Specification Analysis of Functional Autoregressive Models

Maciej Opuchlik
Universidad Carlos III de Madrid
Economics Department

November 30, 2013

Abstract

We propose new tests for the correct specification of functional models in terms of transformed residual sample autocorrelations. The methodology of transformation of empirical residual autocorrelations has been recently used for parametric multivariate time series models and in the special case of VARMA with closed form solutions. This paper is looking to expand it to the functional setting for autoregressive models. We propose the method of eliminating the asymptotic estimation error in the sample autocorrelations in the framework of functional (Hilbertian) autoregressive model. The most promising approach to this setup is based on Karhunen-Loéve expansions, which are approximations of objects in $L^2(0,1)$ in the finitely dimensional R^p space. We show that a functional model where the stochastic process is driven by a Hilbert-Schmidt integral operator may be approximated with multivatiate VAR model without asymptotic estimation error. Our method does not require specific estimators of eigenfunctions and parameters of the estimated operator. We present Monte Carlo experiments performed for the classical Box-Pierce test of lack of correlation and our transformation of Box-Pierce statistic. The application is presented concerning the high frequency financial data.

Keywords: Functional data, Goodness-of-fit test, model checking, portmanteau Ljung-Box and Box-Pierce, recursive residuals, residual autocorrelation.

1 Introduction

Functional data analysis has emerged as a significant tool for modeling large dimension data in the last decade. From the practical standpoint it started as a method of utilizing large data sets that have become available characterized by a large record frequency and a limited number of periods when the data has been obtained. On one hand it was natural to assume that such datasets represent the functions discretized on a minutely/daily/weekly time frame while on the other hand the theoretical analysis has been at first omitted. It may be assumed that the functions under study are of some level of regularity, however only recently the subject has been studied from the strict statistical perspective.

From a functional point of view the set of data is represented by the sequence of functions $\{X_n(t)\}$ that are elements of some Hilbert space. Functional time series arises when a long record $\{X(t), t \in [0,T]\}$ in which t is, a continuous index, can be naturally split into segments of equal length. Then it is possible to set

$$X_n(t) = X\left(\{n-1\}T + t\right),\ t \in [0,T],\ n = 1, 2, \ldots, N,$$

where $X(t)$ is a raw time series data record. The transformed series $X_n(t)$ consists of curves treated as observables for $n = 1, 2, \ldots, N$.

The subject has become very broad recently. As comprehensible introductory expositions we may note Ramsay & Silverman (2002), Ramsay & Silverman (2005), and Ramsay et al. (2009), and more theoretical works by Bosq (2000), Ferraty & Vieu (2006), Bosq & Blanke (2007) and Ferraty & Romain (2011). The functional autoregressive model that we are interested in has been theoretically studied by Bosq (2000) and extensively used in practical and theoretical framework (see *e.g.* Besse & Cardot (1996), Antoniadis & Sapatinas (2003), Horváth et al. (2010), Hörmann & Kokoszka (2010). Applications of functional data in economics are focused on financial data analysis which comes from the typical high frequency data sets available. One could note analysis of periodicity and volatility persistence in financial markets (see Fengler et al., 2003).

The simplest time series model in a functional setting is the FAR(1) model of

Bosq, which extends to the functional setting the usual multivariate VAR model. The model is given by the equation

$$X_{n+1} = \Psi(X_n) + \varepsilon_{n+1}, \tag{1}$$

where errors ε_n and observations X_n are curves and Ψ is a linear operator transforming one curve to another curve. Despite its conceptual simplicity, it is a very flexible modeling and predictive tool because the autoregressive operator acts on a Hilbert space whose elements can exhibit any degree of nonlinearity. Thus, even though FAR(1) is a linear model in a function space, it is in fact nonlinear in the sense of finite dimensional spaces.

The problem we are going to adress is testing for correlation of residuals in a functional model formulated as (1). The main idea is based on the observation that it is possible to represent the functional process with a finite dimensional VAR using Karhunen-Loéve expansions. In a similar manner we may ask if it is possible to use developed methods of model checking in a functional setup. In a general parametric time series framework functions of residuals are a key tool for model checking. In case of testing lack of correlation, the standard method is based on using statistics of empirical residuals autocorrelations. Usually portmanteau tests like Ljung-Box (1976) and Box-Pierce (1970) are being utilized. The same applies for vector autoregressive (VAR) and vector autoregressive moving average models (VARMA) that are a standard econometric tool used for macroeconomic data. This class of models is a natural expansion of univariate ARMA models and thus they were extensively studied during the 90's (see *e.g.* Lütkepohl, 1993). In functional time series the modification of Box-Pierce test has been proposed in Gabrys, Hörmann & Kokoszka (2010). However, this test behaves poorly for functional autoregressive model because of asymptotic estimation error of $\hat{\Psi}$ and may be applied only in case of standard functional regression given by

$$Y_n = \Psi(X_n) + \varepsilon_n, \tag{2}$$

for curves Y_n, X_n and functional errors ε_n with X_n independent of ε_n.
The main problem in proposing an feasible test for correlation of errors is due to the assymptotic effect of estimation error. The most recent solution has been proposed in Delgado & Velasco (2011) for general class of paremetric models. The method is based on a transformation of the residual autocorrelation vector to account for possible serial dependence in order to obtain an asymptotically multivariate standard normal distribution and then a pivotal asymptotic transform of empirical autocorrelations that would orthogonalise the system of m residual serial autocorrelations eliminating parameter estimation effect. However, in the functional setup the linear transformations of processes in functional spaces may show behaviour that does not have its analog in a multivariate setting. In order to propose the test of lack of correlation of residuals we need to construct a computable test statistic of finite dimensional objects that would describe the behaviour of residual functions $\varepsilon_n(t)$. In multivariate setting the basis is implicitly given while in the functional framework we must choose and estimate the most suitable one. Then in order to obtain finite dimensional objects we have to truncate the observables $X_n(t)$ and thus residuals $\varepsilon_n(t)$ and these projections are going to produce additional approximation error. This is one of the differences with multivariate VARMA case because even in ideal setting where the errors are iid the error approximations in Karhunen-Loéve basis vould be weakly dependent.
The article is organized as follows: in Section 2 we are going to introduce the basic method of estimating the Karhunen-Loéve expansion and assumptions of autoregressive functional model and goodness-of-fit portmanteau test being the benchmark given by Gabrys et al. (2011). Section 3 will present the algorithm of pivotal asymptotic orthogonalisation of vectorized autocorrelations vector that would asymptotically eliminate the estimation error. In Section 4 we consider the asymptotic effect of estimation of eigenfunction basis of Karhunen-Loéve expansion. Section 5 is devoted to Monte Carlo simulation experiment that would compare this method to the available alternatives. It also contains the application for financial data. All the technical proofs and mathematical notes are being relegated to Appendix.

2 Diagnostic checking in case of FAR(1) model

The model we analize is the FAR(1) of the following form

$$\begin{aligned} &X_{n+1}(t) = \int_0^1 \varphi(s,t)X_n(s)ds + \varepsilon_n(t),\ t \in [0,1], \\ &\{\varepsilon_n(t)\} \text{ i.i.d. for } n \in \mathbb{Z}, \end{aligned} \tag{3}$$

where $X_n(t), \varepsilon_n(t)$ are functions and imposing assumptions $\varphi(s,t)$ that would grant stationarity of $X_n(t)$. Note that according to (3) the series $X_n(t)$ will be driven by Hilbert-Schmidt integral operator. Theoretical background has been moved to Appendix. The assumptions needed to consistently estimate the FAR(1) model (3) and to formulate the "null" hypothesis on functional errors ε_n are as follows:

A1: $\varepsilon_n(t)$ are iid with and $||\varepsilon_n(t)||_2 < \infty$ for $-\infty \leq n \leq N$. $E\varepsilon_n(t) = 0$.

A2: $\varphi \in \Phi$ is measurable and satisfies

$$||\varphi||_\infty < 1 \text{ for } \varphi \in \Phi, \tag{4}$$

which implies

$$\int_0^1 \int_0^1 |\varphi(s,t)|^2 dsdt < \infty \text{ (Hilbert-Schmidt) } . \tag{5}$$

The assumption **A1** has to be imposed in order to make X_n stationary without estimating the intercept μ. As in the multivariate case independence of curves implies lack of correlation. Condition (4) gives the characterization of X_n. Note that if we define $\mathcal{M}$ as σ-algebra generated by sequence $\{\varepsilon_n(t), \varepsilon_{n-1}(t), \dots\}$ then the sequence $X_n\{\omega\}$ is going to be stationary for $\omega \in \mathcal{M}$. Note that we do not have to impose assumptions on the continuity of φ. Model (3) defines a member of a class of Hilbert-Schmidt operators which are bounded and continuous. Finally Φ is a class of functions that define Ψ given in (1) to be a Hilbert-Schmidt operator.

In order to propose the test for error serial correlation we define the residuals with respect to model (3) as

$$\varepsilon_n[\varphi] = X_{n+1} - \int_0^1 \varphi(X_n), \tag{6}$$

for $\varphi \in \Phi$, while simplifying the notation. Clearly we are treating the residuals as functions of φ. In a general hilbertian setup we say that ε_n and ε_{n+i} are uncorrelated if and only if

$$E\left(\left\langle \varepsilon_n, f \right\rangle \left\langle \varepsilon_{n+i}, g \right\rangle\right) = E \int_0^1 \varepsilon_n(s) f(s) ds \int_0^1 \varepsilon_{n+i}(s) g(s) ds = 0$$

for any $f, g \in L^2(0,1)$ with the standard inner product defined in $L^2(0,1)$ as $\left\langle \varepsilon_n, f \right\rangle = \int_0^1 \varepsilon_n(s) f(s) ds$. In order to test the lack of correlation of curves $\varepsilon_n[\hat{\varphi}]$ we need to reduce the dimension of the estimated kernel φ. Usually the Karhunen-Loéve expansion is utilized but in fact we have that for any $X_n \in L^2(0,1)$ we may perform the orthogonal projection of respectively $X_n(t)$ and implicitly $\varepsilon_n(t)$ on p-dimensional subspace of $L^2(0,1)$ spanned by orthonormal basis $v_1, \ldots, v_p$. The choice of Karhunen-Loéve basis, based on the decomposition of the covariance operator of a series X_n is motivated by the optimal convergence to the true series.

The idea is based on transforming the model (3) to its approximation in R^p where the vectors would represent the coefficients of each function X_n in the given basis. So the first step is approximating the $\varphi(\cdot,\cdot) \in \Phi$ in the finite $v_1, \ldots, v_p$ basis. Then the estimated kernel $\hat{\varphi}_p(\cdot,\cdot) \in \Phi$ may be represented by the matrix $\hat{\varphi}^{(p)}$ transforming coefficients of past observed curves $X_{n-1}(t)$ into coefficients of $X_n(t)$ in the KL basis $B_0 = [v_1, \ldots, v_p]$. Let us assume that $\{v_k\}$ forms the orthonormal KL basis of $L^2(0,1)$. Thus each observation X_n admits the representation

$$X_n = \sum_{i=1}^{\infty} \left\langle v_i, X_n \right\rangle v_i.$$

Now if $\{v_k\}$ is a basis of $L^2(0,1)$ then $\{v_i(t)v_j(s)\}$ is a basis of $L^2([0,1]\times[0,1])$ because φ is Hilbert-Schmidt kernel. It follows that $\varphi(t,s)$ may be decomposed as

$$\varphi(t,s)=\sum_{i,j=1}^{\infty}\varphi_{j,i}v_i(t)v_j(s), \tag{7}$$

for any $\varphi(t,s)\in\Phi$, where $\varphi_{j,i}$ with $0\leq j,i\leq p$ are coefficients of $\varphi_0^{(p)}$ matrix satisfying

$$\varphi_{j,i}=\Big\langle\varphi(t,s),v_i(t)v_j(s)\Big\rangle=\int_0^1\varphi(t,s)v_i(t)v_j(s)dtds. \tag{8}$$

The immediate advantage of Karhunen-Loéve expansion is that we may derive $\varphi_{j,i}=\lambda_i^{-1}E\left\langle X_{n-1},v_i\right\rangle\left\langle X_n,v_j\right\rangle$ where λ_i is an eigenvalue associated to v_i. This equation is the immediate analog of OLS estimation of $\varphi^{(p)}$ that has been presented in the Appendix. Note that having the estimator of $\varphi_0^{(p)}=[\varphi_{j,i}]_{i,j<p}$ we can write the kernel estimator by (7) as

$$\hat{\varphi}_p(t,s)=\sum_{i,j=1}^{p}\hat{\varphi}_{j,i}v_i(t)v_j(s). \tag{9}$$

being an exact multivariate OLS analog (see Appendix). Clearly $\hat{\varphi}_p(\cdot,\cdot)\in\Phi$ is not feasible kernel estimator because true basis $v_1,\ldots,v_p$ is not known and has to be estimated. The feasible estimator would have the form

$$\tilde{\varphi}_p(t,s)=\sum_{i,j=1}^{p}\hat{\varphi}_{j,i}\hat{v}_i(t)\hat{v}_j(s), \tag{10}$$

with distribution of $\tilde{\varphi}^{(p)}\in\Phi$ possibly depending on the error of basis B_0 estimation. Note that $\Phi^{(p)}$ is the space of $p\times p$ matrices and is isometric with previously defined Φ as far as the $v_1,\ldots,v_p$ base functions are taken into consideration. Thus there is one to one relationship between any $\varphi\in\Phi$ and its $p\times p$ dimensional counterpart $\varphi^{(p)}\in\Phi^{(p)}$.

In order to define the residuals for any kernel $\varphi_p \in \Phi$ we may rewrite model (3) using (7) as

$$X_{n+1}(t) = \sum_{i=1}^{p} \sum_{j=1}^{p} \varphi_{j,i} \left\langle X_n, v_j \right\rangle v_i(t) + \gamma_{n+1}^{(p)}[\varphi](t) + \varepsilon_{n+1}[\varphi](t), \tag{11}$$

for any $\varphi \in \Phi$, where

$$\gamma_{n+1}^{(p)}[\varphi](t) = \sum_{i=p+1}^{\infty} \sum_{j=p+1}^{\infty} \varphi_{j,i} \left\langle X_n, v_j \right\rangle v_i(t),$$

so $\gamma_{n+1}^{(p)}[\varphi]$ is an approximation error. Now we may define the residuals $\hat{\varepsilon}_n[\varphi]$ according to equation (11)

$$\varepsilon_{n+1}[\varphi] = X_{n+1} - \begin{bmatrix} v_1 & \dots & v_2 \end{bmatrix} \begin{bmatrix} \varphi_{11} & \dots & \varphi_{1p} \\ \vdots & \ddots & \\ \varphi_{p1} & \dots & \varphi_{pp} \end{bmatrix}' \begin{bmatrix} \left\langle X_n, v_1 \right\rangle \\ \vdots \\ \left\langle X_n, v_p \right\rangle \end{bmatrix} + \gamma_{n+1}^{(p)}[\varphi] \tag{12}$$

for $\varphi \in \Phi$, with its projection given as

$$\varphi^{(p)} = \begin{bmatrix} \varphi_{11} & \dots & \varphi_{1p} \\ \vdots & \ddots & \\ \varphi_{p1} & \dots & \varphi_{pp} \end{bmatrix} \in \Phi^{(p)},$$

Now $\gamma_n^{(p)}[\varphi](t)$ represents distortion curve caused by truncation of higher than p order KL expansions factors of $X_n(t)$ (and $\varepsilon_{n+1}[\varphi](t)$ with respect to grid t^* due to quadrature error if this error is going to be discussed). Note that it is impossible to compute $\gamma_n^{(p)}[\varphi]$ for any $\varphi \in \Phi$. Thus $\varepsilon_n[\varphi](t)$ may not be obtained. However, the truncated residuals $\varepsilon_n[\varphi^{(p)}](t) = \varepsilon_n[\varphi] - \gamma_n^{(p)}[\varphi]$ may be computed for $\varphi \in \Phi$ with respect to p most important eigenfunctions.

In order to decrease the dimension of the problem we have to transform residuals defined as in (12) to the representation dependent on the coefficients in the

$v_1, v_2, \ldots, v_p$ basis. We are going to denote $X_n^{(p)}(t)$ as a Karhunen - Loéve expansion of $X_n(t)$ and $x_{ni} = \left\langle X_n, v_i \right\rangle$ as KL coefficients of $X_n(t)$. Note that x_{ni} does not depend on the p-length of the KL expansion for specific $n \leq N$, $i = 1, \ldots, p$ so that each element of this vector is unique. We will use the notation t^* for the grid where various functions may be observed with $t^* \subset (0, 1)$. Now we may rewrite (12) as

$$\begin{aligned} &\begin{bmatrix} v_1(t^*) & \ldots & v_p(t^*) \end{bmatrix} \begin{bmatrix} \epsilon_{(n+1)1} \\ \vdots \\ \epsilon_{(n+1)p} \end{bmatrix} = \\ &\begin{bmatrix} v_1(t^*) & \ldots & v_p(t^*) \end{bmatrix} \left(\begin{bmatrix} x_{(n+1)1} \\ \vdots \\ x_{(n+1)p} \end{bmatrix} - \begin{bmatrix} \varphi_{11} & \ldots & \varphi_{1p} \\ \vdots & \ddots & \\ \varphi_{p1} & \ldots & \varphi_{pp} \end{bmatrix} \begin{bmatrix} x_{n1} \\ \vdots \\ x_{np} \end{bmatrix} \right), \end{aligned} \tag{13}$$

where $\epsilon_{(n+1)i}$ is a KL expansion coefficient of $\varepsilon_n[\varphi](t)$ with respect to v_i and t^* grid. Note that distortion curve $\gamma_n^{(p)}[\varphi]$ from equation (12) is disappears because it does not have a representation in $B = \{v_1, \ldots, v_p\}$ basis.

Now under assumptions **A1**-**A2** we have that for length of KL expansion of $X_n(t)$, p sufficiently large, the $\gamma_n^{(p)}[\varphi]$ error for each $n \leq N$ may be arbitrary small. The distortion coming from estimating the B_0 basis and the eigenvalues Λ_0 will affect the finite sample distribution of autocorrelations of coefficients vectors X_{ni} for $n \leq N, i \leq p$. However, it will not have the assymptotic effect on the distribution of empirical autocovariances. Let us rewrite the model (13) compactly as

$$e_{n+1}[\varphi^{(p)}] = A_{n+1}^{(p)} - \varphi^{(p)} A_n^{(p)}, \tag{14}$$

in general for any $\varphi^{(p)} \in \Phi^{(p)}$ where

$$e_n[\varphi^{(p)}] = \begin{bmatrix} \epsilon_{n1} \\ \vdots \\ \epsilon_{np} \end{bmatrix}_{[\varphi^{(p)}]} = \begin{bmatrix} \left\langle \varepsilon_n, v_1 \right\rangle \\ \vdots \\ \left\langle \varepsilon_n, v_p \right\rangle \end{bmatrix}_{[\varphi^{(p)}]},$$

$$A_n^{(p)} = \begin{bmatrix} x_{n1} \\ \vdots \\ x_{np} \end{bmatrix} = \begin{bmatrix} \left\langle X_n, v_1 \right\rangle \\ \vdots \\ \left\langle X_n, v_p \right\rangle \end{bmatrix},$$

so $e_n[\varphi^{(p)}]$ and A_n for $n \leq N$ are coefficient vectors of respectively residuals $\varepsilon_n[\varphi]$ and observed series X_n. Equation (14) provides the approximation of a FAR(1) defined in (3) and is the equation we are going to refer in further discussion. This approach has been proposed in Gabrys et al. (2010) in a similar functional model with exogenous regressors. Note that equation (14) does not take into account that the basis $B = \{v_1, \ldots, v_p\}$ is unknown and we may use only the estimate $\hat{B}$. According to the formulation of empirical residuals (14) we may define empirical residual covariances

$$\hat{\Gamma}_{\varphi}^{(p)}(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} \left(e_i[\varphi^{(p)}] - \bar{e}[\varphi^{(p)}] \right) \left(e_{i+j}[\varphi^{(p)}] - \bar{e}[\varphi^{(p)}] \right)' \tag{15}$$

for $\varphi^{(p)} \in \Phi^{(p)}$ defined before as a class of linear transforms isometric with space of linear transforms in vector space R^p. We also define $\bar{e}[\varphi^{(p)}]$ as the average across the $1, \ldots, N-j$ residuals in equation (14). We are denoting $\hat{\Gamma}_{\varphi}^{(p)}(j)$ for any $j \geq 1$ as an estimator of $\hat{\Gamma}_{\varphi_0}^{(p)}(j)$.

The "null" hypothesis may be stated as

H_0: The residuals coefficients $e_n[\varphi_0^{(p)}]$ following (14) are not correlated.

against the alternative

$H_1 : H_0$ is not true.

Clearly in a more general sense we are aiming to test the H_0 hypothesis that process $\{X_n(t)\}_{-\infty}^{N}$ is generated by functional model (3) with the kernel $\varphi_0(s,t) \in \Phi$ satisfying **A2** and the residuals $\{\varepsilon_n[\varphi_0]\}$ satisfy assumptions **A1**-**A2**. The test we use is a goodness-of-fit test for univariate strong VAR model introduced by Box and Pierce (1970). The modification of Box-Pierce statistic which has better finite properties is a Ljung-Box portmanteau test proposed by Box and Ljung (1978). Using the BP (LB) test we are are aiming at testing the lack of correlation of coefficients of Karhunen - Loéve expansion of the series up to arbitrary lag H for $X_n(t)$ and $\varepsilon_n[\varphi^p]$ using up to p most important factors. Thus note that the serial correltions of the form $E\left\langle(\varepsilon_n, v_i)\right\rangle\left\langle(\varepsilon_{n-h}, v_j)\right\rangle$ may not be detected for $i, j > p$ or $h > H$.

The test statistic that is going to be used is a modification of the multivariate version of Box-Pierce statistic, (16) was given by Chitturi (1974)

$$\hat{Q}_H(\hat{\varphi}^{(p)}) = N\sum_{h=1}^{H}\mathrm{tr}\{\hat{\Gamma}_{\hat{\varphi}}^{(p)}(h)'\hat{\Gamma}_{\hat{\varphi}}^{(p)}(0)^{-1}\hat{\Gamma}_{\hat{\varphi}}^{(p)}(h)\hat{\Gamma}_{\hat{\varphi}}^{(p)}(0)^{-1}\} \tag{16}$$

or equivalently by Hosking (1981)

$$\hat{Q}_H(\hat{\varphi}^{(p)}) = N\sum_{h=1}^{H}\{\mathrm{vec}(\hat{\Gamma}_{\hat{\varphi}}^{(p)}(h)'[\hat{\Gamma}_{\hat{\varphi}}^{(p)}(0)\otimes\hat{\Gamma}_{\hat{\varphi}}^{(p)}(0)]^{-1}\mathrm{vec}(\hat{\Gamma}_{\hat{\varphi}}^{(p)}(h))\} \tag{17}$$

with $\hat{Q}_H(\varphi_0^{(p)})$ being the statistic $\hat{Q}_H$ evaluated in $\varphi_0^{(p)} \in \Phi^{(p)}$. By testing lack of autocorrelation of functional residuals using Box-Pierce statistics, (16) and (17) we are testing the specification of a FAR(1) model (3) with the arbitrarily chosen level p of Karhunen-Loéve coefficients.

3 Removing the estimation effect of Portmanteau statistics in FAR(1) setting

The aim of this section is presenting a method of eliminating the estimation error from the distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$. The model we are using in this section follows (14) and assumptions **A1-A2**. We assume that the eigenfunctions of KH expansion B_0 are known with the discussion on this problem moved to next section. The number of KL coefficients p is fixed and to simplify notation we will be writing $\varphi \in \Phi^{(p)}$ instead of $\varphi^{(p)} \in \Phi^{(p)}$. It is known that in general multivariate autoregressive VARMA(a, b) under H_0 the asymptotic distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$ statistic does not follow a $\chi^2_{\mathrm{p}^2H}$ asymptotic distribution due to estimation error. It has been shown by Hosking (1980) that finite sample distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$ may be appoximated by a $\chi^2_{p^2(H-a-b)}$ distribution which is closer to the asymptotic distribution of $\hat{Q}_H(\varphi_0^{(p)})$.
Let us define the autocovariance vector by (15)

$$\hat{\gamma}_{\varphi}^{(m)} = \left[\mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1))', \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(2))', \ldots, \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m))'\right]' \tag{18}$$

and autocorrelation vector

$$\hat{\rho}_{\varphi}^{(m)} = \left[\mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1)\hat{\Gamma}_{\varphi}^{(p)}(0)^{-1})', , \ldots, \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m)\hat{\Gamma}_{\varphi}^{(p)}(0)^{-1})'\right]' \tag{19}$$

for $\varphi \in \Phi^{(p)}$. It is clear that under H_0 the elements of $\hat{\gamma}_{\hat{\varphi}}^{(m)}$ are going to be correlated owing to estimation error $(\hat{\varphi}^{(p)} - \varphi_0^{(p)})$. This implies that the true distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$ is going to be different from $\chi^2_{(p^2H)}$ for arbitrary $0 \leq H \leq m$ (see Box, Pierce, 1970 and Durbin, 1970 for VARMA).

Let us define the matrix of empirical derivatives of $\hat{\gamma}_{\hat{\varphi}}^{(m)}$,

$$\hat{\zeta}_{\varphi}^{(m)} = \begin{bmatrix} \nabla \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}(1) \\ \nabla \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}(2) \\ \vdots \\ \nabla \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}(m) \end{bmatrix} \tag{20}$$

where

$$\nabla \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}(i) = \frac{\partial}{\partial \left(\mathrm{vec}\varphi^{(p)}\right)'} \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}, \quad i = 1, \ldots, m.$$

The derivatives with repect to VAR(1) specification of a functional model (3) are going to have a pivotal role in transforming the $\hat{\gamma}_{\varphi}^{(m)}$ vector for $\varphi \in \Phi^{(p)}$. Our solution of the problem of residual dependence in $\hat{\gamma}_{\hat{\varphi}}^{(m)}$ follows Delgado & Velasco (2011). We are not going to derive the true asymptotic distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$ under the H_0 (see Francq and Raïssi, 2007 for VAR, Francq, Roy and Zakoïan, 2005 for ARMA, Li, 1992 for nonlinear models with iid innovations and Hwang, Basawa and Reeves, 1994) but perform the linear transformation of estimated covariance vector $\hat{\gamma}_{\hat{\varphi}}^{(m)}$, defined as (18). The reasoning is based on the fact that we do not directly estimate $\hat{\gamma}_{\varphi_0}^{(m)}$, which is impossible but perform the transform using orthogonal projection operator.

The fundamental result that allows for transformation of $\hat{\gamma}_{\hat{\varphi}}^{(m)}$ is the following

Proposition 1. *Under* H_0, ***A1-A2*** *and*

$$\hat{\varphi}^{(p)} = \varphi_0^{(p)} + O_p(N^{-\frac{1}{2}}), \tag{21}$$

the following holds

$$\hat{\gamma}_{\hat{\varphi}}^{(m)} = \hat{\gamma}_{\varphi_0}^{(m)} + \hat{\zeta}_{\varphi_0}^{(m)}(\varphi_0^{(p)} - \hat{\varphi}^{(p)}) + o_p\left(N^{-\frac{1}{2}}\right) \tag{22}$$

where $\hat{\zeta}_{\varphi_0}^{(m)}$ *is a derivative of* $\hat{\gamma}_{\varphi}^{(m)}$ *evaluated in true parameter* $\varphi_0^{(p)}$ *without the*

associated noise coming from estimation of basis B_0.

Note that assumption (21) requires the $\sqrt{N}$ consistent estimator of $\varphi_0^{(p)}$. In fact OLS estimator is available and satisfies this condition, however the following reasoning is going to be true for any estimator that satisfies (21). We have presented further discussion in the Appendix.

It is clear that *Proposition 1* may be followed by its analogue concerning residual autocorrelations $\hat{\gamma}_{\hat{\varphi}}^{(m)}$ defined in (19) under the same assumptions. It is easy to show that

$$\hat{\rho}_{\hat{\varphi}}^{(m)} = \hat{\rho}_{\varphi_0}^{(m)} + \hat{\eta}_{\varphi_0}^{(m)}(\varphi_0^{(p)} - \hat{\varphi}^{(p)}) + o_p\left(N^{-\frac{1}{2}}\right), \tag{23}$$

with derivative of m autocorrelations $\hat{\eta}_{\varphi_0}^{(m)}$ satisfying

$$\hat{\eta}_{\varphi_0}^{(m)} = \hat{G}_{\varphi}^{-\frac{1}{2}}\hat{\zeta}_{\varphi_0}^{(m)}, \tag{24}$$

$$\hat{G}_{\varphi} = \operatorname{diag}\left(\operatorname{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1))\operatorname{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1))', \ldots, \operatorname{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m))\operatorname{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m))'\right), \tag{25}$$

for $\varphi \in \Phi^{(p)}$.

Now following equation (14) and disregarding the distortion errors $\gamma^{(p)}[\varphi](t)$ the expected relationship

$$\sqrt{N}\hat{\rho}_{\varphi_0}^{(m)} \xrightarrow{d} \mathcal{N}(0, I_{mp^2}).$$

holds and above relationship is the basis of Box-Pierce test. Note that if we consider the asymptotic distribution of $\hat{\rho}_{\hat{\varphi}}^{(m)}$ instead the above relationship will not hold according to *Proposition 1.* We are proposing the pivotal transform of $\hat{\rho}_{\hat{\varphi}}^{(m)}(j)$ for $j \leq m$, based on recursive projections on the space of orthogonalised sample autocorrelations $\{\hat{\rho}_{\hat{\varphi}}^{(m)}(i)\}_{i=j+1}^{j+r}$ of residual coefficients. This transformation due to *Proposition* 1 is asymptotically distribution free (see Delgado & Velasco, 2011). It is widely known that it is not possible to estimate the set of true autocorrelation vectors $\{\hat{\rho}_{\varphi_0}^{(m)}(i)\}_{i=1}^{m}$ because it depends on distribution of estimation error. The idea of martingale transform was introduced by Brown, Durbin & Evans (1975) for cusum tests for linear models. The theoretical part in case of Gaussian processes

follows Khmaladze (1981). We show that our recursive projection operator estimated using the estimated derivatives is going to transform the vector $\hat{\rho}_{\hat{\varphi}}^{(m)}$ into random variable that has the same asymptotic distribution as $\hat{\rho}_{\varphi_0}^{(m)}$.

Let us consider the sequence of vectors $\{\hat{\rho}_{\hat{\varphi}}^{(m)}(i)\}_{i=1}^{m}$. Now according to drift equation (23) $\{\hat{\rho}_{\hat{\varphi}}^{(m)}(i)\}_{i=1}^{m}$ distribution will depend asymptotically only on the estimation error and the distribution of $\hat{\rho}_{\varphi_0}^{(m)}$. Let us consider recursive LS estimation of β parameter in following set of equations

$$\left\{\begin{array}{l} \hat{\rho}_{\hat{\varphi}}^{(m)}(j) = \hat{\eta}_{\hat{\varphi}}^{(m)}(j) \times \beta + \mu_{j,j}, \\ \hat{\rho}_{\hat{\varphi}}^{(m)}(j+1) = \hat{\eta}_{\hat{\varphi}}^{(m)}(j+1) \times \beta + \mu_{j,j+1}, \\ \vdots \\ \hat{\rho}_{\hat{\varphi}}^{(m)}(m) = \hat{\eta}_{\hat{\varphi}}^{(m)}(m) \times \beta + \mu_{j,m}, \end{array}\right\} \tag{26}$$

for $j = 1, \ldots, m-1$.

Note that the condition $j \leq m-1$ allows the above system of equations to consist of at least two equations.
According to (23) we have that $\hat{\beta}_j$ is the estimate of $\mathrm{vec}(\varphi_0^{(m)} - \hat{\varphi}^{(p)})$ estimation error using the information from the sequence $\{\hat{\rho}_{\hat{\varphi}}^{(m)}(j), \ldots, \hat{\rho}_{\hat{\varphi}}^{(m)}(m)\}$ for $j = 1, \ldots, m-1$ and $\mu_{j,i}$ are errors centered in zero for each j'th iteration of LS. Now the solution to (26) for β_{j+1} is a feasible recursive LS estimator given by

$$\hat{\beta}_{j+1} = \left(\sum_{i=j+1}^{m} \hat{\eta}_{\hat{\varphi}}^{(m)}(i)' \hat{\eta}_{\hat{\varphi}}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{m} \hat{\eta}_{\hat{\varphi}}^{(m)}(i)' \hat{\rho}_{\hat{\varphi}}^{(m)}(i),$$

which may be interpreted as block by block estimation. Now, we may define operator $\Im_{\varphi}^{(m)}$

$$\Im_{\varphi}^{(m)}\{(\hat{\rho}_{\theta}^{(m)}(j))\} = \hat{\rho}_{\theta}^{(m)}(j) - \hat{\eta}_{\varphi}^{(m)}(j) \left(\sum_{i=j+1}^{m} \hat{\eta}_{\varphi}^{(m)}(i)' \hat{\eta}_{\varphi}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{m} \hat{\eta}_{\varphi}^{(m)}(i)' \hat{\rho}_{\theta}{}^{(m)}(i) \tag{27}$$

for $\varphi, \theta \in \Phi^{(p)}$. Note that according to (27) the feasible estimator of $\Im_{\varphi_0}^{(m)}$ would be $\Im_{\hat{\varphi}}^{(m)}$. The motivation for (27) is that by *Proposition* 1 we have

Lemma. *Under* ***A1-A2*** *and* H_0 *we have asymptotically*

$$\Im_{\varphi_0}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) = \Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j)) + o_P(1),\ j = 1, \ldots, m \tag{28}$$

and for N *going to infinity we are going to have*

$$\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) \xrightarrow{p} \Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j)),\ j = 1, \ldots, m, \tag{29}$$

The proof of the *Lemma* has been moved to Appendix. The main problem is consistent estimation of the derivatives by its empirical counterparts. If $\Im_{\varphi_0}^{(m)}(\cdot)$ may be consistently estimated it implies by (28) that $\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))$ is going to converge in probability to $\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j))$. Note that $\Im_{\varphi}^{(m)}(\cdot)$ is a linear transform of $\hat{\rho}_{\varphi}^{(m)}(j+1), \ldots, \hat{\rho}_{\varphi}^{(m)}(m)$ whose coefficients have to be estimated and these coefficients are unique by satisfying the conditions given by the *Lemma*. Now the covariance of $\Im_{\hat{\varphi}}^{(m)}$ has has to be taken into account in calculating Box-Pierce statistic (17). In general we have that $\sqrt{N}\tilde{\rho}_{\varphi_0}^{(m)}(j)$ are distributed as iid standard normals for $j \geq 1$. Note that its martingale transform $\sqrt{N}\Im_{\hat{\varphi}}^{(m)}(\rho_{\varphi_0}^{(m)}(j))$ will asymptotically converge to mean zero normals. By *Lemma* we have that $\sqrt{N}\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))$ are asymptotically distributed as $\sqrt{N}\Im_{\hat{\varphi}}^{(m)}(\rho_{\varphi_0}^{(m)}(j))$ with variance

$$\widehat{Avar}\left(\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))\right) = \Big[I_{d^2} + \hat{\eta}_{\hat{\varphi}}^{(m)}(j)\big(\sum_{i=j+1}^{m} \hat{\eta}_{\hat{\varphi}}^{(m)}(i)'\hat{\eta}_{\hat{\varphi}}^{(m)}(i)\big)^{-1}\hat{\eta}_{\hat{\varphi}}^{(m)}(j)'\Big] \tag{30}$$

for $j \geq 1$.

Now note that in transformed FAR(1) model by (14) we have according to $\Im_{\varphi}^{(m)}$ defined as in (27) that the first autocorrelation $\hat{\rho}_{\hat{\varphi}}^{(m)}(1)$ is transformed by $m-1$ autocorrelations, $\hat{\rho}_{\hat{\varphi}}^{(m)}(2)$ is transformed by $m-2$ autocorrelations and so forth. If we are using explicitly m autocorrelations in this way, then the condition that has

to be met is

$$rank\left(\sum_{i=j+1}^{m}\tilde{\eta}_{\hat{\varphi}}^{(m)}(i)'\tilde{\eta}_{\hat{\varphi}}^{(m)}(i)\right)=p^2 \tag{31}$$

The implication of the above is that first covariances of order less that m could not be corrected because of singularity problems. The simplest solution to this problem is to use constant number of derivatives to transform a subset of m empirical autocorrelations. The number of autocorrelations used in correction algorithm is not bounded. Thus let us assume that we are going to use arbitrary number r of past autocorrelations in each of the equations. This modification of (27) would produce following projection operator

$$\begin{aligned}&\Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))=\hat{\rho}_{\hat{\varphi}}^{(m)}(j)-\hat{\eta}_{\hat{\varphi}}^{(m)}(j)\left(\sum_{i=j+1}^{j+r}\hat{\eta}_{\hat{\varphi}}^{(m)}(i)'\hat{\eta}_{\hat{\varphi}}^{(m)}(i)\right)^{-1}\sum_{i=j+1}^{j+r}\hat{\eta}_{\hat{\varphi}}^{(m)}(i)'\hat{\rho}_{\hat{\varphi}}^{(m)}(i)\\&j=1,\ldots,m-r,\ H\leq m-r,\ r<m.\end{aligned} \tag{32}$$

The rationale for fixing the number of derivatives used for recursive LS algorithm has numerical rather than statistical character. If we assume that m is large then using excessive number of derivatives of autocorrelations is going to decrease the goodness of fit of the $\Im_{\hat{\varphi}}^{(m)}$ operator. Note that given a VAR(1) specification, the autocorrelations and derivatives of autocorrelations of order $j+1, j+2, \ldots$ for any $j \leq m$ are going to converge to zero in expotential way. Thus the estimates of inverses of matrices of the form (31) are not asymptotically bounded. This problem may be treated as a property of resursive projection technique. Thus there is a tredeoff between the precision of the estimate and the number of lags considered r, given the number of observations N.

Now accounting for (30) and lack of estimation effect in $\hat{\rho}_{\hat{\varphi}}^{(m)}(0)$, we have according

to (17)

$$\hat{Q}_H^{\Im}(\hat{\varphi}^{(p)}) = N\sum_{j=1}^{H}\left\{\Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))'\Big[\widehat{Avar}\left(\Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))\right)\Big]^{-1}\Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))\right\} \tag{33}$$

where $\hat{Q}_H^{\Im}(\hat{\varphi})$ is BP statistic transformed by $\Im_{\hat{\varphi}}^{(m,r)}$ defined in (32) and estimated by empirical residuals dependent on $\hat{\varphi} \in \Phi^{(p)}$.
Now we may claim the theorem that is the implication of *Lemma*

Theorem 1. *Under assumptions* ***A1-A2*** *and Proposition 2 and 3 we have the following*

$$\hat{Q}_H^{\Im}(\hat{\varphi}) = \hat{Q}_H^{\Im}(\varphi_0) + o_P(1), \tag{34}$$

$$\hat{Q}_H^{\Im}(\hat{\varphi}) \xrightarrow{D} \chi^2_{(Hp^2)}, \ \textit{with } N \to \infty \tag{35}$$

under H_0 *for* $H = 1, 2, \ldots,\ H \leq m - r,\ r < m.$

The above result is the main reason that justifies our approach and allows for eliminating the estimation error in asymptotic distribution of Box-Pierce statistic. The crucial point is that we do not estimate the $\hat{\rho}_{\varphi_0}^{(m)}$ vector, because it would be impossible without estimating the asymptotic distribution of estimation error. It is clear that $\hat{Q}_H^{\Im}(\hat{\varphi})$ is a martingale transform of BP statistic defined in (17) and the othogonalisation procedure leading to elimination of serial correlation due to estimation error should lead asymptotically to $\chi^2_{(Hp^2)}$ distribution. Interesting point is that in multivariate setup it would be much harder to rewrite the $\Im_{\varphi}^{(m,r)}$ operator to fit the definition (16).
In the end we have to state a technical proposition concerning the consistency of estimation of $\Im_{\varphi_0}^{(m,r)}$. In order to assure it we have to propose the consistent estimators of empirical derivatives for $\varphi \in \Phi^{(p)}$. Note that $\nabla \mathrm{vec}\hat{\Gamma}_{\varphi}^{(p)}(j),\ j = 1, \ldots, m$ for $\varphi \in \Phi^{(p)}$ are obtained using the derivatives of $\hat{e}_n[\varphi]$ for $n = 1, \ldots, N$ with respect to φ'. In the following proposition we are claiming that under H_0 in general strong VAR(1) framework we will obtain the $o_P(1)$ convergence of $\hat{\zeta}_{\hat{\varphi}}^{(m)}$

with the proof moved to the appendix.

Proposition 2. *Under H_0 and **A1**-**A2** we have that derivatives matrix defined in* (20) *will satisfy*

$$\begin{aligned} &\nabla vec\hat{\Gamma}^{(p)}_{\hat{\varphi}}(i) \xrightarrow{p} \nabla vec\hat{\Gamma}^{(p)}_{\varphi_0}(i), \quad \text{for } i = 1, \ldots, m \\ &\hat{\zeta}^{(m)}_{\hat{\varphi}} = \hat{\zeta}^{(m)}_{\varphi_0} + o_P(1) \end{aligned} \tag{36}$$

where the matrix $\hat{\zeta}^{(m)}_{\varphi}$ for $\varphi \in \Phi^{(p)}$ satisfies the following equation

$$\hat{\zeta}^{(m)}_{\varphi} = \begin{bmatrix} \nabla vec\hat{\Gamma}^{(p)}_{\varphi}(1) \\ \vdots \\ \nabla vec\hat{\Gamma}^{(p)}_{\varphi}(m) \end{bmatrix} = -\frac{1}{N} \sum_{n=1}^{N-m+1} \left(\begin{bmatrix} A^{(p)\prime}_{n} \\ A^{(p)\ \prime}_{n+1} \\ \vdots \\ A^{(p)\ \ \prime}_{n+m-1} \end{bmatrix} \otimes I_d \otimes e_n[\varphi] \right). \tag{37}$$

The above result proposes well defined estimator of derivatives of autocorrelations evaluated at unknown $\varphi_0 \in \Phi^{(p)}$. Note that the stationarity of the process $X_n(t)$ following **A1**- **A2** would imply stationarity of the derivatives of residual autocorrelations following equation (14). We would also like to know the probability limit of $\hat{\zeta}^{(m)}_{\varphi_0}$. In order to continue we need to state the following

Proposition 3. *Assuming **A1**-**A2** and H_0 it is true that*

$$\bar{\zeta}^{(m)}_{\varphi_0} \xrightarrow{p} -E \left(\begin{bmatrix} A^{(p)\prime}_{n} \\ A^{(p)\ \prime}_{n+1} \\ \vdots \\ A^{(p)\ \ \prime}_{n+m-1} \end{bmatrix} \otimes I_d \otimes e_n[\varphi_0] \right) = O_p(1).$$

Proposition 3 is one of the ways to suggest the asymptotic effect of estimation error. If the derivatives of empirical autocovariances were converging to zero then by *Proposition 1* the asymptotic effect could not be corrected.

4 Asymptotic effect of eigenfunction basis $\hat{B}$ and eigenvalue $\hat{\Lambda}$ decomposition

In the previous section we presented the method of estimation of orthogonal projection operator $\Im_{\varphi_0}^{(m,r)}$ and the convergence of $\hat{Q}_H^{\Im}(\hat{\varphi}^{(p)})$ to the desired $\chi^2_{p^2H}$ distribution under "null" hypothesis. The problem however is the distortion error that comes from distribution of B_0 basis consisting of $v_1, \ldots, v_p$ functions. The assumption we are using is that the $O_p(N^{-\frac{1}{2}})$ estimator of the basis is available. The estimated eigenfunctions will be denoted as $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$. Now in reality we estimate the multivariate representation of FAR(1) according to (14)

$$e_{n+1}[\varphi^{(p)}] = A_{n+1}^{(p)} - \varphi^{(p)} A_n^{(p)} + \xi_{n+1}^{(p)}(p, \varphi^{(p)}), \tag{38}$$

where ξ_n is unobserved distortion error of the empirical residual coefficients for observed $X_n(t)$. In an obvious way $\xi_n^{(p)}$ affects the coefficients of $X_n(t)$, $A_n^{(p)}$ because we evaluate the coefficients of $X_n(t)$ in the estimated basis $\hat{B}$. In addition we have that it affects also the estimator $\hat{\varphi}^{(p)}$ during the OLS estimation. In general the distribution of $\xi_n^{(p)}$ could have an asymptotic effect on distribution of $\hat{Q}_H[\hat{\varphi}^{(p)}]$ and subsequently $\hat{Q}_H^{\Im}[\hat{\varphi}^{(p)}]$. Let us denote $\tilde{e}_n(\varphi^{(p)})$ as the empirical residual coefficients affected by the distribution of $\xi_n^{(p)}$ errors for $\varphi^{(p)} \in \Phi^{(p)}$.
In order to show the lack of asymptotic effect of $\xi_n^{(p)}$ we need to refer to *Proposition 1.* The drift equation (22) defines the asymptotic effect of estimated $\varphi_0^{(p)}$ and at the same time is a basis of correction algorithm based on recursive LS. Thus we need to state the following

Theorem 2. *Under H_0, A1-A2 and*

$$\hat{B} = B_0 + O_p(N^{-\frac{1}{2}}), \tag{39}$$

the following holds

$$\tilde{\gamma}_{\varphi}^{(m)} = \hat{\gamma}_{\varphi}^{(m)} + o_p\left(N^{-\frac{1}{2}}\right) \tag{40}$$

for $\varphi^{(p)} \in \Phi^{(p)}$*, where* $\tilde{\gamma}_{\varphi}^{(m)}$ *is residual autovovariance vector of* m *autocovariances evaluated in* $\tilde{e}_{n+1}[\varphi^{(p)}]$ *and* $\hat{\gamma}_{\varphi}^{(m)}$ *is evaluated in undistorted residual coefficients* $e_n[\varphi^{(p)}]$.

Note that *Theorem 2* implies that basis estimation errors $\xi_n^{(p)}$ would not matter in the probability limit in distribution of Box-Pierce statistic ie. there would be no asymptotic effect. Assumption (39) may be formulated by the means of relation between true eigenfunctions $B_0 = [v_1, \ldots, v_p]$ and estimated base $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$. The distortion errors $\xi_n^{(p)}$ come from the fact that we may write $\hat{v}_i = \sum_{n=1}^{\infty} \alpha_{ij} v_j$ with α_{ij} not being Kronecker deltas. Thus basis vectors $\hat{v}_i$ are incorrectly specified and the coefficients α_{ij} have some distribution around the true values with $E\alpha_{ii} = 1$ and $E\alpha_{ij} = 0$ for $i \neq j$. Now in order to show the analog of *Proposition 1* we may use the Fourrier decomposition of functions in base $\hat{B}$ in $\hat{B}_0$ and in some sense use the first Fourrier coefficients as a measure of deviation of $\hat{B}$ from the true basis B_0. We may show that

Proposition 4. *Defining the matrix* $\hat{\Delta}$ *of coefficients* α_i *being a Fourrier expansion of order two of functions* $\{\hat{v}_1, \ldots, \hat{v}_p\}$ *in the* $\{v_1, \ldots, v_p\}$ *basis as*

$$\hat{\Delta} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1p} \\ \alpha_{21} & \ddots & & \\ \vdots & & & \\ \alpha_{p1} & \ldots & & \alpha_{pp} \end{bmatrix}, \Delta_0 = I_p \tag{41}$$

the following conditions hold for any $X_n(t)$ *series under assumptions* ***A1-A2***

$$\hat{A}_n^{(p)} = \hat{\Delta} A_n^{(p)}, \tag{42}$$

for any $n \leq N$ *where* $\hat{A}_n^{(p)}$ *are coefficients of* $X_n(t)$ *in distorted basis. In addition for any* $\varphi \in \Phi^{(p)}$ *we have*

$$\tilde{e}_n[\varphi^{(p)}] = \hat{\Delta} A_n^{(p)} - \varphi^{(p)} \hat{\Delta} A_{n-1}^{(p)}, \tag{43}$$

for any $n \leq N$.

Note that in order to obtain equations (42) we have to use the linearity of scalar product and the fact that $L^2(0,1)$ is a Hilbert space. The main asymptotic result concerning eigenfunction estimation has been formulated in Hall & Hosseini-Nassab (2005). By means of generalized Fourrier decomposition it has been shown that the trace coefficients α_{ii} deviations δ_{ii} satisfy

$$\delta_{ii} = -\frac{1}{2}N^{-1}\sum_{k:k\neq i}(\lambda_i - \lambda_k)^{-2}\left(\int_0^1\int_0^1 Zv_i(s)v_k(t)dsdt\right)^2 + O_p(N^{-\frac{3}{2}})$$

with $\alpha_{ii} = 1+\delta_{ii}$. Now, Z is defined as covariance kernel error satisfying $Z = K-\hat{K}$ with

$$K(s,t) = E\{X_n(t) - EX_n(t)\}\{X_n(s) - EX_n(s)\}, \tag{44}$$

$$K(s,t) = \sum_{i=1}^{\infty}\lambda_i v_i(s)v_i(t). \tag{45}$$

and subsequent estimated $\hat{K}$. In practice we may obtain the K operator as $K = EX_n \otimes X_n$ according to a grid of X_n observations while $\hat{\Lambda}$ matrix of eigenvalues is the diagonalised counterpart of $\hat{K}$ (see Appendix). Thus Z in equation (41) is an error of covariance operator estimation which would affect eigenvalues matrix estimator diag $\hat{\Lambda} = [\hat{\lambda}_1, \ldots, \hat{\lambda}_p]$ and eigenfunction estimator $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$.
Now using the intermediate result in *Proposition 4* we may prove *Theorem 2*. In order to show lack of asymptotic effect of $\hat{\Delta}$ we need to show the negative analog of *Proposition 1*. The proof would consider using the derivatives of empirical residuals $\tilde{e}_n[\varphi^{(p)}]$ with respect to vecΔ evaluated in $\hat{\Delta}$.

Proposition 5. *Under **A1-A2** we have for $\varphi_0 \in \Phi^{(p)}$,*

$$\frac{\partial}{vec\Delta'}\tilde{\Gamma}_{\varphi}^{(p)}(i) = o_P(1), \tag{46}$$

for $i \leq m$.

5 Numerical ilustrations

In this section we are investigating size and power properties of our modification of BP test. The procedure that can be used as a comparison was developed in Gabrys et al. (2010) for exogenous regressors, with p-values obtained using $\chi^2(p^2H)$ distribution. In order to account for estimation error we will be also using $\chi^2(p^2(H-1))$ distribution. The characteristical feature of functional data is relatively small number of observed functions N that rarely reach 200 and the grid, ie. the number of observed values of each curve on $(0,1)$ that tends to be large and surely larger than N. Our modification of BP statistic will be denoted as DV transform. The models for Monte Carlo experiments have been proposed in Gabrys et al. (2010), we follow this design in order to get the clear comparison of methods. The motivation for assuming general Brownian Bridge errors is clear. In reality the errors may be any independent series of processes $\varepsilon_n(t)$. However, we do not want the errors to be dominating the entire model. Brownian Motion is more variable than Brownian Bridge under the same specification but we do not know where the Brownian Motion is going to end. Independently simulated Brownian Bridges are the closest counterpart of the iid errors found in the standart multivariate setup because it starts and ends always at zero.

5.1 Empirical size.

Data generating process follows strong FAR(1) model

$$\begin{aligned} X_{n+1}(t^*) &= \int_0^1 \varphi(s,t)X_n(s^*)ds + \varepsilon_{n+1}(t^*), \\ \varphi(s,t) &= 1.2e^{-\frac{(s^2+t^2)}{2}}, \\ \varepsilon_n(t^*) &= \mathrm{BB}_n(t^*). \end{aligned} \tag{47}$$

We simulated $k = 1000$ independent trajectories of autoregressive process following (47) of lenght N with respect to t^* grid. Note that the trace norm of φ has been estimated to be $||\varphi||_{\mathrm{tr}} = 0.85$. For each replication we estimated matrix coefficients of FAR(1) approximation $\varphi^{(p)}$ where p is the number of eigenfunctions considered

Table 1: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (47) *autoregressive model,* $N = 50$, $\#t^* = .05$, $p = 2$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
DV transform, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	4.5	4.7	4.4	5.1	6.9	11.2	19.8	37.0
	k=2	4.5	4.4	5.8	7.7	10.6	18.2	33.3	54.6
	k=3	3.6	4.6	6.6	11.2	16.5	31.2	50.7	74.8
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		0.6	1.4	0.8	0.8	0.7	0.8	1.2	0.9
$\chi^2((H-1)p^2)$		n.a	10.8	6.2	5.1	3.4	3.8	3.0	2.7

and then applied portmanteau test to the residuals. The nominal level of the test is $\alpha = 5\%$. The number of covariances used in computing $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ BP statistic is $H = 1, \ldots, 8$. In the functoinal data setup usually BP statistic may be derived for small number of lags due to relatively small number of observed functions. For DV modification of BP test, the H_0 is rejected when $\tilde{Q}^{\Im}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$. For the standard BP test we are presenting both results so H_0 is rejected when $\tilde{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2(H-1))$ or $\tilde{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$.

In *Table 1* we summarized the results of empirical size simulations for the model (47) with $N = 50$ and grid $\#t^* = .05$. Number of eigenfunctions estimated is $p = 2$, which is the reasonable approach for short series. The results for DV modification of Box-Pierce statistic are unsatisfactory for $H > 4$ due to small number of observed curves $N = 50$. In this case the most important factor are numerical errors of estimating the residuals rather than finite sample distributions. For low number of observations quality of eigenfunctions $\hat{B}$ estimators matter. In the range of low lag number H taken into consideration our test outperforms available BP tests proposed by Gabrys et al.

In *Table 2* we increased the number of grid points to more realistic case of

Table 2: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (47) *autoregressive model,* $N = 100$, $\#t^* = .01$, $p = 2$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	6.6	5.3	3.2	2.9	3.2	2.8	3.0	2.4
	k=2	4.3	3.4	4.9	3.8	2.7	3.3	3.6	2.7
	k=3	2.8	3.4	3.0	2.7	3.0	2.8	3.5	3.8
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		0.7	1.2	0.6	1.1	1.3	0.9	1.0	1.8
$\chi^2((H-1)p^2)$		n.a	11.8	6.6	6.3	5.4	3.3	4.3	4.0

$\#t^* = .01$ and increrased the number of observations. In this case the results of approximating 5% quantile of $\hat{Q}_{\varphi_0}$ improved for larger values of H. For the most important low values of H DV modification of BP test will outperform standart approach because there the estimation error would affect $\tilde{Q}_{\hat{\varphi}}(H)$ the most.

In *Table 3* we presented the result for the case where the grid has been decreased ten times compared to example in *Table 2*. We have shown that grid does not have asymptotical impact on distribution of $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ because sequential OLS is performed after estimation of eigenfunctions. Thus for resonable number of N observations results did not change significantly both for our out transform of BP statistic and the standart approach.

The impact of increasing the number of eigenfunctions p taken into consideration is depicted in *Table 4*. The results for DV transform of BP statistic start to become unreliable for $H > 5$ due to the fact that the number of parameters to be estimated including the derivatives and asymptotic variance increase more than two times compared with the case with $p = 2$.

Table 3: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (47) *autoregressive model,* $N = 100$, $\#t^* = .1$, $p = 2$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	6.1	4.3	4.7	4.2	2.9	3.0	2.4	2.2
	k=2	4.7	4.1	3.3	3.6	3.5	3.0	2.2	3.3
	k=3	3.3	4.2	3.4	3.8	2.9	2.5	3.1	3.8
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		0.7	1.1	1.6	0.8	1.6	1.2	1.1	0.6
$\chi^2((H-1)p^2)$		n.a	11.3	7.8	5.7	6.5	5.1	3.7	3.7

Table 4: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (47) *autoregressive model,* $N = 100$, $\#t^* = .01$, $p = 3$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	4.2	4.6	3.8	2.9	3.7	5.7	12.2	24.2
	k=2	2.6	2.4	3.2	3.6	5.8	13.0	21.7	51.1
	k=3	1.6	2.7	2.1	5.2	9.1	21.9	47.7	99.6
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		0.1	0.1	0.2	0.3	1.0	0.3	0.4	0.4
$\chi^2((H-1)p^2)$		n.a	11.6	7.3	4.5	5.2	4.0	2.8	2.7

Table 5: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (47) *autoregressive model,* $N = 200$, $\#t^* = .01$, $p = 3$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	4.7	3.5	2.6	3.4	2.8	2.3	2.4	1.8
	k=2	3.0	3.0	3.8	3.5	2.8	2.5	2.2	2.2
	k=3	3.7	2.4	3.4	3.3	2.5	1.9	2.5	1.8
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		0.0	0.1	0.0	0.3	0.6	0.7	0.6	0.6
$\chi^2((H-1)p^2)$		n.a	10.8	7.6	6.5	5.5	3.9	4.6	4.2

In *Table 5* we increased the number of observed curves two times from $N = 100$ to $N = 200$ while keeping the same grid and $p = 3$. The results went back to the situation from *Table 2* with $N = 100$, $\#t^* = .01$ and $p = 2$.

In order to see Monte Carlo simulated size characteristics for DV modification of BP statistic $\tilde{Q}_{\hat{\varphi}}(H)$ we introduce weakly dependent FAR(1) model

$$\begin{aligned} X_{n+1}(t^*) &= \int_0^1 \varphi(s,t) X_n(s^*) ds + \varepsilon_{n+1}(t^*), \\ \varphi(s,t) &= 1.2 e^{-\frac{(s^2+t^2)}{2}}, \\ \varepsilon_n(t^*) &= \mathrm{BB}_n(t^*) \times .\mathrm{BB}_{n-1}(t^*) \end{aligned} \tag{48}$$

Clearly model (48) is a FAR(1) but in this case functional errors $\varepsilon_n(t)$ follow the weakly dependent process. In *Table 6* we summarized the results in the intermediate case of $N = 100$ observations and $p = 2$. The results are comparable to the similar case of strong FAR(1) model presented in *Table 2*. Note that in both cases we were estimating the asymptotic variance matrix ie. we were acting agnostically without knowing the true model.

Table 6: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (48) *autoregressive model,* $N = 100$, $\#t^* = .05$, $p = 2$, *Brownian Bridge residuals*

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	4.1	3.7	3.9	4.9	4.1	5.5	6.2	8.9
	k=2	4.7	3.7	3.3	4.6	4.6	5.8	8.6	10.4
	k=3	3.4	4.5	4.6	6.1	4.8	7.0	8.7	9.4
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		2.1	3.9	2.2	3.3	3.3	3.2	2.9	2.2
$\chi^2((H-1)p^2)$		n.a	19.9	12.3	12.1	9.1	8.1	6.8	6.1

5.2 Empirical power

We are simulating the following FAR(1) model with autocorrelated functional errors

$$X_{n+1}(t^*) = \int_0^1 \phi(s,t) X_n(s^*) ds + \varepsilon_{n+1}(t^*), \tag{49}$$

$$\phi(s,t) = 1.2e^{-\frac{(s^2+t^2)}{2}}, \tag{50}$$

$$\varepsilon_n(t^*) = \int_0^1 \theta(s,t) X_n(s^*) ds + \upsilon_n(t^*), \tag{51}$$

$$\theta(s,t) = 0.9e^{-\frac{(s^2+t^2)}{2}}, \tag{52}$$

$$\upsilon_n = BB_n(t^*). \tag{53}$$

As in the size Monte Carlo experiment the number of simulated trajectiories is $k = 1000$ of lenght N with respect to t^* grid. Note that the trace norm of φ is the same as in previous experiments $||\varphi||_{\text{tr}} = .85$ and the norm of operator θ driving the FAR(1) process $\varepsilon_n(t)$ is $||\theta||_{\text{tr}} = .66$. For each replication we estimated matrix coefficients of FAR(1) approximation $\varphi^{(p)}$ where p is the number of eigenfunctions considered and then applied portmanteau test to the residuals. The nominal level of the test is $\alpha = 5\%$. The number of covariances used in computing $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ BP

Table 7: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (48) *autoregressive model,* $N = 80$, $\#t^* = .05$, $p = 2$, *FAR(1) errors* $\varepsilon_n(t)$, *Brownian Bridge distortions* $\upsilon_n(t)$

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	83.3	84.9	76.4	63.2	54.9	45.0	35.2	31.2
	k=2	91.1	87.6	80.6	65.2	52.2	43.7	36.8	29.6
	k=3	91.9	88.4	80.3	66.2	58.0	44.8	38.8	35.0
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		96.7	96.8	95.2	92.0	89.7	89.6	84.6	81.7
$\chi^2((H-1)p^2)$		n.a	99.5	99.2	97.4	95.8	95.8	93.2	89.2

statistic is $H = 1, \ldots, 8$. For DV modification of BP test, the H_0 is rejected when $\tilde{Q}^{\Im}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$. For the standard BP test we are presenting both results so H_0 is rejected when $\tilde{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2(H-1))$ or $\tilde{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$. Clearly model (49) does not satisfy H_0 because $\varepsilon_n(t)$ are serially correlated.

The realistic example with $N = 80$, $p = 2$ and $\#t^* = .05$ has been sumarized in *Table 7*. The rejection rates are comparable to standart method following Gabrys et al. (2011) for low values of evaluated lags of $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$. This characteristic will be more evident with lower values of observations N as shown in *Table 8*. In general DV modification of BP statistic shows better size than power characteristics. However, usually in functional data only the low number of lags of autocorrelations H may be taken into consideration.

In *Table 9* we are presenting results for extreme case with $N = 30$ and improved grid $\#t^* = .01$. Evidently better results for larger values of DV modification of BP statistic are coming from numerical errors of estimation of derivatives. Another aspect is that better grid does not help much as far as results are concerned. On the other hand our results are not worse than estimates given by Gabrys et al. (2011).

In *Table 10* we increased the number of observations from $N = 30$ to $N = 100$. In this case results of our method are comparable to the alternative approaches up

Table 8: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (48) *autoregressive model,* $N = 30$, $\#t^* = .01$, $p = 2$, *FAR(1) errors* $\varepsilon_n(t)$, *Brownian Bridge distortions* $\upsilon_n(t)$

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	25.6	21.3	25.1	37.0	58.7	99.5	100	100
	k=2	29.9	31.5	39.2	58.3	99.9	100	100	100
	k=3	33.3	39.5	59.2	99.8	100	100	100	100
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		34.8	36.1	29.6	21.6	21.5	18.2	15.0	12.3
$\chi^2((H-1)p^2)$		n.a	74.0	57.2	44.4	42.2	33.2	28.7	24.7

Table 9: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (48) *autoregressive model,* $N = 100$, $\#t^* = .01$, $p = 2$, *FAR(1) errors* $\varepsilon_n(t)$, *Brownian Bridge distortions* $\upsilon_n(t)$

	H	1	2	3	4	5	6	7	8
Projection method, number of derivatives used, k									
$\chi^2(Hp^2)$	k=1	95.0	94.5	91.7	87.5	76.5	65.6	58.6	50.0
	k=2	97.6	96.4	93.0	87.3	78.2	66.3	58.3	48.7
	k=3	98.2	96.2	93.0	88.7	77.8	66.9	60.4	50.6
Standart Box Pierce test, Gabrys et al., 2011									
$\chi^2(Hp^2)$		99.5	99.5	98.9	98.3	96.3	96.6	97.0	94.5
$\chi^2((H-1)p^2)$		n.a	99.9	99.8	99.5	99.0	98.8	98.8	98.0

to $H = 4$. With increasing number of observations both methods tend to reject surely, however the treshold N will be higher for DV modification. (I will do also 10% and 20% - it will be probably better and will look nicer...)

5.3 Application to financial markets for minute data

We are going to apply the methods developed in the paper for intradaily data of noted prices of IMB corporation in NASDAQ index in the period $6/05/2006 - 6/23/2006$. There are several approaches available because the data contains Open/High/Low/Close prices (*Figure 1*). The following analysis will be using Open prices, however the alternative data does not differ greatly because the differences are contained in intraminute periods. The minute data allows for very precise estimation of the eigenfunctions of the process so the estimation error of $\hat{B}$ estimation is relatively low. The grid t^* is every minute from 9.31 to 16.00 which gives $\#t^* = 390$. The number of observartions is $N = 15$, however this number seems sufficient to estimate the FAR(1) operator and test the specification of the model. The eigenvalues matrix of diagonalized covariance operator $\hat{\Lambda}$

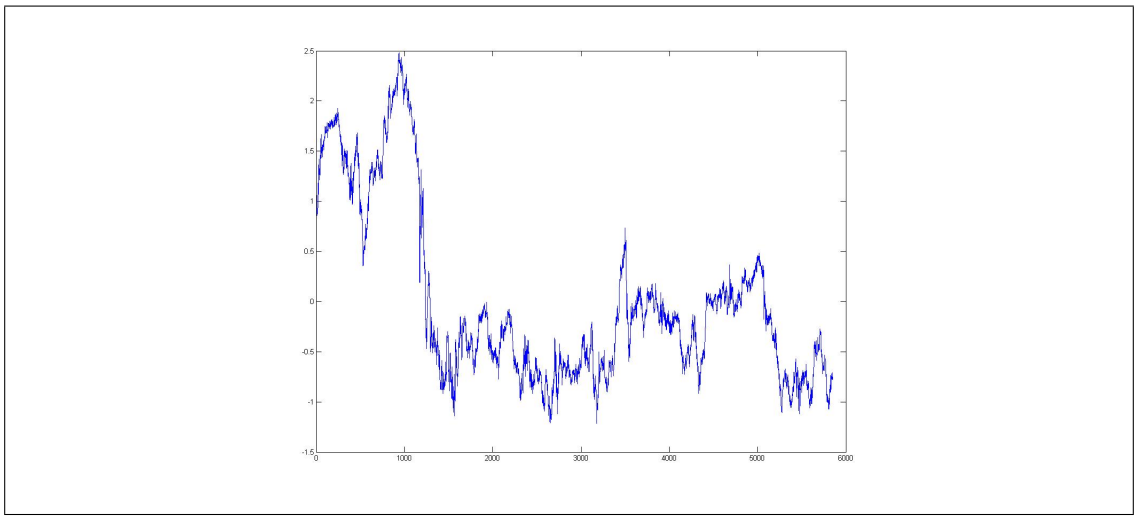

Figure 1: IBM opening prices on NASDAQ with minute frequency 6/05/2006-6/23/2006.

was estimated as

$$\hat{\Lambda} = \begin{bmatrix} .6921, 0, 0, 0 \\ 0, .0421, 0, 0 \\ 0, 0, .0182, 0 \\ 0, 0, 0, .0063 \end{bmatrix},$$

which means that there is one dominating signal with the second one ten times weaker and the rest decreasing in significance approximately three times monotonically. Thus for such low number of observations we estimate $\varphi_0^{(p)}$ with $p = 2$ number of eigenfunctions p. We are setting the hypothesis that the prices of IMB follow the model

$$\begin{aligned} &IBM_{n+1}(t) = \int_{9.30}^{16.00} \varphi(s,t) IBM_n(s) ds + \varepsilon_{n+1}(t), \\ &\varepsilon_n(t) \text{ are uncorrelated,} \end{aligned} \tag{54}$$

where $\{IBM_n(t)\}_{n=1}^{15}$ are the prices of IBM on the considered dates divided into daily curves on a minute grid. Now estimating $\varphi_0^{(p)}$ gives the estimate

$$\hat{\varphi}^{(p)} = \begin{bmatrix} 0.51, 0.87 \\ -0.13, -0.01 \end{bmatrix},$$

which produces the estimate $\tilde{\varphi}(s,t)$ shown in *Figure 2*. The kernel shown in *Figure 2* is typical representation of the gaussian kernel estimated for low number of observations. If trace norm of Ψ Hilbert-Schmidt operator defined by $\varphi(s,t)$ is also low then the estimation error of $\varphi^{(p)}$, concentrates close to the beginning of the grid. We may note the rising convex surface around $(0,0)$ instead of convex surface decreasing expotentially to $(1,1)$.

In order to test the specification of the model we are testing the lack of correlation of the residuals $\tilde{e}[\hat{\varphi}^{(p)}]$ using DV modification of BP statistic $\tilde{Q}_{\hat{\varphi}}^{\Im}(H)$ using the $\chi^2(4H)$ treshold and unmodified BP statistics $\tilde{Q}_{\hat{\varphi}}(H)$ following $\chi^2(4H)$ and $\chi^2(4(H-1))$ tresholds. The Type I Error $\alpha = 5\%$.

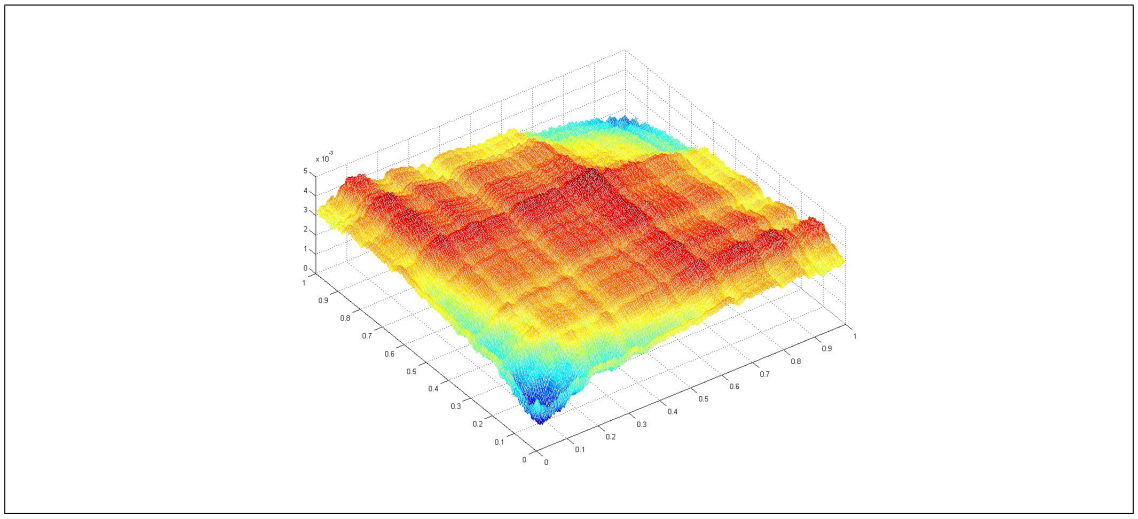

Figure 2: Estimated φ kernel of FAR(1) model (54).

The results of testing H_0 hypothesis were summarized in *Table 10*. Evaluating BP test by both DV modification of $\tilde{Q}_{\hat{\varphi}}(H)$ statistic and using the standart approach does not suggests rejecting H_0 hypothesis. Now for only $N = 15$ observations taking into account more than $H = 2$ lags could lead to wrong conclusions because number of observation used in computing the BP statistics is equal to $N - H$. The reason we did not compute the values of modified values of $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ is the inability to diagonalize asymptotic covariance matrix $\hat{G}_{\hat{\varphi}}$ and subsequently autocovariance of projection operator $\Im_{\hat{\varphi}}$. Such cases may happen for small number of observations and are a numerical rather than statistical problem.
It may be surprising that eliminating the first day would drastically change the results of estimation. Estimating the noncorrelated functional series leads to similar results as in multivariate case. Firstly the approximated operator $\hat{\varphi}^{(p)}$ is close to unity. Secondly the values of BP statistic tend to be low due to overfitting. In the described case we obtained the estimated kernel shown in *Figure 3*.

Table 10: *Evaluation of DV modification of BP statistic and Standart BP statistic for estimated residuals in model* (54)

	H	1	2	3
Projection method BP statistic, number of derivatives used, k				
$\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$	k=1	4.5268	9.1763	not computed
	k=2	2.4644	not computed	not computed
Standart Box Pierce statistic, Gabrys et al., 2011				
$\tilde{Q}_{\hat{\varphi}}(H)$		3.8336	8.4087	10.0941
p-values for asymptotic χ^2 *distributions*				
$\chi^2_{0.95}(Hp^2)$		9.4877	15.5073	21.0261
$\chi^2_{0.95}((H-1)p^2)$		n.a	9.4877	15. 5073

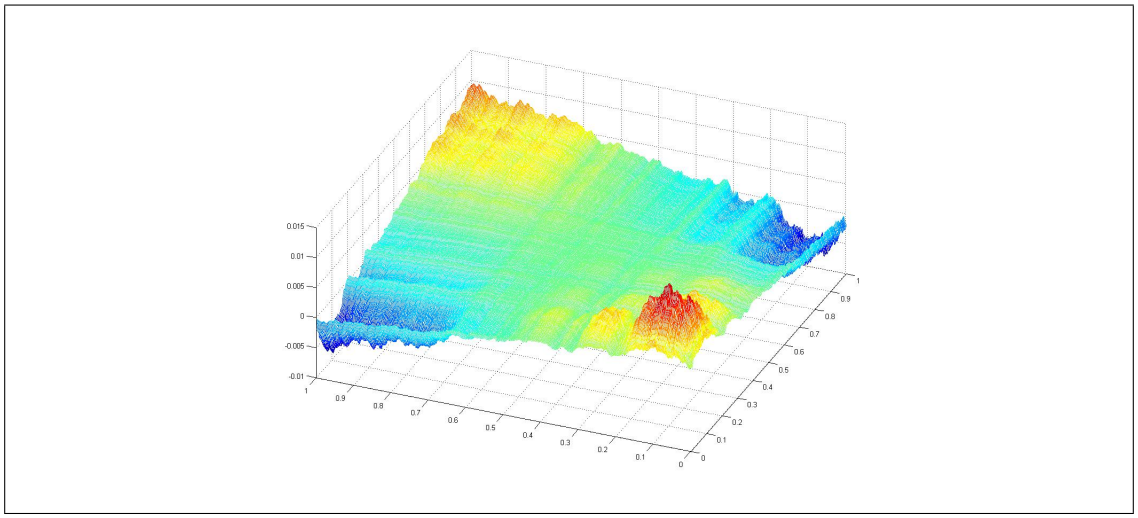

Figure 3: Estimated φ kernel of FAR(1) model (54) in the 6/06/2006-6/23/2006 range.

The interpretation of such a kernel is straightforward. Given that the sequence $\{IBM_n(t)\}_{n=2}^{15}$ is serially noncorrelated the kernel $\varphi(s,t)$ defined in model (54) will satisfy the continuity condition of the observed data, thus only the beginning and the end of the grid of the observed minute period would matter. All the information contained in the observed values of the functions would not matter at all and would be estimated as depending on the noncorrelated errors $\varepsilon_n(t)$. It should be noted that for such a short sequence of observations we cannot draw

final conclusions, however the functional approach could possibly lead to different interpretation of the data compared to multivariate approach.

Appendix

- **Specification of functional FAR(1) model**

Note that formulation (3) may look specific but in fact it is quite general in infinite dimensional approach. In multivariate framework it is possible to rewrite VAR(a) for any $a \in \mathbb{N}$ and $a \geq 1$ as VAR(1), however by the cost of increasing the dimension of the data. In infinite dimension space FAR(a) may be directly represented as FAR(1) using the Hilbert Hotel argument (see Bosq, 2000).
We are treating each element of a sequence X_n as an element of a suitable hilbertian space. Here we are assuming that $X_n(t)$ are elements of the space of square integrable functions $L^2([0,1])$ with the standard product of the form $\left\langle X_n, f \right\rangle = \int_0^1 X_n(s)f(s)ds$, that also implies the euclidean norm $||X_n||_2 = \sqrt{\int_0^1 |X_n(s)|^2 ds}$.
The direct parameter of the model is an operator $\Psi : L^2([0,1]) \to L^2([0,1])$ defined as

$$\Psi(X_n)(t) = \int_0^1 \varphi(s,t)X_n(s)dsdt, \tag{55}$$

with the the assumption that $\Psi \in \mathcal{C}$, where $\mathcal{C}$ is a set of all compact, bounded integral operators defined on $L^2(0,1)$. Such class of operators are known as Hilbert-Schmidt integral operators with $\varphi : [0,1] \times [0,1] \to \mathbb{R}$ being called the Hilbert-Schmidt kernel with the assumption that

$$\int_0^1 \int_0^1 |\varphi(s,t)|^2 dsdt < \infty. \tag{56}$$

which implies that we may choose the parameter space $\varphi \in \Phi$, where Φ is a space of functions that satisfy (56). In general any operator kernel φ defines operator Ψ but there exist such compact operators that do not have their kernel representation. Thus by (55) we are assuming that kernel φ fully parametrizes FAR(1) model.
The reason for formulation of a FAR(1) model using Hilbert-Schmidt operators is twofold. On one hand we would expect the linear operator to be bounded. Then by Banach theorem if it is bounded then it is continuous. Compactness of a linear

operator between Hilbert spaces is sufficient for both continuity and boundedness. Now in general each compact linear operator $T : \mathcal{H} \to \mathcal{H}$ in Hilbert space $\mathcal{H}$ may be represented as

$$T = \sum_{n=1}^{\infty} \lambda_n \left\langle f_n, \cdot \right\rangle g_n, \tag{57}$$

where $f_1, f_2, \ldots,$ and $g_1, g_2, \ldots,$ are orthonormal sets and the sequence λ_n may have the focal point only at 0. Thus choosing Hilbert-Schmidt operator allows for truncation of estimated parameter φ in a finite basis.

In a general functional data setup we have to choose the most suitable base to represent $X_n(t)$ and φ as an infinite sum of time dependent processes. In this sense each function $X_n(t)$ may be represented by infinite series of coefficients in this base thus it is essential that Hilbert-Schmidt operators behave similar to linear transformations in finite dimensional spaces as far as eigenvalue decomposition is concerned. The second argument applies to Ψ which may be represented in the same base implicitly through $\varphi(s,t)$ decomposition. If we would be able to choose the basis of $L^2(0,1)$, then the same would apply to φ which is an element of $L^2((0,1)\times(0,1))$ by (56). Also it is worth noting that if condition $\varphi(s,t) = \overline{\varphi(s,t)}$ is satisfied than spectral theorem applies so we may choose the base with respect to spectral decomposition of Ψ using Karhunen-Loéve decomposition.

- **Estimation of a basis in functional FAR(1) model**

Principal components concept comes from multivariate analysis. The idea is to convert the set of given variables in such a way to get the set of uncorrelated variables. The problem may be described as reducing the dimensionality of data by means of approximation(see Gorban, Zinovyev (2009)). Principal components under mild conditions define the best finite orthonormal basis for the set of observable functions. The problem of finding principal components may be equivalently defined as a problem of finding the modes of highest variation. The method is based on diagonalisation of covariance operator of observed series of curves $X_n(t)$

defined as

$$\mathrm{C}(X, g) = E\left\langle X_n, g \right\rangle X_n$$

given that $EX_n = 0$ and X_n is stationary. Now the key point is that operator C may be directly observed in the data and has the following empirical analogue

$$\hat{\mathrm{C}}(X, g) = \frac{1}{N}\sum_{n=1}^{N}\left\langle X_n, g \right\rangle X_n, \tag{58}$$

for each g, which follows from assumption of $X_n, g \in \mathcal{H}$ and stationarity. Now the following condition holds in general strictly for variance operator (see Ferraty, Vieu (2006))

$$E\left\langle X_n, X_n \right\rangle = E\left(X_n \otimes X_n'\right).$$

Taking the analogue

$$\hat{\mathrm{C}}(X, X_n) = \left(\frac{1}{N}\sum_{i=1}^{N}\left(X_i \otimes X_i'\right)\right) X_n = \hat{D}X_n$$

we obtain the $t^* \times t^*$ matrix $\hat{D}$ where t^* is the grid of the observed curves X_n. The effect of diagonalisation of $\hat{D}$ is the set of eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ and the eigenfunctions observed with respect to t^*-spaced grid $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p$. We may write

$$\hat{D} = \hat{B}^{-1}\hat{\Lambda}\hat{B}$$

where $\hat{B}$ was defined in (**??**) as matrix of eigenfunctions on $\{t^*\}$, while $\hat{\Lambda}$ is an estimator of eigenvalues matrix. Now, in general Λ does not have to be a diagonal but may have an alternative Jacobi form, but in practice it would mean that a process would be comprised of at least two equally strong signals which in practice

could happen in very specific applications. The values

$$\mathrm{diag}(\hat{\Lambda}) = \left[\hat{\lambda}_1, \ldots, \lambda_r\right]$$

represent estimated expected deviations of corresponding eigenfunctions. This algorithm is called Karhunen-Loéve decomposition of a $X_n(t)$ series. The standard procedure of decomposition of stationary process $X_n(t)$ involves choosing the number of functions in KH expansion $p \leq t^*$ which is large enough to approximate $\{X_n(t)\}$ up to 90% of total variation.

It is worth noting that this procedure does not involve assumptions on regresors and is always feasible. The important implication is the simple application of eigenvalue decomposition. We may note that by construction

$$\hat{\mathrm{C}}(X, \hat{v}_i) = \hat{\lambda}_i \hat{v}_i,$$
$$\mathrm{C}(X, v_i) = \lambda_i v_i$$

and by Karhunen-Loéve theorem we have

$$X_n(t) = \sum_{i=0}^{\infty} \left\langle v_i, X_n \right\rangle v_i(t)$$

Now based on the same assumptions we may write for independent stationary series $\{X_n(t)\}$

$$X_n(t) \sim \sum_{i=1}^{\infty} \lambda_i S_i v_i$$

where S_i are independent random variables and v_i are orthonormal. This gives the interpretation of double orthogonalisation of a process, in terms of probability space and orthogonality of eigenfunctions. In the Gaussian case S_i are standard

normal with the following statement holding

$$\lim_{n\to\infty}\sum_{i=1}^{p}\mathrm{C}(X_n, v_i) = X_n \text{ a.s.}$$

- **OLS estimation of $\varphi^{(p)}$**

Following (13) we have

$$\begin{bmatrix}\hat{x}_{(n+1)1}\\ \vdots \\ \hat{x}_{(n+1)p}\end{bmatrix} = \begin{bmatrix}\hat{\varphi}_{11} & \dots & \hat{\varphi}_{1p}\\ \vdots & \ddots & \\ \hat{\varphi}_{p1} & \dots & \hat{\varphi}_{pp}\end{bmatrix}\begin{bmatrix}\hat{x}_{n1}\\ \vdots \\ \hat{x}_{np}\end{bmatrix} + \begin{bmatrix}\hat{\epsilon}_{(n+1)1}\\ \vdots \\ \hat{\epsilon}_{(n+1)p}\end{bmatrix}, \tag{59}$$

where eigenfunctions evaluated on the t^* grid and $\begin{bmatrix}\hat{v}_1(t) & \dots & \hat{v}_p(t)\end{bmatrix}$ are estimates already obtained by eigenvalue decomposition. Thus it is possible to compute the Karhunen-Loéve approximations of $X_n(t^*)$ by taking

$$\hat{x}_{ni} = \int_{\{t^*\}} X_n(t^*)\hat{v}_i(t^*)dt^*,$$

for $n \leq N$, where $\int_{\{t^*\}}$ is any consistent quadrature approximation algorithm on $\{t^*\}$ grid. Numerical integration methods may differ depending on the assumptions imposed on $X_n(t)$. Now we may write

$$\hat{A}_{n+1} = \hat{\varphi}^{(p)}\hat{A}_n + \hat{e}_{n+1}$$

where $\hat{\varphi}^{(p)}$ and $\hat{e}_{n+1}$ which is followed by

$$\hat{A}_{n+1}\hat{A}_n' = \hat{\varphi}^{(p)}\hat{A}_n\hat{A}_n' + \hat{e}_{n+1}\hat{A}_n', \tag{60}$$

$$E\left(\hat{A}_{n+1}\hat{A}_n'\right)E\left(\hat{A}_n\hat{A}_n'\right)^{-1} = E\hat{\varphi}^{(p)}, \tag{61}$$

using that $E\hat{e}_{n+1}\hat{A}_n' = 0$. In order to obtain the estimator we may use empirical

analogue of the form

$$\hat{\varphi}^{(p)} = \sum_{i=1}^{N-1} \frac{1}{N-1} \left(\hat{A}_{i+1} \hat{A}_i' \right) \left(\hat{A}_i \hat{A}_i' \right)^{-1}.$$

(Consistency)

Note that instead of using direct inverses we may consider element by element estimates following (60). We have that

$$\begin{bmatrix} \hat{\varphi}_{11} & \dots & \hat{\varphi}_{1p} \\ \vdots & \ddots & \\ \hat{\varphi}_{p1} & \dots & \hat{\varphi}_{pp} \end{bmatrix} = \sum_{i=1}^{N-1} \frac{1}{N-1} \begin{bmatrix} \hat{x}_{(i+1)1}\hat{x}_{i1} & \dots & \hat{x}_{(i+1)1}\hat{x}_{ip} \\ \vdots & \ddots & \\ \hat{x}_{(i+1)p}\hat{x}_{i1} & \dots & \hat{x}_{(i+1)p}\hat{x}_{ip} \end{bmatrix} \times \frac{1}{N-1} \sum_{i=1}^{N-1} \begin{bmatrix} \hat{x}_{i1}^2 & \dots & \hat{x}_{i1}\hat{x}_{ip} \\ \vdots & \ddots & \\ \hat{x}_{ip}\hat{x}_{i1} & \dots & \hat{x}_{ip}^2 \end{bmatrix}^{-1}, \tag{62}$$

with

$$E \begin{bmatrix} \hat{x}_{n1}^2 & \dots & \hat{x}_{n1}\hat{x}_{np} \\ \vdots & \ddots & \\ \hat{x}_{np}\hat{x}_{n1} & \dots & \hat{x}_{np}^2 \end{bmatrix} = \begin{bmatrix} E\hat{x}_{n1}^2 & & \\ & \ddots & \\ & & E\hat{x}_{np}^2 \end{bmatrix}$$

using the fact that $E\hat{x}_{ni}\hat{x}_{nj} = 0$ for $i \neq j$ from the Karhunen-Loéve theorem and Assumptions [][]. Thus we may write by (62)

$$\hat{\varphi}_{j,i} = \sum_{i=1}^{N-1} \frac{1}{N-1} \hat{x}_{ni}\hat{x}_{nj}(\hat{x}_{ni}^2)^{-1}, \tag{63}$$

which is the analogue of the equation

$$\varphi_{j,i} = \lambda_i^{-1} E \left\langle X_{n-1}, v_i \right\rangle \left\langle X_n, v_j \right\rangle,$$

where λ_i is an eigenvalue associated to v_i. In most cases (63) is the most numerically stable method of estimating $\varphi^{(p)}$. In functional setup the number of observed curves N may be much lower than in the multivariate setup so it is reasonable

to avoid inverting covariance matrices during estimation. Clearly estimator based on (62) would require inverting $p \times p$ matrix. The alternative way to solve this problem is a rout taken in Gabrys, Horváth & Kokoszka (2010). Let us vectorize equation (59)

$$\begin{bmatrix} \hat{x}_{(n+1)1} \\ \vdots \\ \hat{x}_{(n+1)p} \end{bmatrix} = \left[\begin{bmatrix} \hat{x}_{n1} \\ \vdots \\ \hat{x}_{np} \end{bmatrix}' \otimes I_p \right] \text{vec}\varphi^{(p)} + \begin{bmatrix} \hat{\epsilon}_{(n+1)1} \\ \vdots \\ \hat{\epsilon}_{(n+1)p} \end{bmatrix}, \tag{64}$$

where estimated parameter $\text{vec}\varphi^{(p)}$ is the $p^2 \times 1$ vector. Note that a closed solution of above equation would require inverting large $p^2 \times p^2$ matrix, however it would be always invertible by construction following (64) (see Gabrys et al., 2011).

- **Proof of Proposition 1**

Proposition 1 is the statement of asymptotic effect of estimation error on the empirical autocovariances. At the same time it allows for algorithm of correction. This approach to the proof has been proposed in Delgado & Velasco (2011) in univariate case. The multivariate version has been proposed in Opuchlik (2012). In order to lessen the notation load we will write $\varphi \in \Phi^{(p)}$ instead of $\varphi^{(p)} \in \Phi^{(p)}$. We need to show that for each $j = 1, \ldots, m$

$$\text{vec}\left(\hat{\Gamma}^{(p)}_{\hat{\varphi}}(j)\right) - \text{vec}\left(\hat{\Gamma}^{(p)}_{\varphi_0}(j)\right) = \nabla\hat{\Gamma}^{(p)}_{\varphi_0}(j)(\text{vec}\hat{\varphi} - \text{vec}\varphi_0) + D_N(j)$$

with $D_N(j) = o_P(N^{-\frac{1}{2}})$. Now we could use the following approximation

$$D_N(j) = (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)' \left\{ \frac{\partial^2 \text{vec}(\Gamma^{(p)}_{\varphi}(j))}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi^*) \right\} (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)$$

with φ^* satisfying $||\varphi^* - \varphi_0||_2 \leq ||\hat{\varphi} - \varphi_0||_2$ but $\left\{ \frac{\partial^2 \text{vec}(\Gamma^{(p)}_{\varphi}(j))}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi^*) \right\}$ is a second order tensor which is cumbersome to handle. In order to stay in he standard linear setup we may write the following

$$\left[\text{vec}\left(\hat{\Gamma}^{(p)}_{\hat{\varphi}}(j)\right)\right]_{(i)} - \left[\text{vec}\left(\hat{\Gamma}^{(p)}_{\varphi_0}(j)\right)\right]_{(i)} = \left[\nabla\hat{\Gamma}^{(p)}_{\varphi_0}(j)\right]_{i'\text{th row}} (\text{vec}\hat{\varphi} - \text{vec}\varphi_0) + D_N(j)_{(i)}$$

where (i) denotes position in vector and $i = 1, 2, \ldots, p^2$, with $D_N(j)_{(i)} = o_P(N^{-\frac{1}{2}})$. Now we show that

$$D_N(j)_{(i)} = (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)' \frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi^*)(\text{vec}\hat{\varphi} - \text{vec}\varphi_0) \tag{65}$$

Now in order to prove that (65) holds it is enough to argument that $\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi^*) = O_P(1)$. Let us evaluate the expression (65) in $\varphi^* \in \Phi^{(p)}$ using *Proposition 2* and **A1-A2**

$$\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}} = \frac{1}{N-j}\sum_{n=1}^{N-j} \frac{\partial}{\partial \text{vec}\varphi} \left(\begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-p}^{(p)} \end{bmatrix}' \otimes I_d \otimes e_n[\varphi_*^{(p)}] \right)'_{i'\text{th row}} =$$

$$= \frac{1}{N-j}\sum_{n=1}^{N-j} \frac{\partial}{\partial \text{vec}\varphi} \left(\begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-p}^{(p)} \end{bmatrix}' \otimes I_p \otimes \left(A_n^{(p)} - \varphi_*^{(p)} A_{n-1}^{(p)}\right)_{i'\text{th row}} \right)'.$$

Now note that for each $n = 1, \ldots, N$

$$\frac{\partial}{\partial \text{vec}\varphi} \left(\begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-p}^{(p)} \end{bmatrix}' \otimes I_p \otimes \left(A_n^{(p)} - \varphi_*^{(p)} A_{n-1}^{(p)}\right)_{i'\text{th row}} \right)' = f(\{A_n^{(p)}\}) = O_p(1),$$

by **A1-A2** because it would be linear with respect to residuals $e_n[\varphi^{(p)}]$ and non-linear with respect to regressors $A_n^{(p)}$ which implies by LLN and *Proposition 2* that

$$\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}} = \left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_0^{(p)}} + o_P(1). \tag{66}$$

So from (66) and (65) we get that

$$D_N(j)_{(i)} = (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)' \left\{ \frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi_*^{(p)}) + o_P(1) \right\} (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)$$

with

$$D_N(j)_{(i)} = Op(N^{-1}),$$

assuming that $\hat{\varphi}^{(p)}$ is a root N estimator.
QED

Note that in this case which is in fact a standart VAR we could use the linearity of the second derivative of the empirical residuals with respect to φ. The result also applies in the nonlinear setup in VARMA case. However, in this case the proof needs more general motivation.

- **Proof of Proposition 2**

In order to show (36) it is convenient to present the form of derivatives $\hat{\zeta}_\varphi^{(m)}$ according to (37) in general for $\varphi \in \Phi^{(p)}$. From (14) we may write the system of equations

$$\nabla_\varphi e_{n+j}^{(p)}[\varphi] = \frac{\partial e^{(p)}[\varphi]}{\partial \text{vec}\varphi'}(\varphi) = -A_{n-1}^{(p)\,\prime} \otimes I_d, \tag{67}$$

for $n = 2, \ldots, N$. We are interested in obtaining the derivatives of residual autocovariances $\nabla_\varphi \hat{\Gamma}_\varphi(i)$ for $i = 1, \ldots, m$. We may write the following

$$\begin{aligned} \nabla \text{vec}\hat{\Gamma}_{\varphi,n}^{(p)}(j) =& \frac{\partial \text{vec}(e_{n+j}^{(p)}[\varphi] e_n^{(p)}[\varphi]')}{\partial \varphi'} = \\ =& \frac{\partial e_{n+j}^{(p)}[\varphi]}{\partial \varphi'} \otimes e_n^{(p)}[\varphi] + e_{n+j}^{(p)}[\varphi] \otimes \frac{\partial e_n^{(p)}[\varphi]}{\partial \varphi'} = \\ =& \nabla_\varphi e_{n+j}^{(p)}[\varphi] \otimes e_n^{(p)}[\varphi] + e_{n+j}^{(p)}[\varphi] \otimes \nabla_\varphi e_n^{(p)}[\varphi], \end{aligned} \tag{68}$$

with the second factor converging in probability to the limit $Ee^{(p)}_{n+j}[\varphi]\otimes\nabla_{\varphi}e^{(p)}_{n}[\varphi] = 0$ under H_0, so it may be omitted. Now we are proposing an estimator

$$\hat{\zeta}^{(m)}_{\varphi_0} = \begin{bmatrix} \nabla \mathrm{vec}\hat{\Gamma}^{(m)}_{\varphi}(1) \\ \vdots \\ \nabla \mathrm{vec}\hat{\Gamma}^{(m)}_{\varphi}(m) \end{bmatrix}_{(\varphi_0)} = \frac{1}{N}\sum_{n=1}^{N-m+1}\left(\hat{\zeta}^{(m)}_{\varphi_0}\right)_n \tag{69}$$

with

$$\left(\hat{\zeta}^{(m)}_{\varphi_0}\right)_n = \begin{bmatrix} \nabla \mathrm{vec}\hat{\Gamma}^{(p)}_{\varphi,n}(1) \\ \vdots \\ \nabla \mathrm{vec}\hat{\Gamma}^{(p)}_{\varphi,n}(m) \end{bmatrix}_{(\varphi_0)}.$$

Using (68) and (69) gives us (37).

In order to show (36) we need to prove that $\nabla \mathrm{vec}\hat{\Gamma}^{(m)}_{\varphi_0}(h)$ is estimated consistently for $h = 1, \ldots, m$. The sufficient condition that would imply the (36) statement is

$$\frac{1}{N}\sum_{n=i+h}^{N}({A^{(p)}_{n-i}}' \otimes I_p)\otimes(e^{(p)}_{n-h}[\hat{\varphi}] - e^{(p)}_{n-h}[\varphi_0]) = o_P(1) \tag{70}$$

for any $i = 1, \ldots, p$ after using $A_1\otimes B_1 - A_2\otimes B_2 = A_1\otimes(B_1 - B_2) + (A_1 - A_2)\otimes B_2$. Note that we may write

$$\sup_{0\le n\le N}||e^{(p)}_{n-h}[\hat{\varphi}] - e^{(p)}_{n-h}[\varphi_0]|| = O_P(N^{-\frac{1}{2}}), \tag{71}$$

which comes from the observation that estimation error $(\mathrm{vec}\hat{\varphi} - \mathrm{vec}\hat{\varphi})$ is asymptotically normal and at the same time by **A1-A2** and model FAR(1) formulation under the "null" we have that the operator

$$\varphi \in \Phi^{(p)} : \varphi \to \{e^{(p)}_n[\varphi]\}_{n=1}^{N}$$

is is linear. Thus the distance between $e^{(p)}_n[\hat{\varphi}]$ and $e^{(p)}_{n-h}[\varphi_0]$ is going to be bounded

uniformly by the distance between $\hat{\varphi} \in \Phi^{(p)}$ and $\hat{\varphi} \in \Phi^{(p)}$ multiplied by some constant. This result will hold also in nonlinear setting but we would need the assumptions on the higher order derivatives. Now by (71) we have

$$||\frac{1}{N}\sum_{n=i+h}^{N}({A_{n-i}^{(p)}}' \otimes I_p) \otimes (e_{n-h}^{(p)}[\hat{\varphi}] - e_{n-h}^{(p)}[\varphi_0]))|| \leq$$
$$\leq \sup_{0\leq n\leq N} ||{A_{n-i}^{(p)}}' \otimes I_p|| \sup_{0\leq n\leq N} ||e_{n-h}^{(p)}[\hat{\varphi}] - e_{n-h}^{(p)}[\varphi_0]|| = o_P(1)$$

because $\sup_{0\leq n\leq N} ||{A_{n-i}^{(p)}}' \otimes I_p|| = O_P(1)$.
QED

- **Proof of Proposition 3**

The Proposition has been proven in Opuchlik (2012) for general VARMA(a, b) setup.

- **Proof of Lemma**

In order to show (28) it is enough to observe that $\Im_{\varphi_0}^{(m)}$ is continuous and bounded under **A1-A2** and "null" hypothesis with $\hat{\rho}_{\hat{\varphi}}^{(m)} = \hat{\rho}_{\varphi_0}^{(m)} + o_P(1)$. Boundedness in probability of $\Im_{\varphi_0}^{(m)}$ comes from boundedness of $\hat{\eta}_{\varphi_0}$ and $\hat{\eta}_{\varphi_0}(i) = O_P(1)$ for $i = 1, \ldots, m$. The same arguments apply to $\Im_{\varphi_0}^{(m,r)}$.
In order to show (29) it is sufficient to note that $\Im_{\varphi_0}(\cdot) \xrightarrow{p} \Im_{\varphi_0}(\cdot)$ because from *Proposition* 2 and 3 we have that

$$\hat{\eta}_{\hat{\varphi}}(j) = \hat{\eta}_{\varphi_0}(j) + O_p(N^{-\frac{1}{2}}), \quad j = 1, \ldots, m$$

and $\Im_{\varphi_0}(\cdot)$ is a finite sum of derivatives and identity operators. Now $\hat{\rho}_{\hat{\varphi}}^{(m)} \xrightarrow{p} \hat{\rho}_{\varphi_0}^{(m)}$ so also $Avar\left(\hat{\rho}_{\varphi_0}^{(\Im,m)}(j)\right)$ with $\hat{\rho}_{\varphi_0}^{(\Im,m)}(j) = \Im_{\varphi_0}^{(m,r)}(\hat{\rho}_{\varphi_0}^{(m)}(j))$ is estimated consistently for $j = 1, \ldots, m$.
QED

- **Proof of Theorem 1**

Condition (34) follows directly from the *Lemma*. To prove (34) let us derive the asymptotic distribution of $\hat{Q}_H^{\Im}(\varphi_0)$, using the asymptotic limit of projection operator, $\Im_{\theta_0}^{(m,r)}$. Thus from (32) we have

$$\hat{\rho}_{\varphi_0}^{(\Im,m)}(j) = \Im_{\varphi_0}^{(m,r)}(\hat{\rho}_{\varphi_0}^{(m)}(j)) = \hat{\rho}_{\varphi_0}^{(m)}(j) - \hat{\eta}_{\varphi_0}^{(m)}(j) \left(\sum_{i=j+1}^{j+r} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\eta}_{\varphi_0}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{j+r} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\rho}_{\varphi_0}(i)$$

$j = 1, \ldots, m-r,\ H \leq m-r,\ r < m.$

where $\hat{\rho}_{\varphi_0}^{(m)}(i) \sim \mathcal{N}(0, I_{d^2})$ iid for $i = 1, \ldots, m$. However, note that $\hat{\rho}_{\varphi_0}^{(\Im,m)}(i)$ does not have asymptotically normal distribution because of the coefficients of the form

$$\hat{\eta}_{\varphi_0}^{(m)}(j) \left(\sum_{i=j+1}^{j+r} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\eta}_{\varphi_0}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{j+r} \hat{\eta}_{\varphi_0}^{(m)}(i)',$$

which are not identity matrices. Now in order to standarize the vector $\hat{\rho}_{\varphi_0}^{(\Im,m)}(i)$ we may write

$$\begin{aligned} \check{\rho}_{\varphi_0}^{(m)}(j) &= \hat{\rho}_{\varphi_0}^{(\Im,m)}(j) \times Avar\left(\hat{\rho}_{\varphi_0}^{(\Im,m)}(j)\right)^{-1/2}, \\ \check{\rho}_{\varphi_0}^{(m)}(j) &\sim \mathcal{N}(0, I_{p^2}), \end{aligned}$$

for j=1,...,m. Thus, we get that $\hat{Q}_H^{\Im}(\varphi_0) \to \chi^2_{(Hd^2)}$.
QED

- **Proof of Proposition 4**

The proof is simply the use of bilinearity of a scalar product. We have in general univariate case

$$\left\langle \sum_{j=1}^{p} \alpha_{ij} v_j, X_n \right\rangle = \sum_{j=1}^{p} \alpha_{ij} \left\langle v_j, X_n \right\rangle,$$

for any $i \leq p$. This proves (42). (43) follows from (42) used in quasi autoregreesive equation for $e_n[\varphi^{(p)}]$ residuals (14) for $\varphi^{(p)} \in \Phi^{(p)}$.

- **Proof of Proposition 5**

[TO BE WRITTEN]

- **Proof of Theorem 2**

[TO BE WRITTEN]

References

1. D.W.K. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 66 (1991) 817-858

2. A. Aue, S. Hormann, L. Horvath, M. Huskova, Sequential stability test for functional linear models, Univ. of California, 2010

3. G. Aneiros-Perez, R. Cao, J. Vilar- Fernandez, Functional methods for time series prediction, a nonparametric approach.

4. I. Berkes, S. Hormann, J. Schauer, Asymptotic results for the empirical process of stationary sequences, Stochastic Processes and their Applications 119, 2009

5. I. Berkes, R. Gabrys, L. Horvath, P. Kokoszka, Detecting changes in the mean of functional observations, J. R. Statist. Soc. B, 2009

6. Billingsley, P. (1999) *Convergence of probability measures* (2nd ed.), New York: Wiley.

7. D. Bosq Linear processes in function spaces. Theory and applications, Springer Verlag, 2000

8. Box, G.E.P., and Cox, D. (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society*, Ser.B, 26, 211-252.

9. Box, G.E.P., and Pierce, D. A. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, 65, 1509-1526.

10. Box, G.E.P., and Jenkins, M. J. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day.

11. Brockwell, P. J., and Davies, R. A. (1991), *Time Series: Theory and Methods*. New York: Springer Verlag.

12. J. M. Chiou, H. Muller, Diagnostics for functional regression via residual processes. Computational Statistics and Data Analysis, 15, 2007

13. Chittori, R. V. (1974), "Distribution of Residual Autocorrelations in Multiple Autoregressive Schemes", *Journal of the American Statistical Association*, 69, 928-34.

14. M. Delgado, C. Velasco, An asymptotically pivotal transform of the residuals sample autocorrelations with application in model checking, Journal of American Statistical Association, 2011

15. Delgado, M. A., Hidalgo, J., and Velasco, C. (2005), "Distribution Free Goodness-of-Fit Tests for Linear Processes", *Annals of Statistics*, 33, 2568-2609.

16. J.J. Faraway, Regression Analysis for a Functional Response, Technometrics, Vol. 39, No. 3, 1997

17. M. Fengler, W. Hardle, P.Villa, The dynamics of implied volatilities: A common principle components approach, Rev. Derivative Research, 2003

18. F. Ferraty, P.Vieu, Nonparametric functional data analysis: Theory and practice, Springer, 2006

19. P. Hall, J. Horowitz, Methodology and Convergence Rates for Functional Linear Regression, The Annals of Statistics, Vol. 35, No. 1, 2007

20. P. Hall and M. Hosseini-Nasab, On Properties of Functional Principal Components Analysis, Journal of the Royal Statistical Society, 2006

21. L. Horvath, Marie Huskova, P. Kokoszka, Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis 101, 2010

22. R. Gabrys, C. Horvath, P. Kokoszka, Test for error correlation in linear functional model, 2010

23. R. Gabrys, P. Kokoszka, Portmanteau test of independence for functional observations, Journal of the American Statistical Association, Vol. 102, No. 480, 2007

24. H. Cardot, F. Ferraty, P. Sarda, Functional Linear Model, Statistics and Probability Letters 45(1999), 1998

25. W. Gonzalez-Manteiga, A. Martinez-Calvo, Bootstrap calibration in functional linear regression models with applications, COMPSTAT'2010, Paris (France), 2010

26. A. N. Gorban, A. Y. Zinovyev, Principal Graphs and Manifolds, Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, Olivas E.S. et al Eds. Information Science Reference, Hershey, PA, USA, 2009.

27. S. Hormann, P. Kokoszka, Weakly dependent functional data, Annals of Statistics, 2010

28. Hosking, J. R. M. (1980). "The Multivariate Portmanteau Statistic", *Journal of American Statistical Association*, 75, 343-86.

29. Hosking, J. R. M. (1981). "Equivalent Forms of the Multivariate Portmanteau Statistic", *Journal of the Royal Statistical Society*, B43, 219-30.

30. Ljung, G. M. (1986), "Diagnostic Testing of Univariate Time Series Models", *Biometrica*, 73, 725-730.

31. Ljung, G. M., and Box, G. E. P. (1978), "On Measure of Lack Of Fit in Time Series Models", *Biometrica*, 65, 297-303.

32. Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Berlin: Springer Verlag.

33. A. Mas, Weak convergence in the functional autoregressive model, Journal of Multivariate Analysis 98, 2007

34. T. Mourid , Processus autoregressiifs d'ordre superieur. Acad Sci t.317(Ser. I):1167-1172, 1993

35. H. Muller, and U. Stadtmuller, Generalized Functional Linear Models, The Annals of Statistics, 33, 774-805, 2005

36. J. O. Ramsey, B. W. Silverman, Functional Data Analysis, Springer 2005

37. B. W. Silverman, Incorporating Parametric Effects into Functional Principal Components Analysis, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 4, 1995