (1999) and many papers that followed. Despite sharing the same name, these two classes of regression models are very different, and are suitable for different problems.

In the development that follows, all random and deterministic functions are assumed to be in the real Hilbert space L^2 of square integrable functions with the inner product $\langle x, y\rangle = \int x(t)y(t)dt$. The domain of integration is a compact interval. It is usually assumed that it is the interval $[0, 1]$, and we will do so as well. In any application, the interval $[\min t, \max t]$ is transformed to $[0, 1]$ via an affine map. An introduction to the theory of random function in L^2 is presented in Chapter 2 of Horváth and Kokoszka (2012), for a more in–depth treatment see Hsing and Eubank (2015). Precise assumptions and asymptotic statements relevant to our context are provided in Section 3.

We begin by defining the vectors/matrices

$$Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_N(t) \end{bmatrix} \qquad \beta(t) = \begin{bmatrix} \beta_1(t) \\ \beta_2(t) \\ \vdots \\ \beta_p(t) \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & & \\ x_{N,1} & \dots & x_{N,p} \end{bmatrix} \qquad \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_N(t) \end{bmatrix}.$$

We then have the linear relationship $Y(t) = \mathbf{X}\beta(t) + \varepsilon(t)$. We consider two methods for estimating the vector valued function $\beta(t)$. The first is a least squares (LS) approach

which utilizes the sum of squared residuals

$$L_{LS}(\beta) = \sum_{n=1}^{N} \left\| Y_n - \sum_{k=1}^{p} x_{nk}\beta_k \right\|^2 = \int (Y(t) - \mathbf{X}\beta(t))^\top (Y(t) - \mathbf{X}\beta(t))\, dt.$$

The estimate $\hat{\beta}_{LS}(t)$ is the minimizer of $L_{LS}(\beta)$, and takes the familiar form

$$\hat{\beta}_{LS}(t) = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top Y(t). \tag{2.2}$$

The second method is an extension of an approach known as the Generalized Least Squares (GLS), which generally reduces the variance of estimators in the presence of dependence across statistical units. In our setting, the target function takes the form

$$L_{GLS}(\beta) = \int (Y(t) - \mathbf{X}\beta(t))^\top \mathbf{V}(t)(Y(t) - \mathbf{X}\beta(t))\, dt, \tag{2.3}$$

where $\mathbf{V}(t)$ is a matrix valued function of t. Practically, we use a data driven choice for $\mathbf{V}(t)$, which we denote as $\hat{\mathbf{V}}(t)$. As in scalar GLS, a good choice (and optimal under certain assumptions) for the weights is to use the inverse covariance of the error terms. We thus propose using $\mathbf{V}(t) \equiv \mathbf{W}^{-1}(t,t)$ where $W_{ij}(t,s) = \mathrm{Cov}(\varepsilon_i(t), \varepsilon_j(s))$; the inverse denotes a matrix inversion computed at each time point. The matrix $\mathbf{W}(t,s)$ can be estimated using a Bartlett type estimator and is discussed in Appendix A. The minimizer of $L_{GLS}(\beta)$ is denoted as $\hat{\beta}_{GLS}$ and takes the form

$$\hat{\beta}_{GLS}(t) = (\mathbf{X}^\top \hat{\mathbf{V}}(t)\mathbf{X})^{-1}\mathbf{X}^\top \hat{\mathbf{V}}(t)Y(t). \tag{2.4}$$

We note that smooth estimates can be obtained by incorporating a penalty (ridge regression). For example, the right–hand side of (2.2) could be replaced by $(\mathbf{X}^\top \mathbf{X} + \alpha\mathbf{P})^{-1}\mathbf{X}^\top Y(t)$, see e.g. Ramsay *et al.* (2009), but we do not pursue this approach because our focus is on testing the significance of the regression coefficients.

We now turn to the problem of testing if a particular β_k is the zero function. The two methods we propose work for any estimator $\hat{\beta}_k$ whose covariance function

$$C_k(t,s) = \mathrm{Cov}(\hat{\beta}_k(t), \hat{\beta}_k(s)) \tag{2.5}$$

can be estimated. We therefore present our approach in this general setting. In Appendix A, we explain how the covariance functions of the estimators $\hat{\beta}_{LS}(t)$ and $\hat{\beta}_{GLS}(t)$ are obtained.

The first approach is based on the L^2 norm. The test statistic is defined as

$$\widehat{\Lambda}_k = N\|\hat{\beta}_k\|^2 = N \int \hat{\beta}_k^2(t)dt.$$

When $\beta_i(t) \equiv 0$ (under the assumptions of Theorem 3.1), this statistic has approximately a weighted chi–square distribution,

$$\widehat{\Lambda}_k \sim \sum_{j=1}^{N} \hat{\lambda}_{k,j} \chi_j^2(1), \tag{2.6}$$

where $\hat{\lambda}_{k,j}$ are the eigenvalues of $\widehat{C}_k$, an estimate of the C_k given by (2.5). One can only handle the right hand side above numerically, but tools such as the `imhoff` function in `R`, can be used to obtain P-values/critical values. The asymptotic justification for the approximation (2.6) is given in Corollary 3.1. The second approach is based on functional principal components, see e.g. Chapter 3 of Horváth and Kokoszka (2012). Denote by $\hat{v}_{k,1}, \hat{v}_{k,2}, \ldots$ the eigenfunctions of $\hat{C}_k$ corresponding to the eigenvalues $\hat{\lambda}_{k,1} > \hat{\lambda}_{k,2} > \ldots$, and set

$$\widehat{T}_{k,q} = N \sum_{j=1}^{q} \frac{\langle \hat{\beta}_k, \hat{v}_{k,j} \rangle^2}{\hat{\lambda}_{k,j}}.$$

Corollary 3.1 shows that for a fixed q,

$$\widehat{T}_{k,q} \sim \chi^2(q). \tag{2.7}$$

In practice, the truncation level q is chosen as the smallest integer such that $\sum_{j=1}^{q} \hat{\lambda}_{k,j} > 0.85 \sum_{j=1}^{N} \hat{\lambda}_{k,j}$. We thus use a data driven q, and the validity of this choice is justified by a simulation study. The asymptotic theory is used to derived the approximation (2.7), which is then enhanced by a data driven procedure.

To summarize, we will work with the following four test statistics:

$$\begin{aligned}
\widehat{\Lambda}_{LS,k} &= N \int \hat{\beta}_{LS,k}^2(t)\, dt, && (\text{LS}{-}L^2);\\
\widehat{T}_{LS,k,q} &= N \sum_{j=1}^{q} \frac{\langle \hat{\beta}_{LS,k}, \hat{v}_{k,j} \rangle^2}{\hat{\lambda}_{k,j}}, && (\text{LS} - \text{PCA});\\
\widehat{\Lambda}_{GLS,k} &= N \int \hat{\beta}_{GLS,k}^2(t)\, dt, && (\text{GLS}{-}L^2);\\
\widehat{T}_{GLS,k,q} &= N \sum_{j=1}^{q} \frac{\langle \hat{\beta}_{GLS,k}, \tilde{v}_{k,j} \rangle^2}{\tilde{\lambda}_{k,j}}, && (\text{GLS} - \text{PCA}).
\end{aligned}$$

The $\hat{\lambda}_{k,j}, \hat{v}_{k,j}$ in the LS statistics are the eigenelements of $\widehat{C}_{LS,k}$, while the $\tilde{\lambda}_{k,j}, \tilde{v}_{k,j}$ in the GLS statistics are the eigenelements of $\widehat{C}_{GLS,k}$. The construction of the estimators $\widehat{C}_{LS,k}$ and $\widehat{C}_{GLS,k}$ is explained in Section A.

The tests based on the PCA–statistics reject $H_0 : \beta_k = 0$ if the statistic, $\widehat{T}_{k,q}$ exceeds a critical value of the chi–square distribution with q degrees of freedom. To apply the tests based on the LS–statistics, an empirical distribution of the right–hand side of (2.6) must be created for each specific data set. The null hypothesis is rejected if $\widehat{\Lambda}_k$ exceeds a critical value of this empirical distribution.

3 Asymptotic Theory

A key to establishing relations (2.6) and (2.7) is a central limit theorem for sums involving the factors x_{nk} and the error functions ε_n, cf. (2.2) and (2.4). The factors considered in Section 5 show mild departures from stationarity (the volatility outlier representing the October 1987 crash is handled separately). Many econometric time series exhibit similar departures from stationarity, so central limit theorems for nonstationary sequences have received a fair deal of attention in the past decades. Convergence to the normal distribution has been established under various quantifications of the degree of nonstationarity and dependence. The often cited work of Herrndorf (1984) considered α–mixing sequences, Davidson (1993) used near–epoch dependence, Neumann (2013) used the weak dependence concept of Doukhan and Louhichi (1999). This list is by no means exhaustive. In our approach, we use the concepts of a Bernoulli shift representation and approximability, which have been used extensively in recent theoretical work on nonlinear time series, e.g. see Pötscher and Prucha (1997), Wu (2005), Shao and Wu (2007), Aue *et al.* (2009), Hörmann and Kokoszka (2010), Hörmann *et al.* (2013) and Kokoszka and Reimherr (2013). We however allow some nonstationarity. Precise definitions together with a suitable central limit theorem are stated in Section 3.1. Using these results, we establish in Section 3.2 the required asymptotic results for model (2.1), and provide asymptotic justification for the tests based on relations (2.6) and (2.7).

3.1 Central limit theorem for nonstationary approximable sequences

Motivated by the problem studied in this paper and to enhance general applicability to functional time series, we establish a CLT in a general separable Hilbert space $\mathbb{H}$. Let $p \geq 1$ and let $L^p_{\mathbb{H}}$ be the space of $\mathbb{H}$–valued random elements X such that

$$\nu_p(X) = \big(E\|X\|^p\big)^{1/p} < \infty.$$

Definition 3.1 We say that a sequence, $\{Z_n\}$, in a real separable Hilbert space $\mathbb{H}$ is *weakly L^p–approximable* if the following conditions hold:

1. There exists a sequence $\{u_n\}$ of iid elements in an abstract measurable space $\mathcal{U}$ such that
$$Z_n = f_n(u_n, u_{n-1}, \dots),$$
for measurable functions $f_n : \mathcal{U}^\infty \to \mathbb{H}$.
2. For each integer $M > 0$, consider an approximating sequence $Z_{n,M}$ defined by
$$Z_{n,M} = f_n(u_n, u_{n-1}, \dots, u_{n-M}, u^\star_{n-M-1}, u^\star_{n-M-2}, \dots),$$

where the sequences $\{u_n^\star\} = \{u_n^\star(n,m)\}$ are copies of $\{u_n\}$ independent across m and n and independent of the original sequence $\{u_n\}$.

The following condition holds

$$\sum_{M=1}^{\infty} \sup_{1\le n<\infty} \nu_p(Z_n - Z_{n,M}) < \infty. \tag{3.1}$$

Before proceeding, we comment on Definition 3.1. Condition 1 states that every observation Z_n is a, in general nonlinear, moving average. The Z_n are dependent, because they are driven by the same sequence of errors, but they do not have to share the same distributions because the functions f_n may be different. The construction of the approximating sequences $Z_{n,M}$ in condition 2 is always possible, the nontrivial part of this condition is requirement (3.1). We want the distance ν_p between the observations Z_n and their M–dependent approximations $Z_{n,M}$ to be small: the errors u_n far back in the past make a contribution so small that they can be replaced by independent copies. In previous research, the functions f_n have been assumed to be equal, $f_n = f$, implying that $\{Z_n\}$ is stationary. Definition 3.1 permits some degree of nonstationarity; the class of weakly L^p–approximable sequences includes the L^p–m–approximable sequences introduced in Hörmann and Kokoszka (2010), including functional linear processes, Bosq (2000), and functional ARCH sequences, Hörmann *et al.* (2013). We also note the concept of sequences *weakly $\mathcal{M}$–dependent in L^p* introduced by Berkes *et al.* (2011). Their definition also allows some degree of nonstationarity, even though stationarity is imposed as an additional assumption in the statements of their results.

The main contribution of this section is Theorem 3.1. Its statement involves the space, $\mathcal{N}$, of Nuclear (or trace class) operators, a concise account is given in Section 13.5 of Horváth and Kokoszka (2012). For any $z_1, z_2 \in \mathbb{H}$, $z_1 \otimes z_2$ is the operator defined by $z_1 \otimes z_2(u) = \langle z_1, u\rangle\, z_2$.

THEOREM 3.1 *Assume that $\{Z_n\}$ is a zero mean weakly L^2–approximable sequence in a real separable Hilbert space $\mathbb{H}$ such that*

1. $\sup_{1\le n<\infty} E\|Z_n\|^4 < \infty$,
2. $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} E(Z_i \otimes Z_j) \overset{\mathcal{N}}{\to} \Gamma$, *and*
3. *for any fixed* $0 < M_1 \le M_2 < N$

$$N^{-1}\sum_{i=1}^{\lfloor N/M_1\rfloor}\sum_{j=1}^{M_2}\sum_{j'=1}^{M_2} E[Z_{(i-1)M_1+j} \otimes Z_{(i-1)M_1+j'}] \overset{\mathcal{N}}{\to} C_{M_1,M_2}.$$

Then,

$$N^{-1/2}\sum_{n=1}^{N} Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Gamma). \tag{3.2}$$

Theorem 3.1 is proven in Appendix B. Condition 1 replaces the usual fourth moment condition; the supremum is needed because the sequence $\{Z_n\}$ need not be stationary. Condition 2 states that the long–run covariance operator exists and is a nuclear operator (all covariance operators are nuclear). Condition 3 is more technical, but essentially implies that blocks of observations can be systematically dropped and one still has convergence of the covariance operators. Lastly, since these are infinite dimensional objects, tightness is always a central concern. Here tightness is guaranteed by the first condition using a result found in Chen and White (1998).

3.2 Large sample theory for the tests of Section 2

We now apply Theorem 3.1 to derive the asymptotic distributions of $\hat{\beta}_{LS}$. Corresponding results for $\hat{\beta}_{GLS}$ can be obtained, but are quite technical. We have thus moved those results to a supplement, and refer the interested reader there. We begin by considering the least squares estimator. Notice that

$$\sqrt{N}(\hat{\beta}_{LS}(t)-\beta(t)) = (N^{-1}\mathbf{X}^\top\mathbf{X})^{-1}N^{-1/2}\mathbf{X}^\top\varepsilon(t).$$

Under very weak conditions $N^{-1}\mathbf{X}^\top\mathbf{X} \xrightarrow{P} \Sigma_X$, and this will be our assumption. Thus our primary objective is to show that $N^{-1/2}\mathbf{X}^\top\varepsilon(t)$ is asymptotically Gaussian. We have that

$$N^{-1/2}\mathbf{X}^\top\varepsilon(t) = N^{-1/2}\sum_{n=1}^{N} X_n\varepsilon_n(t).$$

Here X_n is vector and $\varepsilon_n(t)$ multiplies each coordinate of X_n. Thus, Theorem 3.1 with $Z_n(t) := X_n\varepsilon_n(t)$ leads to the following Corollary.

COROLLARY 3.1 *If the sequence of functions $Z_n(t) := X_n\varepsilon_n(t)$ satisfies the conditions of Theorem 3.1 and*

$$N^{-1}\mathbf{X}^\top\mathbf{X} \xrightarrow{P} \Sigma_X,$$

then

$$\sqrt{N}(\hat{\beta}_{LS}-\beta) \Rightarrow \mathcal{N}(0, C_{LS}). \tag{3.3}$$

Furthermore when $\beta_k = 0$, then the asymptotic distributions of $\widehat{\Lambda}_{LS,k}$ and $\widehat{T}_{LS,k,q}$ are given in (2.6) and (2.7), respectively. When $\beta_k \neq 0$, $\widehat{\Lambda}_{LS,k} \xrightarrow{P} \infty$. When $\langle \beta_k, v_j\rangle \neq 0$ for some $j \leq q$, then $\widehat{T}_{LS,k,q} \xrightarrow{P} \infty$ as well.

PROOF: Relation (3.3) follows directly from Theorem 3.1. Statements on relations (2.6) and (2.7) follow from standard results on transformations of normal distributions. The consistency of the tests is also easy to establish, if $\beta_k \neq 0$, then $\widehat{\Lambda}_k \sim N \int \beta_k^2(t)dt$, with only a slightly more complex argument applying to $\widehat{T}_{k,q}$.

We emphasize that the assumption that the functional sequence $Z_n(t) := X_n\varepsilon_n(t)$ satisfies the conditions of Theorem 3.1 is very general and includes practically all reasonable assumptions under which asymptotic normality of the LS and GLS can be expected in our context. We now take a closer look at some specific special cases.

The functions $Z_n(t) := X_n\varepsilon_n(t)$ are random elements of the Hilbert space $\mathbb{H} = (L^2)^p$ with the inner product

$$\langle z_1, z_2 \rangle = \sum_{k=1}^{p} \int z_{1k}(t) z_{2k}(t) dt.$$

We assume that the regressors $X_n = [x_{n1}, x_{n2}, \ldots, x_{np}]^\top$ are independent of the the error functions ε_n. Then,

$$E \left\| Z_n \right\|^4 = E \left\| X_n \right\|^4 E \left\| \varepsilon_n \right\|^4 = E \left\{ \sum_{k=1}^{p} x_{nk}^2 \right\}^2 E \left\{ \int \varepsilon_n^2(t) dt \right\}^2.$$

Thus condition 1 of Theorem 3.1 is implied by

$$\sup_{n \geq 1} E \left\{ \sum_{k=1}^{p} x_{nk}^2 \right\}^2 < \infty \quad \text{and} \quad \sup_{n \geq 1} E \left\{ \int \varepsilon_n^2(t) dt \right\}^2 < \infty.$$

If we assume, in addition, that the sequences $\{X_n\}$ and $\{\varepsilon_n\}$ are stationary, condition 1 is implied by $E \left\| X_0 \right\|^4 < \infty$ and $E \left\| \varepsilon_0 \right\|^4 < \infty$.

To obtain a simple special case under which condition 2 is satisfied, suppose the ε_n are iid. It is easy to see that then $E(Z_i \otimes Z_j) = 0$ if $i \neq j$, so that condition 2 requires that $N^{-1} \sum_{i=1}^{N} E(Z_i \otimes Z_i)$ converges to a covariance operator acting on $\mathbb{H}$. This in turn is implied by the requirement that $N^{-1} \sum_{i=1}^{N} E[X_i X_i^\top]$ converges. If the X_i form a stationary sequence, this condition always holds because the average is equal to $E[X_0 X_0^\top]$ for every N.

Condition 3 is also easy to interpret if the ε_n are iid. In the case of two blocks of equal length, $M_1 = M_2 = N/2 = M$, it requires that the operator

$$M^{-1} \sum_{j=1}^{M} E(Z_j \otimes Z_j) + M^{-1} \sum_{j=1}^{M} E(Z_{M+j} \otimes Z_{M+j})$$

converges, as $M \to \infty$. This is implied by the convergence of

$$M^{-1} \sum_{j=1}^{M} E(X_j X_j^\top) + M^{-1} \sum_{j=1}^{M} E(X_{M+j} X_{M+j}^\top).$$

An analogous condition can be formulated for a larger number of blocks, and again holds trivially if the X_n form a stationary sequence. Condition 3 thus intuitively states that while the regressors can be nonstationary, their covariance matrices must not oscillate too much

Various other technical assumptions could be specified, but this is not our objective. The essence of Corollary 3.1 is that asymptotic normality holds under very general assumptions, and what matters are the properties of the time series $X_n\varepsilon_n(\cdot)$.

4 A simulation Study

In this section, we conduct a simulation study in order to evaluate statistical power and calibration of the least squares and the generalized least squares procedures introduced in previous sections. A broad conclusion is that in the presence of temporal correlation the GLS procedures have an advantage over the LS procedures in terms of empirical size and power and the RMSE of the the estimators. In the absence of temporal correlations, all procedures are comparable. A general recommendation is to use the GLS-L^2 procedure as a default.

Data Generating Processes: We generate data using the model

$$Y_n(t) = \beta_1(t)X_{n1} + \beta_2(t)X_{n2} + \beta_3(t)X_{n3} + \varepsilon_n(t), \quad 0 \leq t \leq 1,$$

with $\beta_1(t) = c\sin(2\pi t)$, $\beta_2(t) = t$ and $\beta_3(t) = t^2$. We will test whether $\beta_1 = 0$. The constant c allows us to control the size of β_1 relative to the size of the remaining coefficients. The null hypothesis corresponds to $c = 0$. Increasing c, makes the distance between the alternative and null hypotheses larger.

We consider two types of distributions for the X_{nj}. To explore the performance of the tests in a standard setting, we take $X_{nj} \overset{iid}{\sim} N(0,1)$, $1 \leq j \leq 3$. To imitate the real data used in Section 5, we draw bootstrap samples from the actual factors. Specifically, the X_{n1} follow the bootstrap distribution of MKTRF, the X_{n2} of HML and the X_{n3} of FVIX. These factors are described in Section 5 and plotted in Figure 4. It is seen that their spread is very different than the spread of the standard normal distribution, so each bootstrap sample has been normalized to have mean 0 and variance 1. This is needed in order to compare power using the same range of $c > 0$, which quantify the magnitude of departure from the null hypothesis.

Temporal and cross–sectional dependence is introduced by using error curves defined as

$$\varepsilon_n(t) = \int_0^1 \alpha(t,s)\varepsilon_{n-1}(s)ds + \delta_n(t), \quad t \in [0,1]$$

where $\alpha(t,s) = \gamma e^{-(t-s)^2}/0.8739$, so the Hilbert–Schmidt norm is γ. We use $\gamma = 0.5$ (the errors are correlated) and $\gamma = 0$ (the errors are iid). The $\delta_n(t)$ are iid Gaussian random functions with the Matérn covariance function

$$C(t,s) = \left(1 + \frac{\sqrt{5}|t-s|}{\rho} + \frac{5|t-s|^2}{3\rho^2}\right) \exp\left(-\frac{\sqrt{5}|t-s|}{\rho}\right).$$

We use $\rho = 1/4$.

Each curve is generated at one thousand equally spaced time points between 0 and 1. As for the return data studied in Section 5, the artificial data were converted to functional objects using cubic B-splines with $K = 20$ basis functions. We use sample size $N = 100, 200, 500$ and 10E3 replications for each scenario. The covariance of $\hat{\beta}_1$ is estimated using the method explained in Appendix A using $h = 1/3$. We also ran simulations using $h = 1/5$. The rejection rates are very close to those obtained using $h = 1/3$.

Empirical results: We begin by examining the type 1 error, i.e. the rows corresponding to $c = 0$. All methods have empirical size within 2-3 standard deviations (0.007) of 0.05, with some indication of conservative size for the GLS methods if $\gamma = 0.5$, but only in the case if normal factors. This may thus be due to random variability. Turning to power, we observe that it is smaller in case of the bootstrap distribution (Table 2 than in case of the normal distribution. This is not surprising because the normally distributed factors are an idealized case. We see that for smaller alternatives the PCA method does better than the L^2 method, with an average gain in power of about 10-20% depending on the row. However, the methods equalize for larger alternatives. Under independence, the power is about the same for both LS and GLS. However, in the dependent case, $\gamma = 0.5$, for smaller alternatives we see on average about a 20-30% increase in power when using GLS. This power gain occurs despite a lower empirical size. Regarding the RMSE of the estimators, it is about 0.07 in all cases, except the case of the LS estimator applied to the DGP with $\gamma = 0.5$. In that case, as expected, the RMSE is slightly larger, about 0.08. Based on our small simulation study, we see that all methods have comparable performance, with the PC methods offering power advantage. The LS-PC method emerges as a simple and accurate test.

	$\gamma = 0.5$				$\gamma = 0$			
	LS-L^2	LS-PC	GLS-L^2	GLS-PC	LS-L^2	LS-PC	GLS-L^2	GLS-PC
	$N = 100$							
$c = 0.00$	0.06	0.06	0.06	0.06	0.07	0.06	0.08	0.07
$c = 0.05$	0.08	0.08	0.09	0.09	0.08	0.08	0.09	0.09
$c = 0.10$	0.12	0.17	0.16	0.21	0.15	0.19	0.15	0.20
$c = 0.15$	0.24	0.31	0.33	0.39	0.28	0.34	0.30	0.36
$c = 0.20$	0.43	0.54	0.56	0.61	0.51	0.58	0.54	0.59
$c = 0.25$	0.68	0.75	0.79	0.79	0.75	0.77	0.76	0.78
$c = 0.30$	0.86	0.87	0.93	0.90	0.90	0.90	0.90	0.90
	$N = 200$							
$c = 0.00$	0.05	0.05	0.04	0.03	0.04	0.04	0.04	0.05
$c = 0.05$	0.07	0.10	0.08	0.10	0.09	0.10	0.09	0.10
$c = 0.10$	0.20	0.27	0.27	0.33	0.25	0.31	0.26	0.31
$c = 0.15$	0.48	0.57	0.58	0.66	0.56	0.62	0.57	0.64
$c = 0.20$	0.79	0.83	0.87	0.89	0.86	0.86	0.86	0.87
$c = 0.25$	0.95	0.95	0.97	0.97	0.97	0.96	0.97	0.97
$c = 0.30$	0.99	0.99	1.00	0.99	1.00	0.99	1.00	0.99
	$N = 500$							
$c = 0.00$	0.05	0.05	0.04	0.05	0.05	0.06	0.06	0.06
$c = 0.05$	0.13	0.2	0.19	0.25	0.17	0.22	0.18	0.23
$c = 0.10$	0.55	0.65	0.67	0.75	0.67	0.70	0.67	0.70
$c = 0.15$	0.94	0.95	0.97	0.97	0.97	0.96	0.97	0.96
$c = 0.20$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$c = 0.25$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$c = 0.30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1: The LS-L^2, LS-PC, GLS-L^2, GLS-PC columns correspond to rejection rates for the different combinations of LS and GLS estimation procedures combined with a squared integral (L^2) or PCA test statistics. The first row, $c = 0.00$, is the empirical type 1 error, while the subsequent rows correspond to empirical power. Dependent errors correspond to $\gamma = 0.5$ and independent errors correspond to $\gamma = 0$.

	$\gamma = \frac{1}{2}$				$\gamma = 0$			
	LS-L^2	LS-PC	GLS-L^2	GLS-PC	LS-L^2	LS-PC	GLS-L^2	GLS-PC
	$N = 100$							
$c = 0.00$	0.05	0.06	0.05	0.07	0.05	0.07	0.06	0.07
$c = 0.05$	0.06	0.06	0.06	0.08	0.05	0.07	0.06	0.08
$c = 0.10$	0.08	0.09	0.09	0.10	0.08	0.10	0.09	0.12
$c = 0.15$	0.11	0.15	0.14	0.18	0.12	0.17	0.14	0.18
$c = 0.20$	0.17	0.24	0.23	0.28	0.20	0.27	0.22	0.29
$c = 0.25$	0.25	0.35	0.37	0.42	0.32	0.40	0.35	0.40
$c = 0.30$	0.40	0.49	0.51	0.55	0.48	0.58	0.49	0.56
	$N = 200$							
$c = 0.00$	0.05	0.06	0.06	0.06	0.06	0.06	0.07	0.06
$c = 0.05$	0.07	0.09	0.09	0.09	0.08	0.09	0.09	0.09
$c = 0.10$	0.11	0.15	0.15	0.19	0.13	0.18	0.13	0.18
$c = 0.15$	0.21	0.29	0.28	0.36	0.26	0.34	0.27	0.35
$c = 0.20$	0.39	0.47	0.49	0.56	0.47	0.52	0.47	0.53
$c = 0.25$	0.60	0.66	0.70	0.74	0.68	0.71	0.69	0.72
$c = 0.30$	0.81	0.82	0.87	0.86	0.87	0.86	0.87	0.85
	$N = 500$							
$c = 0.00$	0.04	0.05	0.05	0.06	0.05	0.05	0.05	0.05
$c = 0.05$	0.09	0.10	0.09	0.10	0.08	0.10	0.09	0.11
$c = 0.10$	0.22	0.28	0.28	0.38	0.26	0.32	0.27	0.32
$c = 0.15$	0.53	0.61	0.65	0.71	0.63	0.66	0.64	0.67
$c = 0.20$	0.85	0.87	0.92	0.92	0.92	0.90	0.92	0.91
$c = 0.25$	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99
$c = 0.30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: This table shows information analogous to that shown in Table 1, but with non–normal data generated using the bootstrap procedure described in this section.

5 Application to cross–sectional returns

In this section, we apply the tests introduced in Section 2 and asymptotically justified in Section 3 to the problem of identifying the risk factors, which we introduced in Section 1. The response curves Y_n in (2.1) are the ordered return curves R_n in (1.2) in month n, adjusted for the risk free return. In other words, $Y_n(t)$ is the tth percentile among all excess returns in month n, e.g. $Y_n(0.5)$ is the median excess return in month n. Following the usual practice of empirical finance, the returns were trimmed at the top and bottom 2.5% before computing the Y_n, plots are shown in Figure 3.

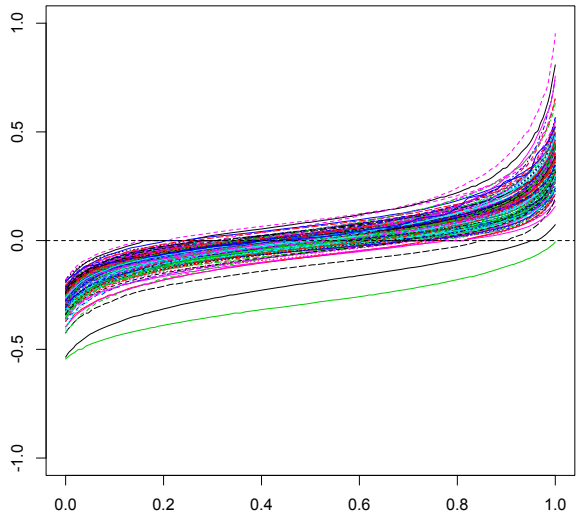

Figure 3: Ordered monthly return curves, with 5% trim.

In our application, regression (1.2) takes the form

$$\begin{aligned} Y_n(t) - \mathrm{RF}_n =& \alpha(t) + \beta_1(t)\mathrm{MKTRF}_n + \beta_2(t)\mathrm{SMB}_n \\ &+ \beta_3(t)\mathrm{HML}_n + \beta_4(t)\mathrm{UMD}_n + \beta_5(t)\mathrm{FVIX}_n + \varepsilon_n(t), \end{aligned} \tag{5.1}$$

with the factors defined as follows:

RF	Risk free return (US Treasury bill),
MKTRF	SP500 return less risk free return,
SMB	Small Minus Big Fama–French factor,
HML	High Minus Low Fama–French factor,
UMD	Up Minus Down Fama–French factor,
FVIX	VIX proxy (volatility proxy).

In this setting, we want to investigate the contemporary relationship between the excess return curves and the same five factors as in Ang *et al.* (2006). Our approach can be applied to any similar factor model. We choose the present setting as an illustration.

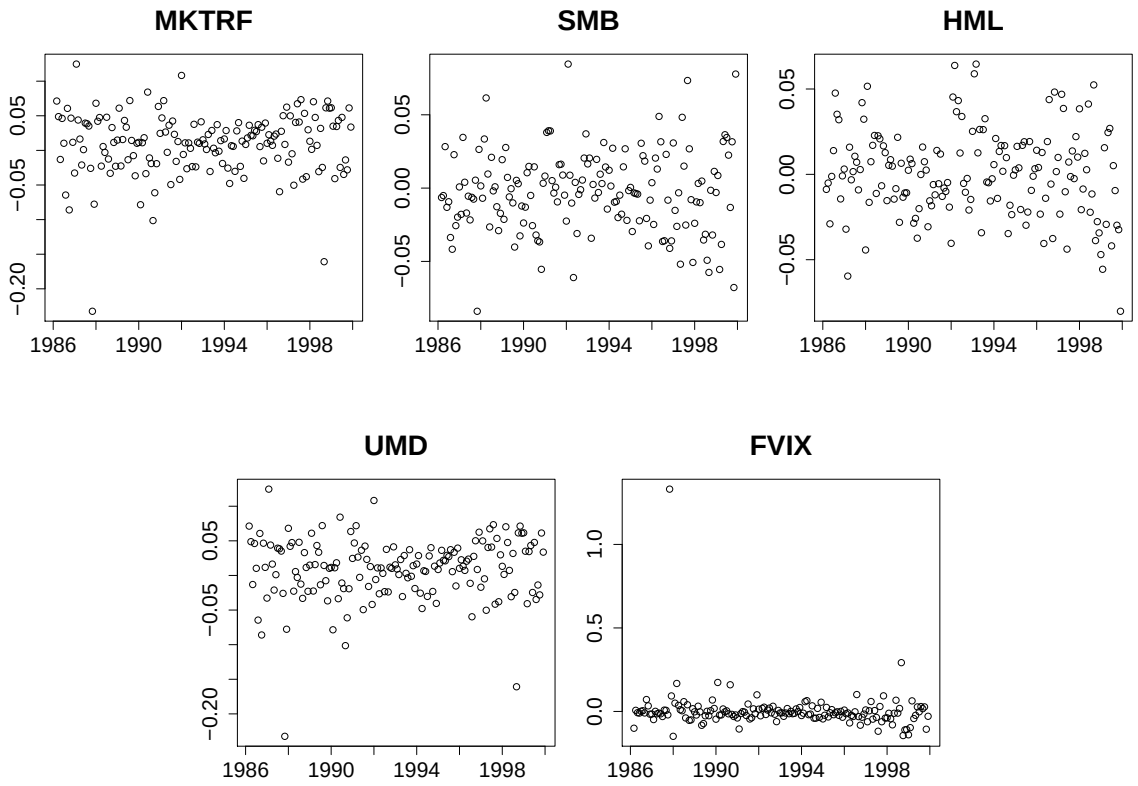

Figure 4: Time series plots of the factors in regression (5.1)

.

The above factors are among the most commonly studied risk factors. The MKTRF, SMB, and HML are the factors of the famous Fama–French three factor model defined in Fama and French (1992), the UMD factor (the momentum factor) was first introduced in Carhart (1997). Those four are the most commonly used factors in the cross sectional literature. The FVIX factor was introduced by Ang *et al.* (2006). It is constructed to mimic the aggregate volatility risk, and is argued to be a good proxy for it. All factors are indexed by the month, n, in the range from 02/1986 to 11/1999, i.e. $N = 166$. Their plots are shown in Figure 4.

Estimated β functions and P-values are displayed, respectively, in Figure 5 and Table 3. Table 3 shows that all factors are significant, Figure 5 leads to the following conclusions. Positive (negative) movements in the market index and the two most commonly used Fama–French factors, SMB and HML, correlate with the movements of individual shares in the same direction, with the effect being the strongest for stocks that perform best in the given month (for MKTRF and SMB factors) or the worst (for the HML factor). The shape of the function β_1 is consistent with the theory of Markowitz. The UMD factor

is negatively correlated with contemporaneous returns. The FVIX factor exhibits the most interesting impact: increased volatility is associated with higher returns for best performing stocks and with lower returns for the worst performing stocks (in a given month). Even though we used five important factors, they explain only a part of cross–sectional returns, the intercept is not a zero function. This agrees with studies which used methods based on portfolio construction.

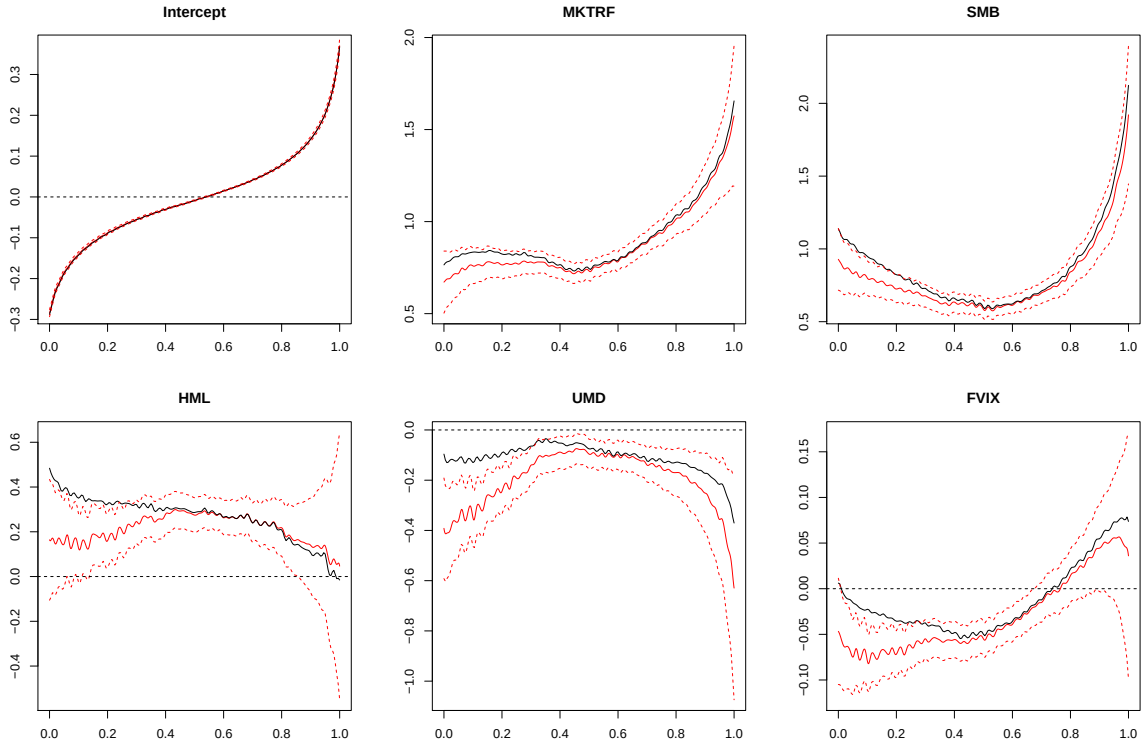

Figure 5: Plots of estimated β curves in model (5.1). Red and black curves indicate GLS and LS estimates, respectively. Red dashed lines show point–wise confidence intervals for the GLS estimate.

A close look of the shape of the beta curves reveals rich information. First, there is a positive relation between the market excess return and the sorted excess return curve, and the shapes of the excess return curves and the shape of the MKTRF beta curve are very similar. Second, although there is a positive relationship between the excess return curve and the SMB factor – the betas are all positve and above 0.5, the shapes for the beta curve and the excess return curves are very different. The SMB beta curve is a skewed U shape. That is middle excess returns are less correlated to SMB than excess returns at the high and low end. Third, the shape of the HML beta is almost the opposite of the excess return curves. That is, the correlation between the low excess returns and HML is higher than the correlation between high excess returns and HML. Fourth, the beta

of the Momentum (UMD) shows a relatively flat shape with step down slow on the high end. In particular, the beta values are much higher for the high end excess return than the low end (roughly the highest is about 4 times the lowest). That is the high excess return stocks are more sensitive to the contemporary UMD factor. Fifth, the relationship between the excess return and the FVIX factor is quite interesting. The low and middle excess returns are negatively correlated with the FVIX factor; the high excess returns are however positively correlated to the FVIX factor. This is consistent with Ang *et al.* (2006) who find negative correlation between low return portfolios and FVIX factor and high positive correlation between high return portfolios and FVIX factor.

	LS-L^2	LS-PCA	GLS-L^2	GLS-PCA
Intercept	5.82e-05	0	2.43e-03	0
MKTRF	7.10e-08	5.48e-105	4.01e-09	0
SMB	5.09e-04	1.32e-88	1.50e-05	6.43e-60
HML	7.56e-04	8.25e-06	1.95e-04	8.10e-08
UMD	5.01e-02	1.76e-02	2.76e-02	8.50e-03
FVIX	4.21e-02	1.66e-01	2.81e-02	2.97e-08

Table 3: P-values for LS and GLS estimates in Figure 5. The tests are based on the four statistics listed in Section 2.

We performed the above analysis with the October 1987 removed. The P–values and the curves turned out to be only minimally different, but the significance statements were not affected. Finite sample performance of the tests was evaluated by a simulation study. For data resembling the cross section of returns, all methods have comparable empirical sizes, in the range 4.5-6.0 % at the nominal level of 5%, with the GLS methods exhibiting some power advantage. The conclusions of the application presented in this section do not depend on the specific method used. As emphasized in the Introduction, this paper focuses on theory and methodology; a more comprehensive empirical study is needed to draw more general conclusions.

A Covariance structure estimation

To understand the derivation of the weighted loss function (2.3), it is instructive to first consider the model $y_n = \sum_{k=1}^{p} x_{nk}\beta_k + \varepsilon_n$, in which all quantities are scalars. The $N \times N$ covariance matrix $\mathbf{\Sigma}$ with entries $\mathbf{\Sigma}[i,j] = \mathrm{Cov}(\varepsilon_i, \varepsilon_j)$ is diagonalized via $\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top = \mathbf{I}$, where $\mathbf{I}$ is the $N \times N$ identity matrix. The transformed errors $\eta_n = \sum_{j=1}^{N} A_{nj}\varepsilon_j$ are uncorrelated, and the GLS estimators are obtained as the LS estimators computed after transforming the original regression by the matrix $\mathbf{A}$. (The matrix $\mathbf{\Sigma}$ is estimated

recursively, Cochrane and Orcutt (1949), Chapter 5 of Shumway and Stoffer (2011).) If the errors are functions, their second order structure is described by the covariances

$$W_{i,j}(t,s) = \mathrm{Cov}(\varepsilon_i(t), \varepsilon_j(s)). \tag{A.1}$$

If $t = s$, the entries $W_{i,j}(t,t)$ form an $N \times N$ covariance matrix, but if $t \neq s$, the matrix $\mathbf{W}(t,s)$ is no longer symmetric nor positive definite, and the Cholesky decomposition used in the scalar case no longer exists. For a fixed t, the GLS approach can however be applied, and is equivalent to the minimization of

$$L_{GLS}(\beta(t);t) = (Y(t) - \mathbf{X}\beta(t))^\top \mathbf{W}^{-1}(t,t)(Y(t) - \mathbf{X}\beta(t)).$$

Based on this argument, we use $\mathbf{V}(t) = \mathbf{W}^{-1}(t,t)$. The integration in (2.3) thus represents averaging over all t's, and leads to a loss function which does not depend on t.

In the actual implementation, we replace $\mathbf{W}(t,t)$ by an estimate $\widehat{\mathbf{W}}(t)$. Estimates $\widehat{\mathbf{W}}(t)$ can be constructed in a number of ways. We use a nonparametric approach. Since the $\varepsilon_i(t)$ are not observed, they must be replaced by residuals $\hat{\varepsilon}_i(t)$, but without knowing $\widehat{\mathbf{W}}(t)$ these residuals cannot be computed. When calculating $\widehat{\mathbf{W}}$ for the LS procedure we use

$$\widehat{R}_\ell(t) = \frac{1}{N-p-\ell} \sum_{n=1}^{N-\ell} \hat{\varepsilon}_{n+\ell}(t)\hat{\varepsilon}_n(t),$$

and set

$$\widehat{W}_{i,j}(t,s) = \begin{cases} \widehat{R}_{i-j}(s,t)K(|i-j|h), & i \geq j, \\ \widehat{R}_{j-i}(t,s)K(|i-j|h), & i < j. \end{cases}$$

We use the Epanechnikov kernel, $K(x) = (1-x^2)_+$, and $h = 1.147(N/4)^{1/3}$. These are usual choices that work well in practice (see (Horváth and Kokoszka, 2012)); in our setting they produce nice estimates $\widehat{\mathbf{W}}(t)$, which are both banded and tapered. Expression (A.1) can be seen as a cross–covariance function of two time series, one indexed by t, the other by s. Unlike autocovariance function, the sample cross-covariance function is defined for both positive and negative lags. However, computationally, one does not want to store more lags than is needed. Making use of the relationship $R_\ell(t,s) = R_{-\ell}(s,t)$, we need only keep track of nonnegative lags.

A direct construction of $\hat{\beta}_{GLS}$ is not possible due to the presence of the matrix $\widehat{\mathbf{W}}$. We therefore use a recursive approach where $\hat{\beta}_{LS}$ is used to get an initial estimate of $\mathbf{W}$, which is then used to get an estimate of $\hat{\beta}_{GLS}$, which is then use to reestimate $\mathbf{W}$. This cycle is continued until a convergence criterion described below is satisfied. The primary task still reduces to the estimation of the covariances in the matrix $\mathbf{W}(t,s)$ defined by (A.1). This can be done in a manner similar to the construction of $\widehat{\mathbf{W}}(t)$. At the rth iteration, we compute

$$\widehat{R}_\ell^{(r)}(t,s) = \frac{1}{N-p-\ell} \sum_{n=1}^{N-\ell} \hat{\varepsilon}_{n+\ell}^{(r)}(t)\hat{\varepsilon}_n^{(r)}(s), \quad 0 \leq \ell < N;$$

and then

$$\widehat{W}_{i,j}^{(r)}(t,s) = \begin{cases} \widehat{R}_{i-j}^{(r)}(s,t)K(|i-j|h), & i \geq j, \\ \widehat{R}_{j-i}^{(r)}(t,s)K(|i-j|h), & i < j. \end{cases}$$

The recursive procedures stop if $\|\hat{\beta}_k^{r-1} - \hat{\beta}_k^r\|^2 < 10^{-4}$, i.e. the convergence criterion is based on the target quantity, the coefficient function of interest.

To implement the tests based on the four statistics introduced in Section 2, we must estimate the covariance functions (2.5) for the estimators $\hat{\beta}_{LS}$ and $\hat{\beta}_{GLS}$. Direct verification shows that

$$\widehat{\mathbf{C}}_{LS}(t,s) = \widehat{\mathrm{Cov}}(\hat{\beta}_{LS}(t), \hat{\beta}_{LS}(s)) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{W}}(t,s)\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1},$$

where the $N \times N$ matrix $\mathbf{W}(t,s)$ is defined by (A.1). While

$$\begin{aligned} \widehat{\mathbf{C}}_{GLS}(t,s) &= \widehat{\mathrm{Cov}}(\hat{\beta}_{GLS}(t), \hat{\beta}_{GLS}(s)) \\ &= (\mathbf{X}^\top\widehat{\mathbf{W}}^{-1}(t)\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{W}}^{-1}(t)\widehat{\mathbf{W}}(t,s)\widehat{\mathbf{W}}^{-1}(s)\mathbf{X}(\mathbf{X}^\top\widehat{\mathbf{W}}^{-1}(s)\mathbf{X})^{-1}. \end{aligned}$$

Here it is understood that the GLS estimator uses the recursive estimate of $\mathbf{W}$.

B Proof of Theorem 3.1

An important intermediate step leading to the proof of Theorem 3.1 is an analogous result for an m-dependent sequence, which is stated in Theorem B.1. We emphasize that stationarity is not assumed in this result. It is therefore of independent interest, as it may be used as a building block for results which involve some other forms of nonstationary weak dependence which include m-dependence as a special case. Following the work of Hoeffding and Robbins (1948), there has been a fair deal of research on the CLT for *scalar* m-dependent sequences, Romano and Wolf (2000) provide a general result which includes many previous results as special cases. Theorem B.1 is designed for Hilbert space valued nonstationary sequences. Before providing the result, we state the following Lemma established by Reimherr and Nicolae (2016).

Lemma B.1 (Reimherr and Nicolae) *For each $1 \leq N < \infty$, let $X_{1,N}, \ldots X_{N,N}$ be independent elements of a separable Hilbert space with mean zero and $E\|X_{n,N}\|^{2+\delta} < \infty$ for some $\delta > 0$ and all $1 \leq n \leq N$. If*

$$\sum_{n=1}^{N} E[X_{n,N} \otimes X_{n,N}] \overset{\mathcal{N}}{\to} C,$$

and

$$\sum_{n=1}^{N} E\|X_{n,N}\|^{2+\delta} \to 0,$$

then

$$\sum_{n=1}^{N} X_{n,N} \xrightarrow{\mathcal{D}} \mathcal{N}(0, C).$$

THEOREM B.1 *Assume that $\{Z_n\}$ is an m-dependent sequence in a real separable Hilbert space $\mathbb{H}$ and that conditions 1,2 and 3 of Theorem 3.1 hold. Then relation* (3.2) *holds.*

PROOF: To set notation, we assume that Z_n is independent of Z_k if $|n-k| \geq M$. Choose $M_0 > 2M$, M_0 will used to block observations, and set $r = \lfloor N/M_0 \rfloor$. Define

$$Y_{NM_0} = N^{-1/2} \sum_{i=1}^{r} \sum_{j=1}^{M_0-M} Z_{(i-1)M+j} = N^{-1/2} \sum_{i=1}^{r} U_i,$$

with the $U_i = \sum_{j=1}^{M_0-M} Z_{(i-1)M+j}$ being independent. Using Lemma B.1, Y_{NM_0} will be asymptotically Gaussian if

$$N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{M_0-M} \sum_{s=1}^{M_0-M} \mathrm{E}[Z_{(i-1)m+j} \otimes Z_{(i-1)m+s}] \xrightarrow{\mathcal{N}} C_{M_0},$$

and for some $\delta > 0$

$$N^{-2-\delta} \sum_{i=1}^{r} \mathrm{E}\|U_i\|^{2+\delta} \to 0.$$

Since the first property is assumed, we need only prove the latter. Set $\delta = 2$, then

$$N^{-4} \sum_{i=1}^{r} \mathrm{E}\|U_i\|^4 = N^{-4} \sum_{i=1}^{r} \mathrm{E}\langle U_i, U_i\rangle^2.$$

Consider $i = 1$, then

$$\mathrm{E}\langle U_1, U_1\rangle^2 = \sum_{j_1=1}^{M_0-M} \sum_{j_2=1}^{M_0-M} \sum_{s_1=1}^{M_0-M} \sum_{s_2=1}^{M_0-M} \mathrm{E}[\langle Z_{j_1}, Z_{s_1}\rangle \langle Z_{j_2}, Z_{s_2}\rangle].$$

Notice that for the summand to be nonzero, each index must be within M of at least one other. One can show that the number of such terms is bounded by $3(2M+1)^2(M_0-M)^2$; sharper rates can be found, but all will have this asymptotic order with respect to M and M_0. Multiple applications of the Cauchy–Schwarz inequality imply that

$$|\mathrm{E}[\langle Z_{j_1}, Z_{s_1}\rangle \langle Z_{j_2}, Z_{s_2}\rangle]| \leq \sup_n \mathrm{E}\|Z_n\|^4.$$

Thus we have the bound

$$N^{-4} \sum_{i=1}^{r} \mathrm{E}\|U_i\|^4 \leq \frac{3r(2M+1)^2(M_0-M)^2}{N^4} \sup_n \mathrm{E}\|Z_n\|^4 \to 0,$$

as $N \to \infty$. We therefore have that $Y_{NM_0} \Rightarrow Y_{M_0}$ as $N \to \infty$, where Y_{M_0} is Gaussian. Clearly $Y_{M_0} \Rightarrow Y$ as $M_0 \to \infty$ as desired. So we need only show that the block approximation and the original sum are, asymptotically, equivalent. To do this, it is enough to show that

$$\lim_{M_0 \to \infty} \limsup_{N \to \infty} \mathrm{E}\|N^{1/2}\bar{Z} - Y_{NM_0}\|^2 = 0,$$

cf. Theorem 3.2 of Billingsley (1999). So consider

$$\begin{aligned}
\mathrm{E}\|N^{1/2}\bar{Z} - Y_{NM_0}\|^2 &= N^{-1}\mathrm{E}\left\|\sum_{i=1}^{r-1}\sum_{j=0}^{M} Z_{iM_0-j}\right\|^2 \\
&= N^{-1}\sum_{i=1}^{r-1}\sum_{j=0}^{M}\sum_{s=0}^{M}\mathrm{E}\langle Z_{iM_0-j}, Z_{iM_0-s}\rangle \\
&\leq \frac{r(M+1)^2 \sup \mathrm{E}\|Z_n\|^2}{N} \sim \frac{(M+1)^2 \sup \mathrm{E}\|Z_n\|^2}{M_0} \to 0,
\end{aligned}$$

as $M_0 \to \infty$, thus the claim holds.

Proof of Theorem 3.1: By assumption, there exists a sequence $Z_{n,M}$,

$$Z_{n,M} = f_n(u_n, u_{n-1}, \ldots, u_{n-M}, u^\star_{n-M-1}, u^\star_{n-M-2}, \ldots),$$

which is M-dependent, decoupled from the original sequence, and such that

$$\sum_{M=1}^{\infty} \sup_{1 \leq n < \infty} \nu_2(Z_n - Z_{n,M}) < \infty. \tag{B.1}$$

Define

$$Y_{N,M} = N^{-1/2}\sum_{n=1}^{N} Z_{n,M}$$

By Theorem B.1 $Y_{N,M} \Rightarrow Y_M$, where Y_M is a mean zero Gaussian element of $\mathbb{H}$ with the covariance operator

$$C_M = \lim_{N \to \infty} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{E}(Z_{i,M} \otimes Z_{j,M}).$$

Clearly, $Y_M \Rightarrow Y$ where Y is a mean zero Gaussian element of $\mathbb{H}$ with covariance operator C. Thus the original sum will have the same limiting distribution if

$$\lim_{M \to \infty} \limsup_{N \to \infty} \mathrm{E}\|N^{1/2}\bar{Z} - Y_{N,M}\|^2 = 0.$$

We have

$$\begin{aligned}
\mathrm{E}\|N^{1/2}\bar{Z} - Y_{N,M}\|^2 &= N^{-1}\mathrm{E}\left\|\sum_{n=1}^{N} Z_n - Z_{n,M}\right\|^2 \\
&= N^{-1}\sum_{n_1}^{N}\sum_{n_2}^{N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2} - Z_{n_2 M}\rangle \\
&= N^{-1}\sum_{n}^{N}\mathrm{E}\|Z_n - Z_{nM}\|^2 + 2N^{-1}\sum_{1\le n_1<n_2\le N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2} - Z_{n_2 M}\rangle.
\end{aligned}$$

Since the sum in (B.1) is finite, it follows that

$$\limsup_{M}\limsup_{N} N^{-1}\sum_{n}^{N}\mathrm{E}\|Z_n - Z_{n,M}\|^2 = 0.$$

Next decompose the second summand into

$$2N^{-1}\sum_{1\le n_1<n_2\le N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2}\rangle - 2N^{-1}\sum_{1\le n_1<n_2\le N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2 M}\rangle. \tag{B.2}$$

We now make use of a slight modification of the arguments given in Horvath, Kokoszka, and Reeder (2013). x If $n_2 > n_1$ then $(Z_{n_1}, Z_{n_1 M})$ is independent of Z_{n_2,n_2-n_1}. We can thus make the insertion

$$\begin{aligned}
&\sum_{1\le n_1<n_2\le N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2}\rangle \\
&= \sum_{1\le n_1<n_2\le N}\mathrm{E}\langle Z_{n_1} - Z_{n_1 M}, Z_{n_2} - Z_{n_2,n_2-n_1}\rangle \\
&\le \sum_{1\le n_1<n_2\le N}\sqrt{\mathrm{E}\|Z_{n_1} - Z_{n_1 M}\|^2}\sqrt{\mathrm{E}\|Z_{n_2} - Z_{n_2,n_2-n_1}\|^2} \\
&\le \sup_n\sqrt{\mathrm{E}\|Z_n - Z_{nM}\|^2}\sum_{1\le n_1<n_2\le N}\sqrt{\mathrm{E}\|Z_{n_2} - Z_{n_2,n_2-n_1}\|^2} \\
&\le \sup_n\sqrt{\mathrm{E}\|Z_n - Z_{nM}\|^2}\sum_{1\le n\le N}\sum_{m=1}^{\infty}\sqrt{\mathrm{E}\|Z_n - Z_{n,m}\|^2} \\
&\le N\sup_n\sqrt{\mathrm{E}\|Z_n - Z_{nM}\|^2}\sum_{m=1}^{\infty}\sup_n\sqrt{\mathrm{E}\|Z_n - Z_{n,m}\|^2},
\end{aligned}$$

and the above is such that

$$\left[\lim_M\limsup_N\sup_{1\le n\le N}\sqrt{\mathrm{E}\|Z_n - Z_{nM}\|^2}\right]\sum_{m=1}^{\infty}\sup_{1\le n\le N}\sqrt{\mathrm{E}\|Z_n - Z_{n,m}\|^2} = 0,$$

by assumption. Similar arguments hold for the second term in B.2 and thus the claim holds.

C Cross–section of returns and portfolio formation

To facilitate the comparison with the new approach we propose, in this section we explain the basic ideas of the approach to the study of the cross–section of returns based on portfolio formation. The most frequently cited early reference to this approach is the work of Fama and MacBeth (1973) whose objective was to test the validity of the efficiency equation:

$$ER(i) = ER_0 + \beta(i)[ER(m) - ER_0], \tag{C.1}$$

derived from the static two–parameter CAPM; R_0 is the risk–free return and the $\beta(i) = \mathrm{Cov}(R(i), R(m))/\mathrm{SD}(R(m))$ is the usual beta of asset i. To test the validity of (C.1), Fama and MacBeth (1973) considered the model

$$R_n(i) = \alpha_n + \beta(i)f_n(1) + \beta^2(i)f_n(2) + s(i)f_n(3) + \varepsilon_n(i), \tag{C.2}$$

where $s(i)$ is an asset specific quantity constructed by the authors, which plays a role of a beta in (1.1). The objective of testing is to determine if any other betas, except the beta in (C.1), should be used to explain the cross–section of returns. The null hypothesis is formulated as $Ef_{2n} = 0, Ef_{3n} = 0$. In such an approach, the quantities $\beta(i), \beta^2(i)$ and $s(i)$ are treated as known. In fact, they are not known, and a number of refinements are needed to carry out the test. We now describe the main aspects of the testing procedure.

The tests use 20 sub–portfolios of a large market portfolio to reduce the variability in the estimates of the β_i. These portfolios change from month to month and are constructed by ordering the individual estimates of $\beta_n(i)$. This procedure leads to the following regression for each month (indexed by n):

$$R_n(p) = \alpha_n + f_n(1)\hat{\beta}_{n-1}(p) + f_n(2)\hat{\beta}^2_{n-1}(p) + f_n(3)\hat{s}_{n-1}(p) + \varepsilon_n(p), \quad p = 1, 2, \ldots, 20. \tag{C.3}$$

In regression (C.3), $\hat{\beta}_{n-1}(p), \hat{\beta}^2_{n-1}(p), \hat{s}_{n-1}(p)$ are treated as the explanatory variables. For each month n the least squares estimates $\hat{\alpha}_n,\ \hat{f}_n(1),\ \hat{f}_n(2),\ \hat{f}_n(3)$ are computed. To test the null hypothesis $Ef_n(3) = 0$, the usual t–test is applied to the $\hat{f}_n(3)$. A possible temporal dependence is ignored; the $\hat{f}_n(3)$ are treated as independent observations with common mean and variance. The above approach has been extended and modified in various directions, see e.g. Fama and French (1992), which is perhaps the most cited work in this field.

Our research has been most directly motivated by the work of Ang *et al.* (2006) who investigate whether stocks with different sensitivities to innovations in aggregate volatility have different returns. The main steps of the analysis can be summarized as follows. At the end of each month n, for each stock i, run the regression

$$r_t(i) = \beta_0(i) + \beta_m(i)M_t + \beta_v(i)V_t + \varepsilon_t(i),$$

where t indexes a trading day in month n, M_t is the excess market return, and V_t is the difference in the VIX index from day $t-1$ to day t. Sort the stocks based on the least squares estimates $\hat{\beta}_v(i)$ and form five quintiles: quintile 1 contains stocks with the lowest 20% of the $\hat{\beta}_v(i)$, quintile 5 those with the highest. Denote by $Q_{p,n}, p = 1, 2, 3, 4, 5$, the set of stocks which fall into quintile portfolio p thus formed. Notice that the composition of these portfolios changes from month to month. Each month, one can compute the value–weighted average of the $\hat{\beta}_v(i)$ for each quintile. For example, for the first quantile, we compute $\bar{\beta}_{v,n}^{(p)} = \sum_{i \in Q_{p,n}} w_{i,n} \hat{\beta}_{v,n}^{i}$. There are $N = 166$ months in the study, and one can compute

$$\bar{\beta}_v^{(p)} = \frac{1}{N-1} \sum_{n=1}^{N-1} \bar{\beta}_{v,n}^{(p)}, \quad p = 1, 2, 3, 4, 5.$$

Table 1 in Ang *et al.* (2006) reports that $\bar{\beta}_v^{(1)} = -2.09, \bar{\beta}_v^{(3)} = 0.03$ and $\bar{\beta}_v^{(5)} = 2.18$. In month $n+1$, we compute the value–weighted return for the stocks in each quintile formed in month n:

$$R_{n+1}(p) = \sum_{i \in Q_{p,n}} w_{i,n} R_{n+1}(i), \quad p = 1, 2, 3, 4, 5.$$

We then compute the averages $\bar{R}(p)$ of the $R_{n+1}(p)$. Table 1 in Ang *et al.* (2006) reports that $\bar{R}(1) = 1.64, \bar{R}(3) = 1.36, \bar{R}(5) = 0.60$ percent. This analysis indicates that stocks with low (large negative) sensitivities to the daily change in VIX have higher expected monthly returns next month. After a more detailed validation of this conclusion, the authors turn to the contemporaneous relationship at the monthly frequency by considering the model

$$\text{(C.4)} \qquad R_n(p) = \alpha(p) + \beta_m(p) M_n + \beta_s(p) S_n + \beta_h(p) H_n + \beta_v(p) V_n + \varepsilon_n(p),$$

where S_n and H_n are the usual Fama–French factors, and V_n is a custom constructed factor which reflects monthly change in aggregate volatility (taking a monthly difference in the VIX is not suitable). The resulting least squares estimates $\hat{\beta}_v(p)$ are highly significant. For example, the t–score of $\hat{\beta}_v(1)$ is -4.06 and that of $\hat{\beta}_v(5)$ is 5.35. The results from regression (2.3) indicate that after controlling the Fama-French factors, there is a negative relationship between the change of aggregate volatility and the the return of portfolio constructed from stocks with low (large negative) sensitivities to the daily change in VIX, whereas the relationship is positive for portfolio constructed from stocks with high (large positive) sensitivities to the daily changes in the VIX.

Acknowledgements We thank an anonymous referee for studying our paper carefully and providing useful and constructive advice. The Editor, Professor Trojani, has directed us to relevant work from which we benefited. This research was partially supported by the NSF grant DMS–1462067.

References

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006). The cross–section of volatility and expected returns. *Journal of Finance*, **61,** 259–299.

Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, **37,** 4046–4087.

Berkes, I., Hörmann, S. and Schauer, J. (2011). Split invariance principles for stationary processes. *The Annals of Probability*, **39,** 2441–2473.

Billingsley, P. (1999). *Convergence of Probability Measures; Second Edition.* Wiley, New York.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, New Jersey.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, **52,** 57–82.

Chen, X. and White, H. (1998). Central limit and functional central limit theorems for Hilbert–valued dependent heterogeneous arrays with applications. *Econometric Theory*, **14,** 260–284.

Chiou, J-M., Müller, H-G. and Wang, J-L. (2004). Functional response models. *Statistica Sinica*, **14,** 675–693.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, **44,** 32–61.

Davidson, J. (1993). The central limit theorem for globally nonstationary near–epoch dependent functions of mixing processes: the asymptotically degenerate case. *Econometric Theory*, **9,** 402–412.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence and applications to moment inequalities. *Stochastic Processes and their Applications*, **84,** 313–343.

Durand, R. B., Lim, D. and Zumwalt, J. K. (2011). Fear and the Fama–French factors. *Financial Management*, **40,** 409–426.

Fama, E. F. and French, K. R. (1992). The cross–section of expected stock returns. *Journal of Finance*, **47,** 427–465.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, **81,** 607–636.

Fan, J. and Zhang, J-T. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, **27,** 1491–1518.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, **39,** 254–261.

Gagliardini, P., Ossola, E. and Scaillet, O. (2016). Time–varying risk premium in large corss-sectional equity data sets. *Econometrica*, **84,** 985–1056.

Harvey, C. R., Liu, Y. and Zhu, H. (2014). ...and the cross-section of expected returns. Technical Report. Duke University. Available at Social Science Research Network.

Hays, S., Shen, H. and Huang, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, **6,** 870–894.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, **12,** 141–153–77.

Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Mathematical Journal*, **15,** 773–780.

Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory*, **29,** 267–288.

Hörmann, S., Kidzinski, L. and Kokoszka, P. (2015). Estimation in functional lagged regression. *Journal of Time Series Analysis*, **00,** 000–000; Forthcoming.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, **38,** 1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, **75,** 103–122.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, **179,** 66–82.

Horváth, L. and Rice, G. (2015). Testing for independence between functional time series. *Journal of Econometrics*, **189,** 371–382.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data analysis, with an Introduction to Linear Operators*. Wiley.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, **99,** 2508–2526.

Kokoszka, P., Miao, H. and Zhang, X. (2015). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics*, **13,** 456–477.

Kokoszka, P. and Reimherr, M. (2013). Predictability of shapes of intraday price curves. *The Econometrics Journal*, **16,** 285–308.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley.

Müller, H-G., Sen, R. and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, **165,** 233–245.

Neumann, M. H. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, **17,** 120–134.

Pötscher, B. and Prucha, I. (1997). *Dynamic Non–linear Econonometric Models. Asymptotic Theory*. Springer.

Ramsay, J., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer.

Ramsay, J. O. and Ramsey, J. B. (2001). Functional data analysis of the dynamics of the monthly index of non durable goods production. *Journal of Econometrics*, **107,** 327–344.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Reimherr, M. and Nicolae, D. (2014). A functional data analysis approach for genetic association studies. *The Annals of Applied Statistics*, **8,** 406–429.

Reimherr, M. and Nicolae, D. (2016). Estimating variance components in functional linear models with applications to genetic heritability. *Journal of the American Statistical Association*, **111,** 407–422.

Renault, E., van der Heijden, T. and Werker, J. M. (2016). Arbitrage pricing theory for idiosyncratic variance factors. Working paper. Brown University.

Romano, J. P. and Wolf, M. (2000). A more general central limit theorem for m–dependent random variables with unbounded m. *Statistics and Probability Letters*, **47,** 115–124.

Romano, J. P. and Wolf, M. (2013). Testing for monotonicity in expected asset returns. *Journal of Empirical Finance*, **23,** 93–116.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, **35,** 1773–1801.

Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications with R Examples*. Springer.

Wu, W. (2005). *Nonlinear System Theory: Another Look at Dependence*. Proceedings of The National Academy of Sciences of the United States, volume 102.