# Specification Analysis of Functional Autoregressive Models

Maciej Opuchlik
Universidad Carlos III de Madrid
Economics Department

August 4, 2014

**Abstract**

We propose new tests for the correct specification of dynamic 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 expands this 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 using orthogonal projection operator on the sample autocorrelations of the reduced 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 while the asymptotic estimation error originating from eigenfunctions basis and eigenvalues vector estimation does not have an asymptotic effect on the distribution of modified Box-Pierce statistic under consideration. Our method is general, it 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. An application is presented concerning 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 $\varepsilon_n$ and observations $X_n$ are curves and $\Psi$ 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 adress in this paper is testing for correlation of residuals in a functional model formulated as (1) for Hilbert-Schmidt operators. 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 $\varepsilon_n$ with $X_n$ independent of $\varepsilon_n$.
The main problem in proposing a 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 parametric 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 a 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 projections of the observables $X_n(t)$ and thus of 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 approximations error in Karhunen-Loéve basis would be weakly dependent.
The article is organized as follows: in Section 2 we introduce the basic method of estimating the Karhunen-Loéve expansion and assumptions on the autoregressive functional model and goodness-of-fit portmanteau test being the benchmark given by Gabrys et al. (2011). Section 3 presents 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 assumption on $\varphi(s,t)$ that would grant stationarity of $X_n(t)$. Note that according to (3) the series $X_n(t)$ will be driven by a 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 $\varepsilon_n$ are as follows:

**A1:** $\varepsilon_n(t)$ are iid and $E||\varepsilon_n(t)||_2 < \infty$ for $-\infty \leq n \leq N$, where $||f||_2$ is defined as $\int_0^1 |f(t)|^2 dt$. We also assume mean zero errors $E\varepsilon_n(t) = 0$. **A2:** The kernel parameter space $\Phi$ satisfies $\varphi \in \Phi$ is measurable and

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

where $||f||_\infty = \sup_{t\in(0,1)} |f(t)|$.

The assumption **A1** has to be imposed in order to guarantee $X_n(t)$ to be stationary by assuming iid error curves $\varepsilon_n(t)$ and finitely variable by assuming finite variability of errors. For simplicity we do not need to estimate the intercept $\mu$ which would be the mean of error curves. As in the multivariate case independence of curves implies lack of correlation. The important condition following **A1** is that the observables sequence $\{X_n(t)\}$ satisfies

$$E\left(\int_0^1 X_n^2(t)dt\right)^2 < \infty. \tag{5}$$

Condition (5) assures asymptotic normality of a quasi-OLS algorithm used for estimation of functional model. It is also necessary for some asymptotic results (see Gabrys et al. 2007). Assumption **A2** is the analogue of sufficient condition for stationarity of observable $X_n(t)$ in multivariate case. It is not a necessary condition however and one may think about relaxing it by adding weaker assumptions. Note that condition (4) gives the characterization of parameter space $\Phi$ and $X_n(t)$ by (3). If we define $\mathcal{M}$ as the $\sigma$-algebra generated by the sequence $\{\dots, \varepsilon_n(t), \varepsilon_{n-1}(t), \dots\}$ 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 $\varphi$. Model (3) defines a member of a class of Hilbert-Schmidt operators which are bounded and continuous. Condition (4) implies

$$\int_0^1 \int_0^1 |\varphi(s,t)|^2 dsdt < \infty \text{ (Hilbert-Schmidt) .} \tag{6}$$

Finally $\Phi$ is a subclass of functions that would define FAR(1) model (3) to be driven by Hilbert-Schmidt integral operator. Note that $\Psi$ given in (1) in order to be Hilbert-Schmidt operator does not have to be defined through $\varphi \in \Phi$. Although the class of Hilbert-Schmidt operators is wider, it is enough to consider the most straightforward setup first.

Clearly in a general sense we are aiming to test the null 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**. In order to propose the test for error serial correlation we define the residuals with respect to model (3) as

$$\varepsilon_n[\varphi] = X_{n+1} - \int_0^1 \varphi(X_n), \tag{7}$$

for $\varphi \in \Phi$, while simplifying the notation. Clearly we are treating the residuals as functions of $\varphi$. In a general hilbertian setup we say that $\varepsilon_n$ and $\varepsilon_{n+i}$ curves are

uncorrelated if and only if

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

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 $\hat{\varphi}$. 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 $X_n(t)$ and thus $\varepsilon_n(t)$ on $p$-dimensional subspace of $L^2(0,1)$ spanned by an 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. We have given the definition of the covariance operator in Hilbert spaces in the Appendix.

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 each $\varphi \in \Phi$ is a Hilbert-Schmidt kernel. It follows that $\varphi \in \Phi$ may be decomposed as

$$\begin{aligned} &\varphi : (0,1)\times(0,1) \to \mathbb{R}, \\ &\varphi(t,s) = \sum_{i,j=1}^{\infty} \varphi_{j,i} v_i(t) v_j(s), \end{aligned} \tag{8}$$

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

$$\varphi_{j,i} = \left\langle \varphi(t,s), v_i(t)v_j(s) \right\rangle = \int_0^1 \varphi(t,s)v_i(t)v_j(s)dtds. \tag{9}$$

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 $\lambda_i$ is the eigenvalue associated to $v_i$. This equation is the immediate analog of OLS estimation of $\varphi^{(p)}$ that is presented in the Appendix. Note that having the estimator of $\varphi_0^{(p)} = [\varphi_{j,i}]_{\{1 \leq i,j \leq p\}}$ we can write the kernel estimator by truncating (8) as

$$\hat{\varphi}_p(t,s) = \sum_{i,j=1}^{p} \hat{\varphi}_{j,i} v_i(t)v_j(s), \tag{10}$$

being an exact multivariate OLS analog (see Appendix). Clearly $\hat{\varphi}_p(\cdot,\cdot) \in \Phi$ is not a 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

$$\tilde{\varphi}_p(t,s) = \sum_{i,j=1}^{p} \hat{\varphi}_{j,i} \hat{v}_i(t)\hat{v}_j(s), \tag{11}$$

with distribution of $\tilde{\varphi}^{(p)} \in \Phi$ possibly depending on the error of basis $B_0 = [v_1, \ldots, v_p]$ estimation. Note that $\Phi^{(p)}$ is the space of $p \times p$ matrices and is isometric with the previously defined $\Phi$ 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 $\varphi^{(p)} \in \Phi^{(p)}$ given the basis $B_0$.

In order to define the residuals for any kernel $\varphi_p \in \Phi$ we may rewrite the model (3) using (8) as

$$X_{n+1}(t) = \sum_{i=1}^{p} \sum_{j=1}^{p} \varphi_{j,i} \left\langle X_n, v_j \right\rangle v_i(t) + \gamma_{n+1}^{(p)}[\varphi](t) + \varepsilon_{n+1}[\varphi](t), \tag{12}$$

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. Note that decomposition of $X_n(t)$ for $n \leq N$ according to (8) holds for any $\varphi \in \Phi$ by fitting the residual curves $\varepsilon_n[\varphi] : \varphi \to \varepsilon_n[\varphi](t)$. Thus now we define the residuals $\varepsilon_n[\varphi]$ according to equation (12)

$$\varepsilon_{n+1}[\varphi] = X_{n+1} - \begin{bmatrix} v_1 & \dots & v_p \end{bmatrix} \begin{bmatrix} \varphi_{11} & \dots & \varphi_{1p} \\ \vdots & \ddots & \\ \varphi_{p1} & \dots & \varphi_{pp} \end{bmatrix}' \begin{bmatrix} \left\langle X_n, v_1 \right\rangle \\ \vdots \\ \left\langle X_n, v_p \right\rangle \end{bmatrix} + \gamma_{n+1}^{(p)}[\varphi] \quad (13)$$

for $\varphi \in \Phi$. Note that in (13) we denote the projection of $\varphi$ 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 the distortion curve caused by truncation of higher than $p$ order KL expansion factors of $X_n(t)$. Still we omit the error of $\varepsilon_{n+1}[\varphi](t)$ with respect to grid $t^*$ quadrature error. We are assuming that the grid $t^*$ chosen is equally spaced. Consideration of different grids would not affect our results, however it can be noted that this grid does not have to be optimal on general. Now, 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 the $p$ most important eigenfunctions.

In order to decrease the dimension of the problem we have to transform residuals defined as in (13) to the representation dependent on the coefficients in the $v_1, v_2, \dots, v_p$ basis. We are going to denote $X_n^{(p)}(t)$ as the Karhunen - Loéve expansion of $X_n(t)$ and $x_{ni} = \left\langle X_n, v_i \right\rangle$ as the 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 (13) as

$$\begin{aligned} &\begin{bmatrix} v_1(t^*) & \ldots & v_p(t^*) \end{bmatrix} \begin{bmatrix} \epsilon_{(n+1)1} \\ \vdots \\ \epsilon_{(n+1)p} \end{bmatrix} = \\ &\begin{bmatrix} v_1(t^*) & \ldots & v_p(t^*) \end{bmatrix} \left( \begin{bmatrix} x_{(n+1)1} \\ \vdots \\ x_{(n+1)p} \end{bmatrix} - \begin{bmatrix} \varphi_{11} & \ldots & \varphi_{1p} \\ \vdots & \ddots & \\ \varphi_{p1} & \ldots & \varphi_{pp} \end{bmatrix} \begin{bmatrix} x_{n1} \\ \vdots \\ x_{np} \end{bmatrix} \right), \end{aligned} \tag{14}$$

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 the distortion curve $\gamma_n^{(p)}[\varphi]$ from equation (13) 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 $\Lambda_0$ 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 assymptotic effect on the distribution of empirical autocovariances. Let us rewrite the model (14) compactly as

$$e_{n+1}[\varphi^{(p)}] = A_{n+1}^{(p)} - \varphi^{(p)} A_n^{(p)}, \tag{15}$$

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

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

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

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 (15) 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 (15) does not take into account that the basis $B_0 = \{v_1, \ldots, v_p\}$ is unknown and we can use only the estimate $\hat{B}$. According to the formulation of empirical residuals (15) we may define the empirical residual covariances

$$\hat{\Gamma}_{\varphi}^{(p)}(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} \left( e_i[\varphi^{(p)}] - \bar{e}[\varphi^{(p)}] \right) \left( e_{i+j}[\varphi^{(p)}] - \bar{e}[\varphi^{(p)}] \right)' \tag{16}$$

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 (15). We are denoting $\hat{\Gamma}_{\varphi}^{(p)}(j)$ for any $j \geq 1$ as an estimator of $\hat{\Gamma}_{\varphi_0}^{(p)}(j)$ where $\varphi_0$ denotes the projection of a true kernel $\varphi_0 \in \Phi$ to $\varphi_0^{(p)} \in \Phi^{(p)}$.

The null hypothesis may be stated as

$H_0$: The residuals coefficients $e_n[\varphi_0^{(p)}]$ following (15) are not correlated for some $\varphi_0 \in \Phi$,

against the alternative

$H_1 : H_0$ is not true.

The test we use is an extension of the 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) in multivariate setup. By testing lack of autocorrelation of functional residuals using Box-Pierce statistics, (17) and (18) we are testing the specification of a FAR(1) model (3) with the arbitrarily chosen level $p$ of Karhunen-Loéve coefficients.

The test statistic we use is a modification of the multivariate version of Box-Pierce statistic (17), given by Chitturi (1974)

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

or equivalently by Hosking (1981)

$$\hat{Q}_H(\hat{\varphi}^{(p)}) = N \sum_{h=1}^{H} \{\mathrm{vec}(\hat{\Gamma}^{(p)}_{\hat{\varphi}}(h)' [\hat{\Gamma}^{(p)}_{\hat{\varphi}}(0) \otimes \hat{\Gamma}^{(p)}_{\hat{\varphi}}(0)]^{-1} \mathrm{vec}(\hat{\Gamma}^{(p)}_{\hat{\varphi}}(h))\} \tag{18}$$

with $\hat{Q}_H(\varphi_0^{(p)})$ being the statistic $\hat{Q}_H$ evaluated in $\varphi_0^{(p)} \in \Phi^{(p)}$. 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 correlations 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 $h > H$.

# 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 (15) 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 from now on so to simplify notation in some cases we will be writing $\varphi \in \Phi^{(p)}$ instead of $\varphi^{(p)} \in \Phi^{(p)}$. It is known that in a general multivariate autoregressive VARMA$(a, b)$ under $H_0$ the asymptotic distribution of $\hat{Q}_H(\hat{\varphi}^{(p)})$ statistic does not follow a $\chi^2_{\mathrm{p}^2 H}$ 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 $\hat{Q}_H(\varphi_0^{(p)})$. It follows from the fact that our case may be represented as a VAR(1) model. Our approach will be based on defining an operator that would act on estimated $\hat{Q}_H(\hat{\varphi}^{(p)})$ such that estimation error would be eliminated.

Let us define the autocovariance vector of residual curves coordinates in $B_0$ basis, $e_i[\varphi^{(p)}]$ introduced in (16) as

$$\hat{\gamma}_{\varphi}^{(m)} = \left[\mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1))', \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(2))', \ldots, \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m))'\right]', \tag{19}$$

with coorrespondent autocorrelation vector

$$\hat{\rho}_{\varphi}^{(m)} = \left[\mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(1)\hat{\Gamma}_{\varphi}^{(p)}(0)^{-1})', , \ldots, \mathrm{vec}(\hat{\Gamma}_{\varphi}^{(p)}(m)\hat{\Gamma}_{\varphi}^{(p)}(0)^{-1})'\right]', \tag{20}$$

for $\varphi \in \Phi^{(p)}$. It is expected that under $H_0$ the elements of $\hat{\gamma}_{\hat{\varphi}}^{(m)}$ are 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 different from $\chi^2_{(p^2 H)}$ 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 $\hat{\gamma}_{\hat{\varphi}}^{(m)}$,

$$\hat{\zeta}_{\varphi}^{(m)} = \begin{bmatrix} \nabla \text{vec}\hat{\Gamma}_{\varphi}^{(p)}(1) \\ \nabla \text{vec}\hat{\Gamma}_{\varphi}^{(p)}(2) \\ \vdots \\ \nabla \text{vec}\hat{\Gamma}_{\varphi}^{(p)}(m) \end{bmatrix} \tag{21}$$

where

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

The derivatives with respect to the VAR(1) specification of a functional model (3) are going to have a pivotal role in transforming the $\hat{\gamma}_{\varphi}^{(m)}$ vector for $\varphi \in \Phi^{(p)}$ following 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 a linear transformation of the estimated covariance vector $\hat{\gamma}_{\hat{\varphi}}^{(m)}$, defined as (19) that asymptotically does not depend on $\hat{\gamma}_{\varphi_0}^{(m)}$. The reasoning is based on the fact that while we estimate $\hat{\gamma}_{\varphi_0}^{(m)}$, we cannot estimate the asymptotic effect on the asymptotic distribution of $\hat{\gamma}_{\hat{\varphi}}^{(m)}$. Instead of estimating the estimation error, which is impossible we perform the transform using an 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*

$$\hat{\varphi}^{(p)} = \varphi_0^{(p)} + O_p(N^{-\frac{1}{2}}), \tag{22}$$

*the following holds*

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

*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 (22) requires only a $\sqrt{N}$ 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 (22). We present 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 (20) under the same assumptions. It is easy to show that

$$\hat{\rho}_{\hat{\varphi}}^{(m)} = \hat{\rho}_{\varphi_0}^{(m)} + \hat{\eta}_{\varphi_0}^{(m)}(\mathrm{vec}\varphi_0^{(p)} - \mathrm{vec}\hat{\varphi}^{(p)}) + o_p\left(N^{-\frac{1}{2}}\right), \tag{24}$$

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

$$\begin{aligned} \hat{\eta}_{\varphi_0}^{(m)} &= G_{\varphi_0}^{-\frac{1}{2}}\hat{\zeta}_{\varphi_0}^{(m)}, \\ G_{\varphi_0} &= (\Gamma_{\varphi_0}^{(p)})^2 \otimes I_p, \end{aligned} \tag{25}$$

for $\varphi_0 \in \Phi^{(p)}$ with $\hat{\eta}_{\varphi}^{(m)}, \hat{\zeta}_{\varphi}^{(m)}, G_{\varphi}$ counterparts defined in the entire parameter set $\varphi \in \Phi^{(p)}$ with the following $E\hat{G}_{\varphi_0} = G_{\varphi_0}$ and $\hat{\Gamma}_{\varphi_0}^{(p)} = E\Gamma_{\varphi_0}^{(p)}$ holding. Note that it is possible to derive $G_{\varphi_0}$ in a more general way (see Opuchlik 2012). However, if we think about iid error curves then (25) is sufficient. In fact the largest problem are not the weaker assumptions per se but their interpretation and possible consequences in $L^2$ space. Now, following equation (15) and disregarding the distortion errors $\gamma^{(p)}[\varphi](t)$ the expected relationship

$$\sqrt{N}\hat{\rho}_{\varphi_0}^{(m)} \xrightarrow{d} \mathcal{N}(0, I_{mp^2}), \tag{26}$$

holds and above relationship is the basis of Box-Pierce type tests. Note that if we consider the asymptotic distribution of $\hat{\rho}_{\hat{\varphi}}^{(m)}$ instead, then the relationship (26) will not hold according to *Proposition 1.* Now, let us define $\tilde{\rho}_{\hat{\varphi}}^{(m)}$ as a correlation vector evaluated in estimated basis $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$ instead of $B_0$. If we consider the asymptotic distribution of $\tilde{\rho}_{\hat{\varphi}}^{(m)}$ which also takes into account the error of ba-

sis estimation then condition (26) would also not hold. However, we will show that the asymptotic error of $\hat{\rho}_{\hat{\varphi}}^{(m)}$ and $\tilde{\rho}_{\hat{\varphi}}^{(m)}$ is going to be asymptotically the same under **A1-A2**. In different terms we can speak of distortion error as an error originating from substracting least important portion of $L^2$ space. It is an error of representation that applies to representations of eigenfunctions $\{\hat{v}_i(t)\}_{i=1}^{p}$ as well as representations of residual curves $\varepsilon_n[\varphi^{(p)}](t)$ and observables $X_n(t)$. However, we can show that the distortion error does not have the asymptotic effect on the distribution of the BP statistic.

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, Opuchlik, 2012). It is not possible to estimate the set of true autocorrelation errors $\{\hat{\rho}_{\varphi_0}^{(m)}(i)\}_{i=1}^{m}$ 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 the drift equation (24) distribution of $\{\hat{\rho}_{\hat{\varphi}}^{(m)}(i)\}_{i=1}^{m}$ depends asymptotically only on the estimation error and the distribution of $\hat{\rho}_{\varphi_0}^{(m)}$. Let us consider recursive LS estimation of $\beta$ parameter in the following set of equations

$$\left\{\begin{array}{l} \hat{\rho}_{\hat{\varphi}}^{(m)}(j) = \hat{\eta}_{\hat{\varphi}}^{(m)}(j) \times \beta_{j,m} + \mu_{j,j}, \\ \hat{\rho}_{\hat{\varphi}}^{(m)}(j+1) = \hat{\eta}_{\hat{\varphi}}^{(m)}(j+1) \times \beta_{j,m} + \mu_{j,j+1}, \\ \vdots \\ \hat{\rho}_{\hat{\varphi}}^{(m)}(m) = \hat{\eta}_{\hat{\varphi}}^{(m)}(m) \times \beta_{j,m} + \mu_{j,m}, \end{array}\right\} \tag{27}$$

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

Note that the condition $j \leq m-1$ together with $m \geq 2$ is just a necessary condi-

tion, it allows the above system of equations to consist of at least two equations. According to (24) we have that $\hat{\beta}_{j,m}$ is the estimate of $\text{vec}(\varphi_0^{(p)} - \hat{\varphi}^{(p)})$ 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. Note that the second index of $\mu_{j,i}$ is motivated by the fact that the errors are dependent on $\beta_{j,m}$ which is defined by the set of equations (27). Now the solution to (27) for $\beta_{j,m}$ is a feasible recursive LS estimator given by

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

which may be interpreted as a recursive (forward) estimation. According to the above equation and (27) we have to consider dimensions of matrices under consideration. By (20) we have that $\hat{\rho}_{\varphi}^{(m)}$ is the $mp^2 \times 1$ vector in general for any $\varphi \in \Phi^{(p)}$ while $\hat{\rho}_{\varphi}^{(m)}(j)$ is the $p^2 \times 1$ subvector for $0 \leq j \leq m$. Now by (21) and (25) we have that $\hat{\eta}_{\varphi}^{(m)}$ is the $mp^2 \times p^2$ matrix for $\varphi \in \Phi^{(p)}$ and $\hat{\eta}_{\varphi}^{(m)}(j)$ is the $p^2 \times p^2$ submatrix for $j = 1, \ldots, m$. If we are using explicitly $m$ autocorrelations then the condition that has to be met is

$$rank \left( \sum_{i=j+1}^{m} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\eta}_{\varphi_0}^{(m)}(i) \right) = p^2, \tag{28}$$

and denote it as **A3**. In practice we have to use the estimates $\hat{\varphi} \in \Phi^{(p)}$ but if (28) holds then the rank condition evaluated in $\hat{\varphi} \in \Phi^{(p)}$ will also hold with high probability as $N \to \infty$. This problem has not been theoretically studied in this paper, it can be called a problem of numerical identification which was rarely studied using the simulation methods even in classic cases like QML estimation.Note that $\hat{\eta}_{\varphi}^{(m)}(j)$ is a square matrix so the invertibility of $\hat{\eta}_{\hat{\varphi}}^{(m)}(i)' \hat{\eta}_{\hat{\varphi}}^{(m)}(i)$ is going to be satisfied. Because $\hat{\eta}_{\varphi}^{(m)}(j)$ is the matrix of the derivatives of $\text{vec}\hat{\Gamma}_{\varphi}(j)$ evaluated at $\varphi \in \Phi^{(p)}$ which is constructed by kronecker product multiplication. As a result under the null $\hat{\eta}_{\varphi}^{(m)}(j)$ is invertible because rank condition will hold almost sure under the weak assumption that Hilbert-Schmidt operator $\varphi \in \Phi$ does not transform

each of first $p$ eigenfunctions $v_1, v_2, \ldots, v_p$ satisfies $v_i \notin \ker\varphi$. It is interesting to note that **A3** holds always only in the case of FAR(1) model, which is considered in this paper. Now, we may define linear operator $\Im_{\varphi}^{(m)}$,

$$\Im_{\varphi}^{(m)}\{(\hat{\rho}_{\theta}^{(m)}(j))\} = \hat{\rho}_{\theta}^{(m)}(j) - \hat{\eta}_{\varphi}^{(m)}(j)\beta_{j,m}, \tag{29}$$

for $\varphi, \theta \in \Phi^{(p)}$. Note that both parameters $\varphi \in \Phi^{(p)}$ and $\theta \in \Phi^{(p)}$ should be converging in probability to $\varphi_0^{(p)}$. Thus in practice we will have according to (29) that the feasible estimator of $\Im_{\varphi_0}^{(m)}$ would be $\Im_{\hat{\varphi}}^{(m)}$ with the estimate of composite $m$ autocorrelations $\hat{\rho}_{\hat{\varphi}}^{(m)}(i)$ for $i = j, \ldots, m$ ie. satisfying $\hat{\theta} = \hat{\varphi}$. The motivation for (29) is that by *Proposition* 1 we may expect that applying $\Im_{\hat{\varphi}}^{(m)}$ to empirical correlations $\hat{\rho}_{\hat{\varphi}}^{(m)}(j)$ for $j = 1, \ldots, m-1$ would produce the vector $\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))$ satisfying condition (26) under $H_0$. To show that this is the case we formulate the following lemma.

**Lemma.** *Under **A1**-**A3** and* (22) *we have that under* $H_0$ *for* $N \to \infty$

$$\Im_{\varphi_0}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) = \Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j)) + o_P(1),\ j = 1, \ldots, m-1 \tag{30}$$

*and*

$$\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) \xrightarrow{p} \Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j)),\ j = 1, \ldots, m-1. \tag{31}$$

The proof of the lemma has been moved to Appendix. Note that the problem of estimation stated in the lemma considers convergence of the operator $\Im_{\hat{\varphi}}^{(m)}$ and its behaviour in the neighborhood of the true arguments of the form $\hat{\rho}_{\varphi_0}^{(m)}(j)$, $j = 1, \ldots, m-1$ which are not known. 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 (30) 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 to be taken into account in calculating Box-Pierce statistic (18). In our case we have

that $\sqrt{N}\hat{\rho}^{(m)}_{\varphi_0}(j)$ are distributed as independent standard normals for $j \geq 1$. The lemma states that its martingale transform $\sqrt{N}\Im^{(m)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\varphi_0}(j))$ will asymptotically converge to independent mean zero normals. It is followed by $\sqrt{N}\Im^{(m)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))$ being asymptotically distributed as $\sqrt{N}\Im^{(m)}_{\hat{\varphi}}(\rho^{(m)}_{\varphi_0}(j))$. Let us define the asymptotic variance of $\sqrt{N}\Im^{(m)}_{\varphi_0}(\rho^{(m)}_{\varphi_0}(j))$ in following proposition

**Proposition 2.** *$\sqrt{N}\Im^{(m)}_{\varphi_0}(\hat{\rho}^{(m)}_{\varphi_0}(j))$ is asymptotically zero mean normal and its asymptotic variance is defined as*

$$Avar\left(\Im^{(m)}_{\varphi_0}(\hat{\rho}^{(m)}_{\varphi_0}(j))\right) = \Big[I_{p^2} + \eta^{(m)}_{\varphi_0}(j)\Big(\sum_{i=j+1}^{m} \eta^{(m)}_{\varphi_0}(i)'\eta^{(m)}_{\varphi_0}(i)\Big)^{-1}\eta^{(m)}_{\varphi_0}(j)'\Big], \qquad (32)$$

*for $j \geq 1$ with the estimated counterpart $\widehat{Avar}\left(\Im^{(m)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))\right)$ evaluated in $\hat{\varphi} \in \Phi^{(p)}$. In addition the sequence $\{\sqrt{N}\Im^{(m)}_{\varphi_0}(\hat{\rho}^{(m)}_{\varphi_0}(j))\}_{j=1}^{H}$ is asymptotically uncorrelated.*

Now note that in the transformed FAR(1) model by (15) we have according to $\Im^{(m)}_{\varphi}$ defined as in (29) that the first autocorrelation $\hat{\rho}^{(m)}_{\hat{\varphi}}(1)$ is transformed by $m-1$ autocorrelations, $\hat{\rho}^{(m)}_{\hat{\varphi}}(2)$ is transformed by $m-2$ autocorrelations and so forth with $m$'th correlation $\hat{\rho}^{(m)}_{\hat{\varphi}}(m)$ not being transformed by design of (29). We have already argued that rank condition (28) is going to be satisfied with probability one for $N \to \infty$ under the null but in finite case and large dimensionality it could not hold so the use of pseudoinverses is strongly advised in the most demanding cases. In our case by the block construction of derivatives $\hat{\eta}^{(m)}_{\varphi_0}(i)$ we may consider the rank for each matrix instead of considering rank of the sum of squares. However, for example in general VAR($q$) model with $q \geq 2$ rank condition (28) would not hold with probability one under the null in some cases (see Opuchlik et al. 2012). Note that due to non square dimension of derivatives $\hat{\eta}^{(m)}_{\hat{\varphi}}(i)$ in general VARMA case, the rank condition (28) will become more complicated and instead of invertibility we would have to consider the rank of each matrix $\hat{\eta}^{(m)}_{\varphi}(i)$. The rank problem constitutes the most striking feature of univariate VARMA setup and has an interesting interpretation. Because we limit ourselves to the VAR(1) approximation of a functional autoregressive model however, we will not comment

on it further on.

Now, note that the number of autocorrelations used in correction algorithm is not bounded, thus we may use autocorrelation vector $\hat{\rho}_{\hat{\varphi}}^{(m)}$ which contains subvectors $\hat{\rho}_{\hat{\varphi}}^{(m)}(j)$ with $j = 1, \ldots, m$ but at the same time we may also use higher order autocorrelation vectors to define the variant of $\Im_{\hat{\varphi}}^{(m)}$ operator. Let us assume that we want to use arbitrary number $r$ of past autocorrelations in each of LS iterations. This modification of (29) would produce the following projection operator

$$\Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) = \hat{\rho}_{\hat{\varphi}}^{(m)}(j) - \hat{\eta}_{\hat{\varphi}}^{(m)}(j) \left( \sum_{i=j+1}^{j+r} \hat{\eta}_{\hat{\varphi}}^{(m)}(i)' \hat{\eta}_{\hat{\varphi}}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{j+r} \hat{\eta}_{\hat{\varphi}}^{(m)}(i)' \hat{\rho}_{\hat{\varphi}}^{(m)}(i)$$
$$j = 1, \ldots, m-r, \ H < m-r, \ r < m. \tag{33}$$

The reason for fixing the number of derivatives used for recursive LS algorithm in case of FAR(1) model has numerical rather than statistical character. Note that we are aiming at using the $H$ number of autocorrelations in the Box-Pierce statistic (18). If we assume that $m$ is large with $m \geq H$ then using excessive number of derivatives of autocorrelations is going to decrease the goodness of fit of the $\Im_{\hat{\varphi}}^{(m)}$ operator. Thus autocorrelation vectors of small order $j = 1, 2, \ldots$ would be corrected with excessive number of matrices of relatively small norm. Note that given a VAR(1) specification, the autocorrelations and derivatives of autocorrelations of order $j+1, j+2, \ldots$ for any $j \leq H$ are going to converge to zero in expotential way. Thus the estimates of inverses of matrices of the form (28) are not asymptotically bounded which would result in large errors. On the other hand we want to correct the further away autocorrelations close to preset $m$ values with the controlled value $r$ of future autocorrelations. This problem may be treated as a property of resursive projection technique, there is a tradeoff between the precision of the estimate and the number of lags considered $H, m$ and lags used in LS recursive algorithm in each iteration $r$, given the number of observations $N$.

Now accounting for (32) and lack of estimation effect in $\hat{\Gamma}^{(m)}_{\hat{\varphi}}(0)$, we can define according to (18)

$$\hat{Q}^{\Im}_H(\hat{\varphi}^{(p)}) = N\sum_{j=1}^{H}\left\{\Im^{(m,r)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))'\Big[\widehat{Avar}\left(\Im^{(m,r)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))\right)\Big]^{-1}\Im^{(m,r)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))\right\} \tag{34}$$

where $\hat{Q}^{\Im}_H(\hat{\varphi})$ is BP statistic transformed by $\Im^{(m,r)}_{\hat{\varphi}}$ defined in (33) 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*** *as* $N \to \infty$,

$$\widehat{Avar}\left(\Im^{(m,r)}_{\hat{\varphi}}(\hat{\rho}^{(m)}_{\hat{\varphi}}(j))\right) \xrightarrow{p} Avar\left(\Im^{(m,r)}_{\varphi_0}(\hat{\rho}^{(m)}_{\varphi_0}(j))\right), \tag{35}$$

*for* $j = 1, \ldots, H$ *with*

$$\hat{Q}^{\Im}_H(\hat{\varphi}) = \hat{Q}^{\Im}_H(\varphi_0) + o_P(1), \tag{36}$$

$$\hat{Q}^{\Im}_H(\hat{\varphi}) \xrightarrow{D} \chi^2_{(Hp^2)}, \ \textit{with } N \to \infty \tag{37}$$

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

The above result is the main reason that justifies our approach and allows for eliminating the estimation error in the asymptotic distribution of Box-Pierce statistic. The crucial point is that we do not approximate directly the $\hat{\rho}^{(m)}_{\varphi_0}$ vector because we would need the estimate of estimation error. Also the estimation of the asymptotic distribution of estimation error is not necessary. It is clear that $\hat{Q}^{\Im}_H(\hat{\varphi})$ statistic following (18) is a function of a martingale transform of $\hat{\hat{\rho}}^{(m)}_{\hat{\varphi}}$. The orthogonalisation procedure leading to elimination of serial correlation due to estimation error should lead asymptotically to $\chi^2_{(Hp^2)}$ distribution of our transformed BP statistic. Interesting point is that in multivariate setup it would be much harder to rewrite the $\Im^{(m,r)}_{\varphi}$ operator to fit the definition (17). Note that consistency of estimator of $Avar\left(\Im^{(m,r)}_{\varphi_0}(\hat{\rho}^{(m)}_{\varphi_0}(j))\right)$ for $j = 1, \ldots, H$ is a necessary

condition for asymptotic analysis of $\hat{Q}_H^{\Im}(\hat{\varphi})$, which implies (36) and (37).
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 $\nabla \text{vec}\hat{\Gamma}_{\varphi}^{(p)}(j)$, $j = 1, \ldots, m$ for $\varphi \in \Phi^{(p)}$ are obtained using the derivatives of $\hat{e}_n[\varphi]$ for $n = 1, \ldots, N$ with respect to $\varphi'$. In the following proposition we are claiming that under $H_0$ in general strong VAR(1) framework we will obtain the consistency of $\hat{\zeta}_{\hat{\varphi}}^{(m)}$ with the proof moved to the Appendix.

**Proposition 3.** *Under $H_0$ and **A1**-**A3** we have that derivatives matrix defined in* (21) *will satisfy*

$$\begin{aligned} &\nabla vec\hat{\Gamma}_{\hat{\varphi}}^{(p)}(i) \xrightarrow{p} \nabla vec\hat{\Gamma}_{\varphi_0}^{(p)}(i), \quad \text{for } i = 1, \ldots, m \\ &\hat{\zeta}_{\hat{\varphi}}^{(m)} = \hat{\zeta}_{\varphi_0}^{(m)} + o_P(1), \end{aligned} \tag{38}$$

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

$$\hat{\zeta}_{\varphi}^{(m)} = \begin{bmatrix} \nabla vec\hat{\Gamma}_{\varphi}^{(p)}(1) \\ \vdots \\ \nabla vec\hat{\Gamma}_{\varphi}^{(p)}(m) \end{bmatrix} = -\frac{1}{N-m+1} \sum_{n=1}^{N-m+1} \left( \begin{bmatrix} A_n^{(p)\prime} \\ A_{n+1}^{(p)\prime} \\ \vdots \\ A_{n+m-1}^{(p)\prime} \end{bmatrix} \otimes I_p \otimes e_n[\varphi^{(p)}] \right). \tag{39}$$

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**- **A3** would imply stationarity of the derivatives of residual autocorrelations following equation (15). We would also like to know the probability limit of $\hat{\zeta}_{\varphi_0}^{(m)}$. In order to continue we need to state the following

**Proposition 4.** *Assuming* ***A1-A3*** *and* $H_0$ *it is true that*

$$\hat{\zeta}_{\varphi_0}^{(m)} \xrightarrow{p} -E\left(\begin{bmatrix} A_n^{(p)\prime} \\ A_{n+1}^{(p)\prime} \\ \vdots \\ A_{n+m-1}^{(p)\prime} \end{bmatrix} \otimes I_p \otimes e_n[\varphi_0^{(p)}]\right). \tag{40}$$

Note that there are several ways to look at the *Proposition 4*. On one hand it is a technical result allowing to explicitly apply the $\Im_{\hat{\varphi}}^{(m,r)}$ operator because *Proposition 3* gives directly the form of derivatives in multivatiate setup dependent in general on $p$ number of eigenvalues. On the other hand note that if $\hat{\zeta}_{\varphi_0}^{(m)}$ would converge to zero in probability then by *Proposition 1* the asymptotic effect could not be corrected by our algorithm because it would not be possible to distinguish it from the modulus of the order $o_P(N^{-\frac{1}{2}})$. Following *Proposition 1* if the derivative of autocovariances would satisfy $\hat{\zeta}_{\varphi_0}^{(m)} = o_P(1)$ for $\varphi_0 \in \Phi^{(p)}$, then we would have

$$\begin{aligned} &\hat{\zeta}_{\varphi_0}^{(m)}(\mathrm{vec}\varphi_0^{(p)} - \mathrm{vec}\hat{\varphi}^{(p)}) = o_P\left(N^{-\frac{1}{2}}\right), \\ &\hat{\gamma}_{\hat{\varphi}}^{(m)} = \hat{\gamma}_{\varphi_0}^{(m)} + o_P\left(N^{-\frac{1}{2}}\right), \end{aligned} \tag{41}$$

under the assumption of an $\sqrt{N}$ estimator of $\varphi_0^{(p)}$. Thus *Proposition 1* is a necessary condition to consider this type of transform we are using while *Proposition 4* gives a sufficient condition. In the next section we show that the effect of basis estimation does not have asymptotic effect using this argument.

# 4 Asymptotic effect of the estimation of eigenfunction basis and eigenvalue vector decomposition

In the previous section we have 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 the null hypothesis assuming that we know the true basis $B_0$ and the true eigenvalue vector $\Lambda_0$. The further problem however is the distortion

error that comes from the estimation of $B_0$ basis consisting of $v_1, \ldots, v_p$ functions. The assumption we are using is that an $O_p(N^{-\frac{1}{2}})$ estimator of the basis is available. The estimated eigenfunctions will be denoted as $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$ and the corresponding eigenvalues as $\hat{\Lambda} = \text{diag}[\hat{\lambda}_1, \ldots, \hat{\lambda}_p]$. Now in reality we estimate the multivariate representation of FAR(1) according to (15)

$$\tilde{e}_{n+1}[\varphi^{(p)}] = A_{n+1}^{(p)} - \varphi^{(p)} A_n^{(p)} + \xi_{n+1}^{(p)}(p, \varphi^{(p)}), \tag{42}$$

where $\xi_n$ is unobserved distortion error of the empirical residual coefficients for observed $X_n(t)$. We 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 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 the distribution of $\hat{Q}_H[\hat{\varphi}^{(p)}]$ and subsequently on $\hat{Q}_H^{\Im}[\hat{\varphi}^{(p)}]$. In order to show the lack of asymptotic effect of $\xi_n^{(p)}$ estimation we need to refer to *Proposition 1.* The drift equation (23) 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*

$$\hat{B} = B_0 + O_p(N^{-\frac{1}{2}}), \tag{43}$$

$$\hat{\Lambda} = \Lambda_0 + O_p(N^{-\frac{1}{2}}), \tag{44}$$

*the following holds*

$$\tilde{\gamma}_{\varphi_0}^{(m)} = \hat{\gamma}_{\varphi_0}^{(m)} + o_p\left(N^{-\frac{1}{2}}\right) \tag{45}$$

*for $\varphi_0^{(p)} \in \Phi^{(p)}$, where $\tilde{\gamma}_{\varphi_0}^{(m)}$ is residual autovovariance vector of $m$ autocovariances evaluated in $\tilde{e}_{n+1}[\varphi_0^{(p)}]$ and $\hat{\gamma}_{\varphi_0}^{(m)}$ is evaluated with undistorted residual coefficients $e_n[\varphi_0^{(p)}]$.*

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 (43) may be formulated by means of relation between true eigenfunctions $B_0 = [v_1, \ldots, v_p]$ and estimated base $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$. Thus, we may write in relation to equation (42)

$$\tilde{e}_{n+1}[\varphi^{(p)}] = \hat{A}_{n+1}^{(p)} - \varphi^{(p)} \hat{A}_n^{(p)}. \tag{46}$$

for $\varphi^{(p)} \in \Phi^{(p)}$ with $\hat{A}_n^{(p)}$ being approximates of $A_n^{(p)}$ in estimated basis $\hat{B}$.

In order to analyse the asymptotic effect of $\xi_n^{(p)}(p, \varphi^{(p)})$ we need to find the representation of this error in the $B_0$ basis. Note that the distortion errors $\xi_n^{(p)}$ come from the fact that we may write $\hat{v}_i = \sum_{j=1}^{\infty} \hat{\alpha}_{ij} v_j$ with $\hat{\alpha}_{ij}$ not being Kronecker deltas. Thus basis vectors $\hat{v}_i$ are incorrectly specified and the coefficients $\hat{\alpha}_{ij}$ have some distribution around the true values. Given that the estimator $\hat{B}$is not biased we have $E\hat{\alpha}_{ii} = 1$ and $E\hat{\alpha}_{ij} = 0$ for $i \neq j$. We need to give the argument why relating distortion errors $\xi_n^{(p)}$ with $\hat{A}_n^{(p)}$ is sufficient. First we have that basis decomposition estimate will also affect estimation of $\varphi_0^{(p)}$, however it will not change the speed of convergence given that both $\hat{B}$ and $\hat{\varphi}^{(p)}$ estimated in $B_0$ are $\sqrt{N}$ estimators. Secondly $\gamma_n^{(p)}[\varphi]$ truncation error will play no role in testing the lack of correlation of $\tilde{e}_n[\hat{\varphi}^{(p)}]$ because its rationale is testing exactly $p$ most important KL expansion coefficients, which however need to be estimated.
Now in order to show the analog of *Proposition 1* we may use the Fourrier decomposition of functions being elements of $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$ in the true $B_0 = [v_1, \ldots, v_p]$. In some sense we use the first Fourrier coefficients as a measure of deviation of $\hat{B}$ from the true basis $B_0$.

**Proposition 5.** *Defining the matrix* $\hat{\Delta}^{(p)}$ *of coefficients* $\hat{\alpha}_i$ *being a Fourrier ex-*

*pansion of order two of functions* $\{\hat{v}_1, \ldots, \hat{v}_p\}$ *in the* $\{v_1, \ldots, v_p\}$ *basis as*

$$\hat{\Delta}^{(p)} = \begin{bmatrix} \hat{\alpha}_{11} & \hat{\alpha}_{12} & \ldots & \hat{\alpha}_{1p} \\ \hat{\alpha}_{21} & \ddots & & \\ \vdots & & & \\ \hat{\alpha}_{p1} & \ldots & & \hat{\alpha}_{pp} \end{bmatrix}, \tag{47}$$

$$\Delta_0^{(p)} = I_p, \tag{48}$$

*the following conditions hold for any* $X_n(t)$ *series under assumptions* ***A1-A3***

$$\hat{A}_n^{(p)} = \hat{\Delta}^{(p)} A_n^{(p)}, \tag{49}$$

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

$$\tilde{e}_n[\varphi^{(p)}] = \hat{\Delta}^{(p)} A_n^{(p)} - \varphi^{(p)} \hat{\Delta}^{(p)} A_{n-1}^{(p)}, \tag{50}$$

*for any* $n \leq N$*.*

Equation (50) is a direct implication of (49) and (46). Equation (49) is a neat result resembling the scalar case. However, note that in order to obtain equations (49) we have to use the linearity of scalar product and the fact that $L^2(0,1)$ is a Hilbert space so that in a different functional space it would not hold. The persisting question is whether $\hat{\Delta}^{(p)}$ is dependent on $\hat{\Delta}^{(p+1)}$. The answer is negative because coefficients of $\hat{\Delta}^{(p)}$ are unique in a following sense. If we write

$$\hat{\Delta}^{(p)} = \begin{bmatrix} \hat{\alpha}_{11}^{(p)} & \hat{\alpha}_{12}^{(p)} & \ldots & \hat{\alpha}_{1p}^{(p)} \\ \hat{\alpha}_{21}^{(p)} & \ddots & & \\ \vdots & & & \\ \hat{\alpha}_{p1}^{(p)} & \ldots & & \hat{\alpha}_{pp}^{(p)} \end{bmatrix}, \hat{\Delta}^{(p+1)} = \begin{bmatrix} \hat{\alpha}_{11}^{(p+1)} & \hat{\alpha}_{12}^{(p+1)} & \ldots & \hat{\alpha}_{1,p+1}^{(p+1)} \\ \hat{\alpha}_{21}^{(p+1)} & \ddots & & \\ \vdots & & & \\ \hat{\alpha}_{p+1,1}^{(p+1)} & \ldots & & \hat{\alpha}_{p+1,p+1}^{(p+1)} \end{bmatrix}, \tag{51}$$

then we have that $\hat{\alpha}_{ij}^{(p)} = \hat{\alpha}_{ij}^{(p+1)}$ for $i, j \leq p$. It means that considering more $q \geq p$ functions in the $\hat{\Delta}^{(q)}$ matrix would not change coefficients of $\hat{\Delta}^{(p)}$ submatrix. The

reason for this is the fact that we estimate the orthogonal set of functions $v_1, \ldots, v_p$ with $\hat{v}_1, \ldots, \hat{v}_p$ which is also orthogonal, which implies uniqueness of $\hat{\Delta}^{(p)}$.

The next problem we have to adress is the error originating from truncation of $\hat{B}$ representation. In order to show (49) we have to use the following

$$\hat{v}_i = \sum_{j=1}^{p} \hat{\alpha}_{ij} v_j + \varsigma_i^{(p)}, \text{ for } i = 1, \ldots, p,$$

$$\begin{bmatrix} \hat{v}_1 \\ \vdots \\ \hat{v}_p \end{bmatrix} = \hat{\Delta}^{(p)} \begin{bmatrix} v_1 \\ \vdots \\ v_2 \end{bmatrix} + \varsigma^{(p)}.$$

Now, $\varsigma^{(p)}$ is an error of basis representation which suggests the omitted dependence effect. In fact $\varsigma^{(p)}$ may be described as a factor affecting each of distortion errors $\xi_n^{(p)}$ defined in (42). The main point is that the error of representation of observed curves and the errors of estimated basis originating from truncation has the same source. With the estimated basis $\hat{B} = [\hat{v}_1, \ldots, \hat{v}_p]$ we face the error of representation of observed curves $X_n(t)$ due to truncation and we also lose the possibility to ideally represent the true basis $B_0 = [v_1, \ldots, v_p]$ even knowing $\{\hat{\alpha}_{ij}\}_{i,j=1}^{\infty}$. Note that in practice these arguments lose much of its significance, because the errors of representation become negligible very fast with the number $p$ of considered eigenfunctions, however it is interesting to see that both observables and estimated eigenfunctions are elements of the same function space and the errors of representation affect both objects in identical manner.

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 $\hat{\alpha}_{ii}$ deviations $\delta_{ii} = \hat{\alpha}_{ii} - 1$

satisfy under **A1**,

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

with $Z(s,t)$ for $0 \leq s,t \leq 1$ defined as covariance kernel error satisfying $Z(s,t) = K(s,t) - \hat{K}(s,t)$ with

$$K(s,t) = E\{X_n(t) - EX_n(t)\}\{X_n(s) - EX_n(s)\}, \tag{52}$$

$$K(s,t) = \sum_{i=1}^{\infty}\lambda_i v_i(s)v_i(t), \tag{53}$$

and subsequent estimated $\hat{K}$. Covariance operator $K$ is fundamental in estimation of $B_0$ basis and subsequently FAR(1) model while $\lambda_i$ are elements of eigenvalue matrix $\Lambda_0$ associated with $v_1, \ldots, v_p$ eigenfunctions. Again the result concerning the distribution of $\delta_{ii}$ is not dependent on the number $p$ of eigenfunctions considered. The argument is the same as shown in (51) ie. the estimated $\hat{\alpha}_{ij}$ coefficients of $\hat{\Delta}^{(p)}$ are unique. The result of Hall & Hosseini-Nassab is the analogue of first order approximation applied to each trace element of $\hat{\Delta}^{(p)}$ and in this sense is not an asymptotic limit for $p \to \infty$. From our perspecive the most important is the $\sqrt{N}$ convergence $\hat{\Delta}$ estimator.

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 (47) 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 5* we may prove *Theorem 2*. In order to show lack of asymptotic effect of $\hat{\Delta}^{(p)}$ 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 vec$\Delta^{(p)}$ evaluated in $\hat{\Delta}^{(p)}$.

**Proposition 6.** *Under assumptions **A1**-**A2** and $\sqrt{N}$ convergence of $\hat{\Delta}^{(p)}$ we have for $\varphi_0 \in \Phi^{(p)}$,*

$$\frac{\partial}{\partial vec\Delta_0^{(p)\prime}}\tilde{\Gamma}_{\varphi_0}^{(p)}(i) = o_P(1), \tag{54}$$

*for $i \leq m$.*

In order to understand why *Proposition 6* is sufficient condition for lack of asymptotic effect of $\hat{\Delta}^{(p)}$ we may claim the analog of equation (41) that

$$\tilde{\zeta}_{\Delta_0}^{(m)}(\mathrm{vec}\Delta_0^{(p)} - \mathrm{vec}\hat{\Delta}^{(p)}) = o_P\left(N^{-\frac{1}{2}}\right), \tag{55}$$

with $\tilde{\zeta}_{\Delta_0}^{(m)}$ defined as

$$\tilde{\zeta}_{\Delta_0}^{(m)} = \begin{bmatrix} \nabla_\Delta \mathrm{vec}\hat{\Gamma}_{\varphi_0}^{(p)}(1) \\ \nabla_\Delta \mathrm{vec}\hat{\Gamma}_{\varphi_0}^{(p)}(2) \\ \vdots \\ \nabla_\Delta \mathrm{vec}\hat{\Gamma}_{\varphi_0}^{(p)}(m) \end{bmatrix}_{\Delta_0^{(p)}} \tag{56}$$

In the end *Theorem 2* we show that there is a fundamental difference between the effect of $\hat{\varphi}^{(p)} \in \Phi^{(p)}$ and $\hat{\Delta}^{(p)}$ on the distribution of Box-Pierce statistic despite the fact that $\hat{\Delta}^{(p)}$ could be also treated as parameter estimated in some parameter space. The effect of $\hat{\varphi}^{(p)}$ has an asymptotic significance for distribution of empirical autocorrelations of functional residuals $\tilde{e}_n[\hat{\varphi}^{(p)}]$ while $\hat{\Delta}^{(p)}$ asymptotic effect is converging to zero in the $\sqrt{N}$ order of precision, so no special accounting for it is needed.

# 5 Numerical ilustrations

In this section we investigate size and power properties of our modificated 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 $\chi^2(p^2H)$ distribu-

tion. In order to account for estimation error we will be also using $\chi^2(p^2(H-1))$ distribution for uncorrected BP statistic. The key 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 a 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 standard multivariate setup because it starts and ends always at zero.

## 5.1 Empirical size.

Data generating process follows strong FAR(1) model

$$\begin{aligned} X_{n+1}(t^*) &= \int_0^1 \varphi(s,t)X_n(s^*)ds + \varepsilon_{n+1}(t^*), \\ \varphi(s,t) &= 1.2e^{-\frac{(s^2+t^2)}{2}}, \\ \varepsilon_n(t^*) &= \mathrm{BB}_n(t^*). \end{aligned} \tag{57}$$

We simulated 1000 independent trajectories of autoregressive process following (57) of length $N$ with respect to $t^*$ grid. Note that the trace norm of $\varphi$ 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 and then applied portmanteau test to the residuals. The nominal level of the test is $\hat{\alpha} = 5\%$. The number of covariances used in computing $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ BP statistic is $H = 1, \ldots, 8$. In the functional data setup usually BP statistic is derived for small number of lags $H \leq 4$ due to relatively small number of observed functions.

Table 1: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (57) *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 |
| *standard 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 |

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 (57) with $N = 50$ and grid $\#t^* = .05$. The number of eigenfunctions estimated is $p = 2$, which is a 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 $\#t^* = .01$ and increased the number of observations. In this case the results of approximating 5% quantile of $\hat{Q}_{\varphi_0}(H)$ improved for larger values of $H$. For the most important low values of $H$ we have that DV modification of BP test will outperform standard approach. For low values of $H$ the estimation error would

Table 2: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (57) *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 |
| *standard 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 |

affect $\hat{Q}_{\hat{\varphi}}(H)$ in a most significant manner.

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 $\hat{Q}^{\Im}_{\hat{\varphi}}(H)$ 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 standard 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* (57) *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 |
| *standard 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* (57) *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 |
| *standard 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* (57) *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 |
| *standard 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 $\hat{Q}_{\hat{\varphi}}(H)$ we introduce weakly dependent FAR(1) model

$$\begin{aligned} X_{n+1}(t^*) &= \int_0^1 \varphi(s,t) X_n(s^*) ds + \varepsilon_{n+1}(t^*), \\ \varphi(s,t) &= 1.2 e^{-\frac{(s^2+t^2)}{2}}, \\ \varepsilon_n(t^*) &= \mathrm{BB}_n(t^*) \times .\mathrm{BB}_{n-1}(t^*) \end{aligned} \tag{58}$$

Clearly model (58) is a FAR(1) but in this case functional errors $\varepsilon_n(t)$ follow a weakly dependent process. The residuals following (58) are uncorrelated but not independent. Clearly above defined model does not follow **A1-A2** but we applied the method developed in Opuchlik et al. (2012) for weakly dependent VARMA models of computing the corrected BP statistic. 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

Table 6: *Empirical size (in %) of the modified DV of 5% nominal level in the case of the* (58) *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 |
| *standard 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 |

were acting agnostically without knowing the true model.

## 5.2 Empirical power

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

$$X_{n+1}(t^*) = \int_0^1 \phi(s,t) X_n(s^*) ds + \varepsilon_{n+1}(t^*), \tag{59}$$

$$\phi(s,t) = 1.2 e^{-\frac{(s^2+t^2)}{2}}, \tag{60}$$

$$\varepsilon_{n+1}(t^*) = \int_0^1 \theta(s,t) X_{n+1}(s^*) ds + \upsilon_{n+1}(t^*), \tag{61}$$

$$\theta(s,t) = 0.9 e^{-\frac{(s^2+t^2)}{2}}, \tag{62}$$

$$\upsilon_n = BB_n(t^*). \tag{63}$$

As in the size Monte Carlo experiment we simulated 1000 trajectories is of length $N$ with respect to $t^*$ grid. Note that the trace norm of $\varphi$ is the same as in previous experiments $||\varphi||_{\text{tr}} = .85$ and the norm of operator $\theta$ 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

Table 7: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (58) *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 |
| *standard 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 |

and then applied portmanteau test to the residuals. The nominal level of the test is $\hat{\alpha} = 5\%$. The number of covariances used in computing $\hat{Q}^{\Im}_{\hat{\varphi}}(H)$ BP statistic is $H = 1, \ldots, 8$. For DV modification of BP test, the $H_0$ is rejected when $\hat{Q}^{\Im}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$. For the standard BP test we are presenting both results so $H_0$ is rejected when $\hat{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2(H-1))$ or $\hat{Q}_{\hat{\varphi}}(H) > \chi^2_{0.95}(p^2H)$. Clearly model (59) 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 the standard 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 the DV modified 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).

Table 8: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (58) *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 |
| *standard 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 |

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 to $H = 4$. With increasing number of observations both methods tend to reject surely, however the treshold $N$ will be higher for DV modification.

## 5.3 Application to financial markets for minute data

We 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

Table 9: *Empirical power (in %) of the modified DV of 5% nominal level in the case of the* (58) *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 |
| *standard 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 |

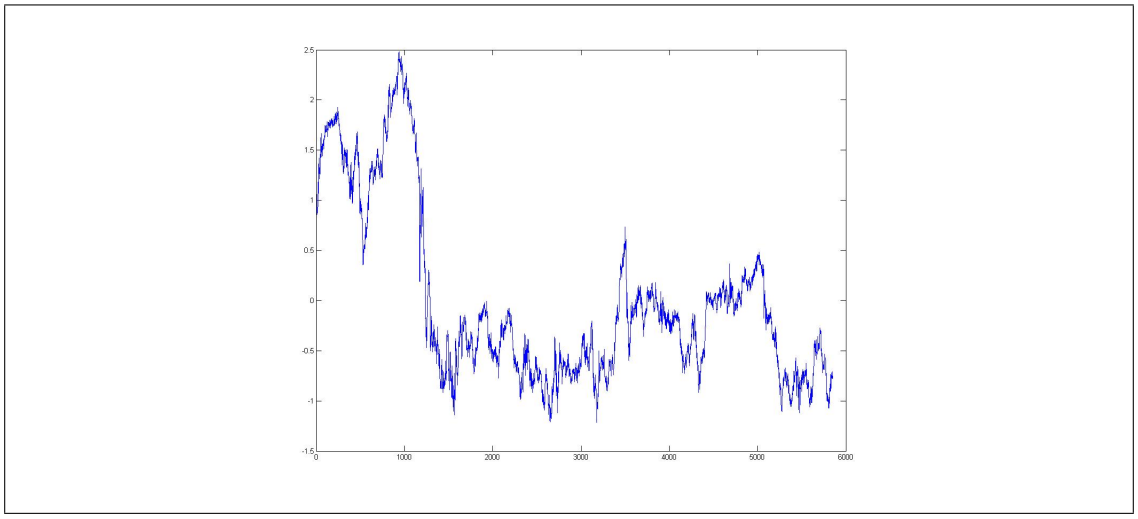


Figure 1: IBM opening prices on NASDAQ with minute frequency 6/05/2006-6/23/2006.

eigenvalues matrix of diagonalized covariance operator $\hat{\Lambda}$ 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 is 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 IBM follow the model

$$\begin{aligned} IBM_{n+1}(t) &= \int_{9.30}^{16.00} \varphi(s,t) IBM_n(s) ds + \varepsilon_{n+1}(t), \\ &\varepsilon_n(t) \text{ are uncorrelated,} \end{aligned} \tag{64}$$

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 *Fig-*

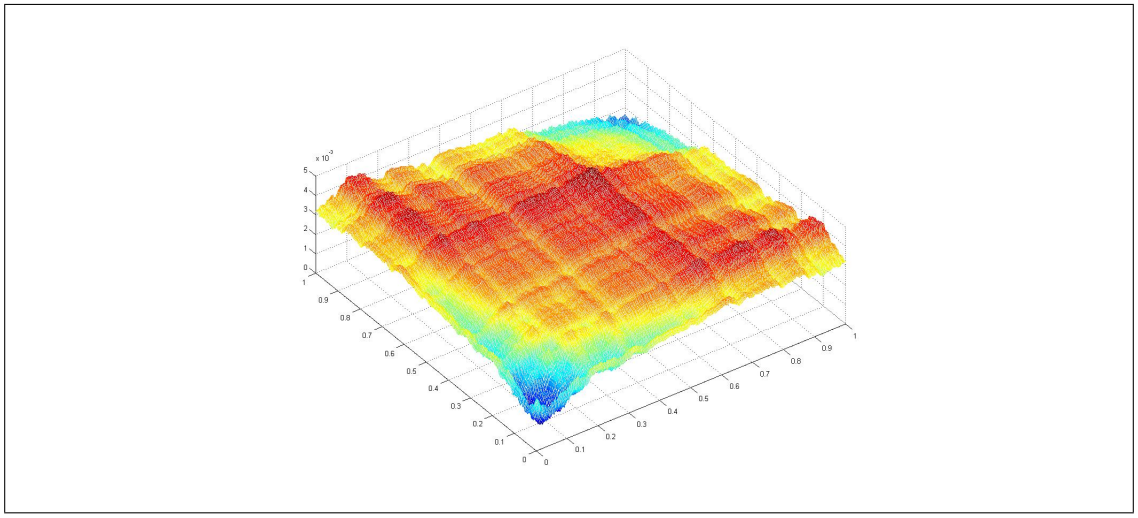

Figure 2: Estimated $\varphi$ kernel of FAR(1) model (64).

*ure 2* is typical representation of the gaussian kernel estimated for low number of observations. If trace norm of $\Psi$ 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

Table 10: *Evaluation of DV modification of BP statistic and standard BP statistic for estimated residuals in model* (64)

| | 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 |
| *standard Box Pierce statistic, Gabrys et al., 2011* | | | | |
| $\tilde{Q}_{\hat{\varphi}}(H)$ | | 3.8336 | 8.4087 | 10.0941 |
| *Critical values for asymptotic $\chi^2$ 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 |

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 $\tilde{Q}^{\Im}_{\hat{\varphi}}(H)$ using the $\chi^2(4H)$ critical value and unmodified BP statistics $\tilde{Q}_{\hat{\varphi}}(H)$ following $\chi^2(4H)$ and $\chi^2(4(H-1))$ critical value. The Type I Error is $\hat{\alpha} = 5\%$.

The results of testing $H_0$ hypothesis were summarized in *Table 10*. Evaluating BP test by both DV modification of the $\tilde{Q}_{\hat{\varphi}}(H)$ statistic and using the standard approach does not suggest 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 an statistical problem. In fact we have used the version of an algorithm of correction developed for general setup of weakly dependent VARMA models.
It may be surprising that eliminating the first day would drastically change the results of estimation. Estimating the noncorrelated functional series leads to sim-

ilar 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*.

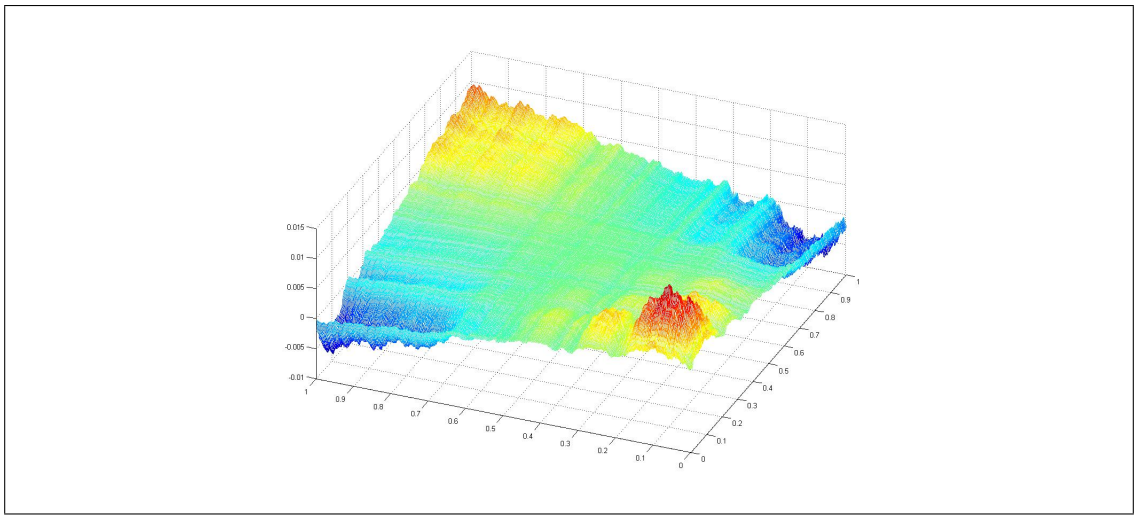


Figure 3: Estimated $\varphi$ kernel of FAR(1) model (64) in the 6/06/2006-6/23/2006 range.

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 (64) 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

$$\Psi(X_n)(t) = \int_0^1 \varphi(s,t)X_n(s)dsdt, \tag{65}$$

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

$$\int_0^1 \int_0^1 |\varphi(s,t)|^2 dsdt < \infty. \tag{66}$$

which implies that we may choose the parameter space $\varphi \in \Phi$, where $\Phi$ is a space of functions that satisfy (66). In general any operator kernel $\varphi$ defines operator $\Psi$ but there exist such compact operators that do not have their kernel representation. Thus by (65) we are assuming that kernel $\varphi$ 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

$$T = \sum_{n=1}^{\infty} \lambda_n \left\langle f_n, \cdot \right\rangle g_n, \tag{67}$$

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

In a general functional data setup we have to choose the most suitable base to represent $X_n(t)$ and $\varphi$ 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 $\Psi$ 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 $\varphi$ which is an element of $L^2((0,1)\times(0,1))$ by (66). 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 $\Psi$ 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

$$\hat{\mathrm{C}}(X, g) = \frac{1}{N}\sum_{n=1}^{N}\left\langle X_n, g \right\rangle X_n, \tag{68}$$

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 $\hat{B}$ is denoted as the estimated matrix of eigenfunctions on $\{t^*\}$, while $\hat{\Lambda}$ is an estimator of the eigenvalues matrix. Now, in general $\Lambda$ 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

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

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

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

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 (14) we have

$$\begin{bmatrix}\hat{x}_{(n+1)1}\\ \vdots \\ \hat{x}_{(n+1)p}\end{bmatrix} = \begin{bmatrix}\hat{\varphi}_{11} & \dots & \hat{\varphi}_{1p}\\ \vdots & \ddots & \\ \hat{\varphi}_{p1} & \dots & \hat{\varphi}_{pp}\end{bmatrix}\begin{bmatrix}\hat{x}_{n1}\\ \vdots \\ \hat{x}_{np}\end{bmatrix} + \begin{bmatrix}\hat{\epsilon}_{(n+1)1}\\ \vdots \\ \hat{\epsilon}_{(n+1)p}\end{bmatrix}, \tag{69}$$

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 $\hat{\varphi}^{(p)}$ is an estimated parameter matrix and $\hat{e}_n$ are estimated residuals equal to $e_n[\hat{\varphi}^{(p)}]$. It is followed by

$$\hat{A}_{n+1}\hat{A}'_n = \hat{\varphi}^{(p)}\hat{A}_n\hat{A}'_n + \hat{e}_{n+1}\hat{A}'_n, \tag{70}$$

$$E\left(\hat{A}_{n+1}\hat{A}'_n\right)E\left(\hat{A}_n\hat{A}'_n\right)^{-1} = E\hat{\varphi}^{(p)}, \tag{71}$$

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

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

Note that instead of using direct inverses we may consider element by element estimates following (70). We have that

$$\begin{bmatrix} \hat{\varphi}_{11} & \dots & \hat{\varphi}_{1p} \\ \vdots & \ddots & \\ \hat{\varphi}_{p1} & \dots & \hat{\varphi}_{pp} \end{bmatrix} = \sum_{i=1}^{N-1} \frac{1}{N-1} \begin{bmatrix} \hat{x}_{(i+1)1}\hat{x}_{i1} & \dots & \hat{x}_{(i+1)1}\hat{x}_{ip} \\ \vdots & \ddots & \\ \hat{x}_{(i+1)p}\hat{x}_{i1} & \dots & \hat{x}_{(i+1)p}\hat{x}_{ip} \end{bmatrix} \times \frac{1}{N-1} \sum_{i=1}^{N-1} \begin{bmatrix} \hat{x}_{i1}^2 & \dots & \hat{x}_{i1}\hat{x}_{ip} \\ \vdots & \ddots & \\ \hat{x}_{ip}\hat{x}_{i1} & \dots & \hat{x}_{ip}^2 \end{bmatrix}^{-1}, \tag{72}$$

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 **A1-A2**. Thus we may write by (72)

$$\hat{\varphi}_{j,i} = \sum_{i=1}^{N-1} \frac{1}{N-1} \hat{x}_{ni}\hat{x}_{nj}(\hat{x}_{ni}^2)^{-1}, \tag{73}$$

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 $\lambda_i$ is an eigenvalue associated to $v_i$. In most cases (73) 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 (72) 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 (69)

$$\begin{bmatrix} \hat{x}_{(n+1)1} \\ \vdots \\ \hat{x}_{(n+1)p} \end{bmatrix} = \left[ \begin{bmatrix} \hat{x}_{n1} \\ \vdots \\ \hat{x}_{np} \end{bmatrix}' \otimes I_p \right] \text{vec}\varphi^{(p)} + \begin{bmatrix} \hat{\epsilon}_{(n+1)1} \\ \vdots \\ \hat{\epsilon}_{(n+1)p} \end{bmatrix}, \tag{74}$$

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 (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 $\varphi^*$ 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

$$D_N(j)_{(i)} = (\text{vec}\hat{\varphi} - \text{vec}\varphi_0)' \frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}(\varphi^*)(\text{vec}\hat{\varphi} - \text{vec}\varphi_0) \tag{75}$$

Now in order to prove that (75) 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 (75) in $\varphi^* \in \Phi^{(p)}$ using *Proposition 3* and **A1-A3**

$$\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}} = \frac{1}{N-j}\sum_{n=1}^{N-j} \frac{\partial}{\partial \text{vec}\varphi} \left( \begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-m}^{(p)} \end{bmatrix}' \otimes I_p \otimes e_n[\varphi_*^{(p)}] \right)'_{i'\text{th row}} = \tag{76}$$

$$= \frac{1}{N-j}\sum_{n=1}^{N-j} \frac{\partial}{\partial \text{vec}\varphi} \left( \begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-m}^{(p)} \end{bmatrix}' \otimes I_p \otimes \left(A_n^{(p)} - \varphi_*^{(p)} A_{n-1}^{(p)}\right)_{i'\text{th row}} \right)'. \tag{77}$$

Now note that for each $n = 1, \ldots, N$ and $j = 1, \ldots, m$ each element of above sum is a finite function of $\{A_{n+j}^{(p)}, \ldots, A_{n+j-m}^{(p)}\}$ or $\{A_{n+j}^{(p)}, \ldots, A_{n-1}^{(p)}\}$ and using iid properties of $\{A_n^{(p)}\}_{n=1}^N$ (from **A1-A2**) by LLN we have

$$\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}} \xrightarrow{p} E\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}}$$

and

$$E\frac{\partial}{\partial \text{vec}\varphi}\left(\begin{bmatrix} A_{n+j}^{(p)} \\ A_{n+j-1}^{(p)} \\ \vdots \\ A_{n+j-m}^{(p)} \end{bmatrix}' \otimes I_p \otimes \left(A_n^{(p)} - \varphi_*^{(p)} A_{n-1}^{(p)}\right)_{i'\text{th row}}\right)' = Ef(\{A_n^{(p)}\}) = O_p(1), \tag{78}$$

because function $f$ is bounded. We do not need to show exactly the form of $f$, for some combinations of $p$ and $j$ it is linear while for others it may be a quadratic form. Now from (78), (76), compactness of $\Phi^{(p)}$ and the fact that $\varphi_0^{(p)}$ and $\varphi_*^{(p)}$ are close to each other, we get the following

$$\left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_*^{(p)}} = \left.\frac{\partial^2 \text{vec}(\hat{\Gamma}_\varphi(j))_{(i)}}{\partial \text{vec}\varphi \partial \text{vec}\varphi'}\right|_{\varphi_0^{(p)}} + o_P(1). \tag{79}$$

Finally from (79) and (75) we get that

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

with

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

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

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

- **Proof of Proposition 2**

From (33) we have

$$\hat{\rho}_{\varphi_0}^{(\Im,m)}(j) = \Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j)) = \hat{\rho}_{\varphi_0}^{(m)}(j) - \hat{\eta}_{\varphi_0}^{(m)}(j) \left( \sum_{i=j+1}^{m} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\eta}_{\varphi_0}^{(m)}(i) \right)^{-1} \sum_{i=j+1}^{m} \hat{\eta}_{\varphi_0}^{(m)}(i)' \hat{\rho}_{\varphi_0}^{(m)}(i), \tag{80}$$

$$j = 1, \ldots, m-r, \ H < m-r, \ r < m. \tag{81}$$

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

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

are not identity matrices so we need a method to standarize it. Secondly we need to argument why there is no dependence between subvectors $\hat{\rho}_{\varphi_0}^{(m)}(i)$ and $\hat{\rho}_{\varphi_0}^{(m)}(j)$ in (80) for $j \leq m-1$ and $i = j+1, \ldots, m$. Lack of dependence follows from independence of error curves assumed in **A1**. Note that the fact that we evaluate the correlations $\hat{\rho}_{\varphi_0}^{(m)}(j)$ in $\varphi_0 \in \Phi^{(p)}$ is essential. Now in order to standarize the vector $\hat{\rho}_{\varphi_0}^{(\Im,m)}(i)$ we may write

$$\sqrt{N}\check{\rho}_{\varphi_0}^{(m)}(j) = \sqrt{N}\hat{\rho}_{\varphi_0}^{(\Im,m)}(j) \times Avar\left(\Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j))\right)^{-1/2},$$
$$\sqrt{N}\check{\rho}_{\varphi_0}^{(m)}(j) \xrightarrow{d} \mathcal{N}(0, I_{p^2}),$$

for j=1,...,m. QED

- **Proof of Proposition 3**

In order to show (38) it is convenient to present the form of derivatives $\hat{\zeta}_{\varphi}^{(m)}$ according to (39) in general for $\varphi^{(p)} \in \Phi^{(p)}$. From (15) we may write the system of equations

$$\nabla_{\varphi} e_{n+j}[\varphi^{(p)}] = \frac{\partial e[\varphi^{(p)}]}{\partial \mathrm{vec}\varphi^{(p)'}}(\varphi^{(p)}) = -{A_{n+j-1}^{(p)}}' \otimes I_p, \tag{82}$$

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

$$\begin{aligned}\nabla \text{vec}\hat{\Gamma}_{\varphi,n}^{(p)}(j) =& \frac{\partial \text{vec}(e_{n+j}[\varphi^{(p)}]e_n[\varphi^{(p)}]')}{\partial \text{vec}\varphi^{(p)\prime}} = \\ =& \frac{\partial e_{n+j}[\varphi^{(p)}]}{\partial \text{vec}\varphi^{(p)\prime}} \otimes e_n[\varphi^{(p)}] + e_{n+j}[\varphi^{(p)}] \otimes \frac{\partial e_n[\varphi^{(p)}]}{\partial \text{vec}\varphi^{(p)\prime}} = \\ =& \nabla_\varphi e_{n+j}[\varphi^{(p)}] \otimes e_n[\varphi^{(p)}] + e_{n+j}[\varphi^{(p)}] \otimes \nabla_\varphi e_n[\varphi^{(p)}],\end{aligned} \tag{83}$$

Now we are proposing an estimator

$$\hat{\zeta}_{\varphi_0}^{(m)} = \begin{bmatrix} \nabla \text{vec}\hat{\Gamma}_\varphi^{(m)}(1) \\ \vdots \\ \nabla \text{vec}\hat{\Gamma}_\varphi^{(m)}(m) \end{bmatrix}_{(\varphi_0)} = \frac{1}{N-m+1} \sum_{n=1}^{N-m+1} \left(\hat{\zeta}_{\varphi_0}^{(m)}\right)_n \tag{84}$$

with

$$\left(\hat{\zeta}_{\varphi_0}^{(m)}\right)_n = \begin{bmatrix} \nabla \text{vec}\hat{\Gamma}_{\varphi,n}^{(p)}(1) \\ \vdots \\ \nabla \text{vec}\hat{\Gamma}_{\varphi,n}^{(p)}(m) \end{bmatrix}_{(\varphi_0)}. \tag{85}$$

Using (84) and (85) gives us (39). Note that the factor

$$\frac{1}{N-m+1} \sum_{n=1}^{N-m+1} e_{n+j}[\varphi^{(p)}] \otimes \nabla_\varphi e_n[\varphi^{(p)}] \xrightarrow{p} E e_{n+j}[\varphi^{(p)}] \otimes \nabla_\varphi e_n[\varphi^{(p)}] = 0$$

under $H_0$ and the probability limit proven in *Proposition 4*. As a result it may be omitted in the implementation of (84).

In order to show (38) we need to prove that $\nabla \text{vec}\hat{\Gamma}_{\varphi_0}^{(m)}(h)$ is estimated consistently for $h = 1, \ldots, m$. The sufficient condition that implies the statement (38) is $\hat{\varphi}^{(p)} \xrightarrow{p} \varphi_0^{(p)}$ and continuous mapping theorem. Note, that if we consider the

function $\tau$,

$$\tau : \varphi^{(p)} \in \Phi^{(p)} \to \hat{\zeta}_{\varphi}^{(m)},$$

then by previous arguments $\tau$ is continuous and by compactness of $\Phi^{(p)}$ with finite dimension of $\Phi^{(p)}$ also bounded.

- **Proof of Proposition 4**

In order to show (40) we may use (39). We have from LLN that

$$-\frac{1}{N}\sum_{n=1}^{N-m+1}\left(\begin{bmatrix} A_n^{(p)\prime} \\ A_{n+1}^{(p)\ \prime} \\ \vdots \\ A_{n+m-1}^{(p)\ \prime} \end{bmatrix} \otimes I_p \otimes e_n[\varphi_0^{(p)}]\right) \xrightarrow{a.s} -E\left[\left(\begin{bmatrix} A_n^{(p)\prime} \\ A_{n+1}^{(p)\ \prime} \\ \vdots \\ A_{n+m-1}^{(p)\ \prime} \end{bmatrix} \otimes I_p \otimes e_n[\varphi_0^{(p)}]\right)\right],$$

for $\varphi_0^{(p)} \in \Phi^{(p)}$, which implies convergence in probability. To see that $E\hat{\zeta}_{\varphi_0}^{(m)} \neq 0$ it is sufficient to note that the vectors $[A_n^{(p)\prime}, \ldots, A_{n+m-1}^{(p)\ \prime}]'$ and $e_n[\varphi_0^{(p)}] = A_n^{(p)} - \varphi_0^{(p)} A_{n-1}^{(p)}$, are correlated.

- **Proof of Lemma**

In order to show (30) it is enough to observe that $\Im_{\varphi_0}^{(m)}$ is continuous and bounded under **A1-A3** 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 (31) it is sufficient to note that $\Im_{\varphi_0}(\cdot) \xrightarrow{p} \Im_{\varphi_0}(\cdot)$ because from *Proposition* 3 and 4 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**

Let us define

$$\hat{Q}_H^{\Im,m}(\hat{\varphi}^{(p)}) = N \sum_{j=1}^{H} \left\{ \Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))' \left[ \widehat{Avar} \left( \Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) \right) \right]^{-1} \Im_{\hat{\varphi}}^{(m,r)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j)) \right\}, \tag{86}$$

which is a transform of standard BP statistic using $\Im_{\hat{\varphi}}^{(m)}$ operator defined in (29). Note that we have that $\Im_{\varphi}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j))$ and $\Im_{\varphi_0}^{(m,r)}(\hat{\rho}_{\varphi_0}^{(m)}(j))$ have the same asymptotic distribution for any $j \leq H$ and $\varphi_0 \in \Phi^{(p)}$ and similar result for $\hat{\varphi}$ follows in the limit. Thus we may consider $\hat{Q}_H^{\Im,m}(\hat{\varphi}^{(p)})$ instead of $\hat{Q}_H^{\Im}(\hat{\varphi}^{(p)})$. The proof now follows from consecutive propositions. *Proposition 1* allows to apply $\Im_{\varphi}^{(m)}(\hat{\rho}_{\varphi}^{(m)}(j))$ operator. By the lemma it is sufficient to consider the distribution of $\sqrt{N}\Im_{\varphi_0}^{(m)}(\hat{\rho}_{\varphi_0}^{(m)}(j))$ instead of $\sqrt{N}\Im_{\hat{\varphi}}^{(m)}(\hat{\rho}_{\hat{\varphi}}^{(m)}(j))$. Condition (35) and (36) hold trivially from the lemma. In order to show (37), from *Proposition 2* we obtain that $\hat{Q}_H^{\Im,m}(\hat{\varphi}^{(p)}) \xrightarrow{D} \chi^2_{(Hp^2)}$ but we argued that $\hat{Q}_H^{\Im}(\hat{\varphi}^{(p)})$ has the same asymptotic distribution.
QED

- **Proof of Proposition 5**

The proof requires bilinearity of a scalar product. We have in general that

$$\left\langle \hat{v}_j, X_n \right\rangle = \left\langle \textstyle\sum_{j=1}^{p} \hat{\alpha}_{ij} v_j, X_n \right\rangle = \displaystyle\sum_{j=1}^{p} \hat{\alpha}_{ij} \left\langle v_j, X_n \right\rangle,$$

for any $0 \leq i \leq p$. Extending this result in univariate case we have

$$\hat{A}_n = \begin{bmatrix} \left\langle \sum_{j=1}^{p} \hat{\alpha}_{1j} v_j, X_n \right\rangle \\ \left\langle \sum_{j=1}^{p} \hat{\alpha}_{2j} v_j, X_n \right\rangle \\ \vdots \\ \left\langle \sum_{j=1}^{p} \hat{\alpha}_{pj} v_j, X_n \right\rangle \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{11} & \hat{\alpha}_{12} & \dots & \hat{\alpha}_{1p} \\ \hat{\alpha}_{21} & \ddots & & \\ \vdots & & & \\ \hat{\alpha}_{p1} & \dots & & \hat{\alpha}_{pp} \end{bmatrix} \begin{bmatrix} \left\langle v_1, X_n \right\rangle \\ \left\langle v_2, X_n \right\rangle \\ \vdots \\ \left\langle v_p, X_n \right\rangle \end{bmatrix} = \hat{\Delta}^{(p)} A_n^{(p)}$$

This proves (49). (50) follows from (49) used in quasi autoregressive equation for $e_n[\varphi^{(p)}]$ residuals (15) for $\varphi^{(p)} \in \Phi^{(p)}$. Note that we may interpret the result as an analogue of scalar case, however in functional spaces Hilbertian assumptions are required.

- **Proof of Proposition 6**

We may start the argument by following the proof of *Proposition 3*. By (50) we have

$$\nabla_{\Delta} \tilde{e}_n^{(p)}[\varphi_0^{(p)}] = \frac{\partial \tilde{e}^{(p)}}{\partial \mathrm{vec} \Delta_0^{(p)\prime}}[\varphi_0^{(p)}] = A_n^{(p)\prime} \otimes I_p - A_{n-1}^{(p)\ \prime} \otimes \varphi_0^{(p)}, \tag{87}$$

for $\varphi_0^{(p)} \in \Phi^{(p)}$. Now, in general for $0 \leq j \leq m$ we have

$$\begin{aligned} \nabla_{\Delta} \mathrm{vec} \tilde{\Gamma}_{\varphi,n}^{(p)}(j) =& \frac{\partial \mathrm{vec}(\tilde{e}_{n+j}^{(p)}[\varphi] \tilde{e}_n^{(p)}[\varphi]')}{\partial \mathrm{vec} \Delta_0^{(p)\prime}} = \\ =& \frac{\partial \tilde{e}_{n+j}^{(p)}[\varphi]}{\partial \mathrm{vec} \Delta_0^{(p)\prime}} \otimes \tilde{e}_n^{(p)}[\varphi] = \nabla_{\Delta} \tilde{e}_{n+j}^{(p)}[\varphi] \otimes \tilde{e}_n^{(p)}[\varphi]. \end{aligned} \tag{88}$$

In order to obtain the derivative of $\tilde{\Gamma}_{\varphi_0}^{(p)}(j)$ with respect to $\mathrm{vec}\Delta'$ in $\varphi_0 \in \Phi^{(p)}$ for $0 \leq j \leq m$ we may use the average

$$\nabla_{\Delta} \mathrm{vec} \tilde{\Gamma}_{\varphi_0}^{(p)}(j) = \frac{1}{N-j-1} \sum_{n=1}^{N-j-1} \nabla_{\Delta} \mathrm{vec} \tilde{\Gamma}_{\varphi_0,n}^{(p)}(j). \tag{89}$$

with following condition satisfied

$$\nabla_\Delta \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j) \xrightarrow{p} E\nabla_\Delta \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j). \tag{90}$$

We need to show that $\nabla_\Delta \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j) = o_p(1)$ and the sufficient condition is lack of correlation between $\tilde{e}^{(p)}_n[\varphi]$ and $\nabla_\Delta \tilde{e}^{(p)}_{n+j}[\varphi_0]$ for any lag $0 \leq j \leq m$. Note that in *Proposition 3* this condition would not hold but here we are using the derivatives from (87). We may derive for $n \leq N$

$$\begin{aligned}
&\nabla_\Delta \tilde{e}^{(p)}_n[\varphi^{(p)}_0]\text{vec}I_p = \\
&= \left(A^{(p)\prime}_n \otimes I_p - A^{(p)\ \prime}_{n-1} \otimes \varphi^{(p)}_0\right)\text{vec}I_p = \\
&= \text{vec}I_p I_p A^{(p)}_n - \text{vec}\varphi^{(p)}_0 I_p A^{(p)}_{n-1} = \\
&= A^{(p)}_n - \varphi^{(p)}_0 A^{(p)}_{n-1} = \\
&= e_n[\varphi^{(p)}_0],
\end{aligned} \tag{91}$$

using equation (87) in the first transition followed by the kronecker product rule $(B' \otimes A)\text{vec}(C) = ACB$. Now, equations (91) lead to the observation that the derivative $\nabla_\Delta \tilde{e}^{(p)}_n[\varphi^{(p)}_0]$ may be lineary transformed into residuals $e_n[\varphi^{(p)}_0]$ evaluated in true parameter matrix $\varphi^{(p)}_0 \in \Phi^{(p)}$ and in $\Delta_0$ ie. without distortion of the basis. Thus if we have any vector noncorrelated with $e_n[\varphi^{(p)}_0]$ it will also not be correlated with $\nabla_\Delta \tilde{e}^{(p)}_n[\varphi^{(p)}_0]$. Note that in scalar case the above holds trivially, while here we may write

$$\nabla_\Delta \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j) = E\nabla_\Delta \tilde{e}^{(p)}_{n+j}[\varphi_0] \otimes \tilde{e}^{(p)}_n[\varphi_0],$$

and by (91) it is sufficient to show that $e_n[\varphi^{(p)}_0]$ and $\tilde{e}_{n-j}[\varphi^{(p)}_0]$ for $0 \leq j \leq m$ are noncorrelated. By **A1-A2** residual vectors $e_n[\varphi^{(p)}_0]$ and $e_{n-j}[\varphi^{(p)}_0]$ are zero mean and independent while distortion error $\xi^{(p)}_n$ sequence is not dependent on past observations of either residual vectors $\{e_i[\varphi^{(p)}_0]\}^{i=n-1}_{-\infty}$ or $\{\xi^{(p)}_i\}^{i=n-1}_{-\infty}$. Now using the probability limit we get

$$E\nabla_\Delta \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j) \xrightarrow{p} 0,$$

for any $0 \leq j \leq m$ so that $\nabla_{\Delta} \text{vec}\tilde{\Gamma}^{(p)}_{\varphi_0}(j) = o_P(1)$.
QED

- **Proof of Theorem 2**

In order to prove statement (45) we will follow the steps from *Proposition 1.* Note that assumption (43) refering to basis $\hat{B}$ and eigenvalues matrix $\hat{\Lambda}$ estimators implies the condition $\Delta_0^{(p)} = \hat{\Delta}^{(p)} + O_P(N^{-\frac{1}{2}})$. Thus the $\sqrt{N}$ estimation of the basis parameters, eigenfunctions and eigenvalues implies $\sqrt{N}$ estimation of the first first order deviations matrix $\Delta_0^{(p)}$. First we need to show that under $H_0$ we have

$$\tilde{\gamma}^{(m)}_{\varphi_0} = \hat{\gamma}^{(m)}_{\varphi_0} + \tilde{\zeta}^{(m)}_{\Delta_0}(\text{vec}\Delta_0^{(p)} - \text{vec}\hat{\Delta}^{(p)}) + o_p\left(N^{-\frac{1}{2}}\right),$$

with $\tilde{\zeta}_{\Delta_0}$ defined as in (56). The proof would be identical with proof of *Proposition 1* but we use the expansion of $\tilde{\gamma}^{(m)}_{\varphi_0}$ around $\hat{\gamma}^{(m)}_{\varphi_0}$ using the approximation of the true basis represented by $\hat{\Delta}^{(p)}$. By *Proposition 5* we could use the estimate of the $\tilde{\zeta}^{(m)}_{\Delta_0}$ following (89) and the probability limit (90) in $\varphi_0^{(p)} \in \Phi^{(p)}$ and the neighbourhood of $\Delta_0^{(p)}$. Now by *Proposition 6* we have that $\tilde{\zeta}^{(m)}_{\Delta_0} = o_P(\left(N^{-\frac{1}{2}}\right))$ and this implies the theorem statement using root $N$ converge in probability of $\hat{\Delta}^{(p)}$.

# 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