Simultaneous Statistical Inference in Dynamic Factor Models

Thorsten Dickhaus[1] and Markus Pauly[2]

Based on the theory of multiple statistical hypotheses testing, we elaborate likelihood-based simultaneous statistical inference methods in dynamic factor models (DFMs). To this end, we work up and extend the methodology of Geweke and Singleton (1981) by proving a multivariate central limit theorem for empirical Fourier transforms of the observable time series. In an asymptotic regime with observation horizon tending to infinity, we employ structural properties of multivariate chi-square distributions in order to construct asymptotic critical regions for a vector of Wald statistics in DFMs, assuming that the model is identified and model restrictions are testable. A model-based bootstrap procedure is proposed for approximating the joint distribution of such a vector for finite sample sizes. Examples of important multiple test problems in DFMs demonstrate the relevance of the proposed methods for practical applications.

MSC 2010 classification numbers: Primary 62J15, 62M10; secondary 62M07, 62F05, 62F03.
JEL Classification: C12, C32, C52.
Key words: Empirical Fourier transform, family-wise error rate, false discovery rate, likelihood ratio statistic, multiple hypothesis testing, multivariate chi-square distribution, resampling, time series regression, Wald statistic.

1. INTRODUCTION AND MOTIVATION

Dynamic factor models are multivariate time series models of the form

$$\mathbf{X}(t) = \sum_{s=-\infty}^{\infty} \Lambda(s)\,\mathbf{f}(t-s) + \varepsilon(t),\ 1 \leq t \leq T. \tag{1.1}$$

Thereby, $\mathbf{X} = (\mathbf{X}(t) : 1 \leq t \leq T)$ denotes a p-dimensional, covariance-stationary stochastic process in discrete time with mean zero, $\mathbf{f}(t) = (f_1(t), \ldots, f_k(t))^\top$ with $k < p$ denotes a k-dimensional vector of so-called "common factors" and $\varepsilon(t) = (\varepsilon_1(t), \ldots, \varepsilon_p(t))^\top$ denotes a p-dimensional vector of "specific factors", to be regarded as error or remainder terms. Both $\mathbf{f}(t)$ and $\varepsilon(t)$ are assumed to be centered and the error terms are modeled as noise in the sense that they are mutually uncorrelated at every time point and, in addition, uncorrelated with $\mathbf{f}(t)$ at all leads and lags. The error terms $\varepsilon(t)$ may, however, exhibit non-trivial (weak) serial autocorrelations. The model dimensions p and k are assumed to be fixed, while the sample size T may tend to infinity.

The underlying interpretation of model (1.1) is that the dynamic behavior of the process $\mathbf{X}$ can already be described well (or completely) by a lower-dimensional "latent" process. The entry (i, j) of the matrix $\Lambda(s)$ quantitatively reflects the influence of the j-th common factor

[1] Thorsten Dickhaus is junior professor (E-mail: dickhaus@math.hu-berlin.de) at Humboldt-University Berlin, Department of Mathematics, Unter den Linden 6, D-10099 Berlin, Germany. Thorsten Dickhaus thanks Reinhard Meister and Jens Stange for fruitful discussions. This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

[2] Markus Pauly is scientific staff member (E-mail: pauly@math.uni-duesseldorf.de) at Heinrich Heine University Düsseldorf, Mathematics department, Universitätsstr. 1, D-40225 Düsseldorf, Germany.

at lead or lag s, respectively, on the i-th component of $\mathbf{X}(t)$, where $1 \leq i \leq p$ and $1 \leq j \leq k$. Recently, Park et al. (2009) and van Bömmel et al. (2013) studied the case where factor loadings may depend on covariates and discussed applications in economics and neuroimaging.

A special case of model (1.1) results if the influence of the common factors on $\mathbf{X}$ is itself without dynamics, i. e., if the model simplifies to

$$\mathbf{X}(t) = \Lambda\, \mathbf{f}(t) + \varepsilon(t),\ 1 \leq t \leq T. \tag{1.2}$$

Peña and Box (1987) were concerned with methods for the determination of the (number of) common factors in a factor model of the form (1.2) and derived a canonical transformation allowing a parsimonious representation of $\mathbf{X}(t)$ in (1.2) in terms of the common factors. Statistical inference in static factor models for longitudinal data has been studied, for instance, by Jöreskog (1969) who developed an algorithm for computing maximum likelihood estimators in models with factorial structure of the covariance matrix of the observables. For further references and developments regarding the theory and the interrelations of different types of (dynamic) factor models we defer the reader to Breitung and Eickmeier (2005), Hallin and Lippi (2013) and references therein.

Statistical inference methods for dynamic factor models typically consider the time series in the frequency domain, cf., among others, Forni et al. (2000, 2009) and references therein, and analyze decompositions of the spectral density matrix of $\mathbf{X}$. Robinson (1991) discussed nonparametric estimators of the latter matrix by kernel smoothing. In a parametric setting, Geweke and Singleton (1981) developed a likelihood-based framework for statistical inference in dynamic factor models by making use of central limit theorems for time series regression in the frequency domain by Hannan (1973). Their inferential considerations rely on the asymptotic normality of the maximum likelihood estimator (MLE) $\hat{\vartheta}$ of the (possibly very high-dimensional) parameter vector ϑ in the frequency-domain representation of the model. We will provide more details in Section 3. To this end, it is essential that the time series model (1.1) is identified in the sense of Geweke and Singleton (1981), which we will assume throughout the paper. If the model is not identified, the individual contributions of the common factors cannot be expressed unambiguously and, consequently, testing for significance or the construction of confidence sets for elements of ϑ is obviously not informative.

In the present work, we will extend the methodology by Geweke and Singleton (1981). Specifically, we will be concerned with simultaneous statistical inference in dynamic factor models under the likelihood framework by considering families of linear hypotheses regarding parameters of the frequency-domain representation of (1.1). As we will demonstrate in Section 3, the following two problems, which are of practical interest, are examples where our methodology applies.

Problem 1 (Which of the specific factors have a non-trivial autocorrelation structure?)**.** *Solving this problem is substantially more informative than just testing a single specific factor for trivial autocorrelations as considered by Geweke and Singleton (1981). Presence of many colored noise components may hint at further hidden common factors and therefore, the solution to Problem 1 can be utilized for the purpose of model diagnosis in the spirit of a residual analysis.*

Problem 2 (Which of the common factors have a lagged influence on $\mathbf{X}$?)**.** *In many economic applications, it is informative if certain factors (such as interventions) have an instantaneous or* *a lagged effect. By solving Problem 2, this can be answered for several of the common factors simultaneously, accounting for the multiplicity of the test problem.*

Solving problems of these types requires multiple testing of several hypotheses simultaneously. In our case, likelihood ratio statistics (or, asymptotically equivalently, Wald statistics) will build the basis for the respective decision rules.

The paper is organized as follows. In Section 2, we provide a brief introduction to multiple testing, especially under positive dependence. In particular, we will analyze structural properties of multivariate chi-square distributions and provide a numerical assessment of type I error control for standard multiple tests when applied to vectors of multivariate chi-square distributed test statistics. This section is meant to contribute to multiple testing theory and practice in general. Although it is known for a longer time that the components of a multivariate chi-square distributed random vector necessarily exhibit pairwise positive correlations, such vectors in general do not fulfill higher-order dependency concepts like multivariate total positivity of order 2 (MTP_2), cf. Example 3.2. in Karlin and Rinott (1980). However, for instance the extremely popular linear step-up test by Benjamini and Hochberg (1995) for control of the false discovery rate (FDR) is only guaranteed to keep the FDR level strictly if the vector of test statistics or p-values, respectively, is MTP_2 (or at least positively regression dependent on subsets, PRDS). Hence, a question of general interest is how this and related tests behave for multivariate chi-square distributed vectors of test statistics. Section 3 demonstrates how such vectors of test statistics arise naturally in connection with likelihood-based solutions to simultaneous inference problems for dynamic factor models of the form (1.1) when the observation horizon T tends to infinity. To this end, we revisit and extend the methodology of Geweke and Singleton (1981). Specifically, we prove a multivariate central limit theorem for empirical Fourier transforms of the observable time series. The asymptotic normality of these Fourier transforms leads to the asymptotic multivariate chi-square distribution of the considered vector of Wald statistics. In Section 4, we propose a model-based resampling scheme for approximating the finite-sample distribution of this vector of test statistics. We conclude with a discussion in Section 5.

2. MULTIPLE TESTING UNDER POSITIVE DEPENDENCE

The general setup of multiple testing theory assumes a statistical model $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ parametrized by $\vartheta \in \Theta$ and is concerned with testing a family $\mathcal{H} = (H_i, i \in I)$ of hypotheses regarding the parameter ϑ with corresponding alternatives $K_i = \Theta \setminus H_i$, where I denotes an arbitrary index set. We identify hypotheses with subsets of the parameter space throughout the paper. Let $\varphi = (\varphi_i, i \in I)$ be a multiple test procedure for $\mathcal{H}$, meaning that each component φ_i, $i \in I$ is a (marginal) test for the test problem H_i versus K_i in the classical sense. Moreover, let $I_0 \equiv I_0(\vartheta) \subseteq I$ denote the index set of true hypotheses in $\mathcal{H}$ and $V(\varphi)$ the number of false rejections (type I errors) of φ, i. e., $V(\varphi) = \sum_{i\in I_0} \varphi_i$. The classical multiple type I error measure in multiple hypothesis testing is the family-wise error rate, FWER for short, and can (for a given $\vartheta \in \Theta$) be expressed as $\mathrm{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta(V(\varphi) > 0)$. The multiple test φ is said to control the FWER at a pre-defined significance level α, if $\sup_{\vartheta\in\Theta} \mathrm{FWER}_\vartheta(\varphi) \leq \alpha$. A simple, but often conservative method for FWER control is based on the union bound and is referred to as Bonferroni correction in the multiple testing literature. Assuming that $|I| = m$, the Bonferroni correction carries out each individual test $\varphi_i, i \in I$, at (local) level α/m. The

"Bonferroni test" $\varphi = (\varphi_i, i \in I)$ then controls the FWER. In case that joint independence of all m marginal test statistics can be assumed, the Bonferroni-corrected level α/m can be enlarged to the "Šidák-corrected" level $1-(1-\alpha)^{1/m} > \alpha/m$ leading to slightly more powerful (marginal) tests. Both the Bonferroni and the Šidák test are single-step procedures, meaning that the same local significance level is used for all m marginal tests.

An interesting other class of multiple test procedures are stepwise rejective tests, in particular step-up-down tests, introduced by Tamhane et al. (1998). They are most conveniently described in terms of p-values $p_1, \ldots, p_m$ corresponding to test statistics $T_1, \ldots, T_m$. It goes beyond the scope of this paper to discuss the notion of p-values in depth. Therefore, we will restrict attention to the case that every individual null hypothesis is simple, the distribution of every T_i, $1 \leq i \leq m$, under H_i is continuous and each T_i tends to larger values under alternatives. The test statistics considered in Section 3 fulfill these requirements, at least asymptotically. Then, we can calculate (observed) p-values by $p_i = 1-F_i(t_i)$, $1 \leq i \leq m$, where F_i is the cumulative distribution function (cdf) of T_i under H_i and t_i denotes the observed value of T_i. The transformation with the upper tail cdf brings all test statistics to a common scale, because each p-value is supported on $[0, 1]$. Small p-values are in favor of the corresponding alternatives.

Definition 1 (Step-up-down test of order λ in terms of p-values, cf. Finner et al., 2012). *Let $p_{1:m} < p_{2:m} < \ldots < p_{m:m}$ denote the ordered p-values for a multiple test problem. For a tuning parameter $\lambda \in \{1, \ldots, m\}$ a step-up-down test $\varphi^\lambda = (\varphi_1, \ldots, \varphi_m)$ (say) of order λ based on some critical values $\alpha_{1:m} \leq \cdots \leq \alpha_{m:m}$ is defined as follows. If $p_{\lambda:m} \leq \alpha_{\lambda:m}$, set $j* = \max\{j \in \{\lambda, \ldots, m\} : p_{i:m} \leq \alpha_{i:m}$ for all $i \in \{\lambda, \ldots, j\}\}$, whereas for $p_{\lambda:m} > \alpha_{\lambda:m}$, put $j* = \sup\{j \in \{1, \ldots, \lambda-1\} : p_{j:m} \leq \alpha_{j:m}\}$ $(\sup \emptyset = -\infty)$. Define $\varphi_i = 1$ if $p_i \leq \alpha_{j*:m}$ and $\varphi_i = 0$ otherwise $(\alpha_{-\infty:m} = -\infty)$.*

A step-up-down test of order $\lambda = 1$ or $\lambda = m$, respectively, is called step-down (SD) or step-up (SU) test, respectively. If all critical values are identical, we obtain a single-step test.

In connection with control of the FWER, SD tests play a pivotal role, because they can often be considered a shortcut of a closed test procedure, cf. Marcus et al. (1976). For example, the famous SD procedure of Holm (1979) employing critical values $\alpha_{i:m} = \alpha/(m-i+1)$, $1 \leq i \leq m$ is, under the assumption of a complete system of hypotheses, a shortcut of the closed Bonferroni test, see, for instance, Sonnemann (2008), and hence controls the FWER at level α.

In order to compare concurring multiple test procedures, also a type II error measure or, equivalently, a notion of power is required under the multiple testing framework. To this end, we define $I_1 \equiv I_1(\vartheta) = I \setminus I_0$, $m_1 = |I_1|$, $S(\varphi) = \sum_{i \in I_1} \varphi_i$ and refer to the expected proportion of correctly detected alternatives, i. e., $\mathrm{power}_\vartheta(\varphi) = \mathbb{E}_\vartheta[S(\varphi)/\max(m_1, 1)]$, as the multiple power of φ under ϑ. If the structure of φ is such that $\varphi_i = \mathbf{1}_{p_i \leq t^*}$ for a common, possibly data-dependent threshold t^*, then the multiple power of φ is increasing in t^*. For step-up-down tests, this entails that index-wise larger critical values lead to higher multiple power.

Gain in multiple power under the constraint of FWER control is only possible if certain structural assumptions for the joint distribution of $(p_1, \ldots, p_m)^\top$ or, equivalently, $(T_1, \ldots, T_m)^\top$ can be established, cf. Example 1 below. In particular, positive dependency among $p_1, \ldots, p_m$ in the sense of MTP_2, see Karlin and Rinott (1980), or PRDS, see Benjamini and Yekutieli (2001), allows for enlarging the critical values $(\alpha_{i:m})_{1 \leq i \leq m}$. To give a specific example, Sarkar (1998) proved that the critical values $\alpha_{i:m} = i\alpha/m$, $1 \leq i \leq m$ can be used as the basis for an FWER-controlling closed test procedure, provided that the joint distribution of p-values is MTP_2. These critical values have originally been proposed by Simes (1986) in connection with a global test for the intersection hypothesis $H_0 = \bigcap_{i=1}^m H_i$ and are therefore often referred to as Simes' critical values. Hommel (1988) worked out a shortcut for the aforementioned closed test procedure based on Simes' critical values; we will refer to this multiple test as φ^{Hommel} in the remainder of this work.

Simes' critical values also play an important role in connection with control of the false discovery rate (FDR). The FDR is a relaxed type I error measure suitable for large systems of hypotheses. Formally, it is defined as $\mathrm{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta[\mathrm{FDP}(\varphi)]$, where $\mathrm{FDP}(\varphi) = V(\varphi)/\max(R(\varphi), 1)$ with $R(\varphi) = V(\varphi) + S(\varphi)$ denoting the total number of rejections of φ under ϑ. The random variable $\mathrm{FDP}(\varphi)$ is called the false discovery proportion. The meanwhile classical linear step-up test by Benjamini and Hochberg (1995), φ^{LSU} (say), is an SU test with Simes' critical values. Under joint independence of all p-values, it provides FDR-control at (exact) level $m_0\alpha/m$, where $m_0 = m - m_1$, see, for instance, Finner et al. (2009). Independently of each other, Benjamini and Yekutieli (2001) and Sarkar (2002) proved that $\sup_{\vartheta\in\Theta} \mathrm{FDR}_\vartheta(\varphi^{\text{LSU}}) \leq m_0\alpha/m$ if the joint distribution of $(p_1, \ldots, p_m)^\top$ is PRDS on I_0 (notice that MTP_2 implies PRDS on any subset). The multiple test φ^{LSU} is the by far most popular multiple test for FDR control and is occasionally even referred to as *the* FDR procedure in the literature.

2.1 MULTIVARIATE CHI-SQUARE DISTRIBUTED TEST STATISTICS

Asymptotically, the vectors of test statistics that are appropriate for testing the hypotheses we are considering in the present work follow a multivariate chi-squared distribution in the sense of the following definition.

Definition 2. *Let $m \geq 2$ and $\vec{\nu} = (\nu_1, \ldots, \nu_m)^\top$ be a vector of positive integers. Let $(Z_{1,1}, \ldots, Z_{1,\nu_1}, Z_{2,1}, \ldots, Z_{2,\nu_2}, \ldots, Z_{m,1}, \ldots, Z_{m,\nu_m})$ denote $\sum_{k=1}^m \nu_k$ jointly normally distributed random variables with joint correlation matrix $R = (\rho(Z_{k_1,\ell_1}, Z_{k_2,\ell_2}) : 1 \leq k_1, k_2 \leq m, 1 \leq \ell_1 \leq \nu_{k_1}, 1 \leq \ell_2 \leq \nu_{k_2})$ such that for any $1 \leq k \leq m$ the random vector $\mathbf{Z}_k = (Z_{k,1}, \ldots, Z_{k,\nu_k})^\top$ has a standard normal distribution on $\mathbb{R}^{\nu_k}$. Let $\mathbf{Q} = (Q_1, \ldots, Q_m)^\top$, where*

$$Q_k = \sum_{\ell=1}^{\nu_k} Z_{k,\ell}^2 \quad \textit{for all} \quad 1 \leq k \leq m. \tag{2.1}$$

Then we call the distribution of $\mathbf{Q}$ a multivariate (central) chi-square distribution (of generalized Wishart-type) with parameters m, $\vec{\nu}$ and R and write $\mathbf{Q} \sim \chi^2(m, \vec{\nu}, R)$.

Well-known special cases arise if all marginal degrees of freedom are identical, i. e., $\nu_1 = \nu_2 = \ldots = \nu_m \equiv \nu$ and the vectors $(Z_{1,1}, \ldots, Z_{m,1})^\top$, $(Z_{1,2}, \ldots, Z_{m,2})^\top$, …, $(Z_{1,\nu}, \ldots, Z_{m,\nu})^\top$ are independent random vectors. If, in addition, the correlation matrices among the m components of these latter ν random vectors are all identical and equal to $\Sigma \in \mathbb{R}^{m\times m}$ (say), then the distribution of $\mathbf{Q}$ is that of the diagonal elements of a Wishart-distributed random matrix $S \sim W_m(\nu, \Sigma)$. This distribution is for instance given in Definition 3.5.7 of the textbook by Timm (2002). The case of potentially different correlation matrices $\Sigma_1, \ldots, \Sigma_\nu$ has been studied by Jensen (1970). Multivariate chi-square distributions play an important role in several multiple testing problems. In Section 3 below, they occur as limiting distributions of vectors of Wald statistics. Further applications comprise statistical genetics (analysis of many contingency tables simultaneously) and multiple tests for Gaussian variances; see for instance Dickhaus and Stange (2012) for more details.

The following lemma shows that among the components of a (generalized) multivariate chi-square distribution only non-negative pairwise correlations can occur.

Lemma 1. *Let* $\mathbf{Q} \sim \chi^2(m, \vec{\nu}, R)$*. Then, for any pair of indices* $1 \leq k_1, k_2 \leq m$ *it holds*

$$0 \leq Cov(Q_{k_1}, Q_{k_2}) \leq 2\sqrt{\nu_{k_1}\, \nu_{k_2}}. \tag{2.2}$$

Proof. Without loss of generality, assume $k_1 = 1$ and $k_2 = 2$. Simple probabilistic calculus now yields

$$\begin{aligned} \text{Cov}(Q_1, Q_2) &= \text{Cov}\left(\sum_{i=1}^{\nu_1} Z_{1,i}^2, \sum_{j=1}^{\nu_2} Z_{2,j}^2\right) \\ &= \sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_2} \text{Cov}(Z_{1,i}^2, Z_{2,j}^2) = 2\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_2} \rho^2(Z_{1,i}, Z_{2,j}) \geq 0. \end{aligned}$$

The upper bound in (2.2) follows directly from the Cauchy-Schwarz inequality, because the variance of a chi-squared distributed random variable with ν degrees of freedom equals 2ν. □

In view of the applicability of multiple test procedures for positively dependent test statistics that have been discussed in Section 2, Lemma 1 points into the right direction. However, as outlined in the introduction, the MTP_2 property for multivariate chi-square or, more generally, multivariate gamma distributions could up to now only be proved for special cases as, for example, exchangeable gamma variates (cf. Example 3.5. in Karlin and Rinott (1980), see also Sarkar and Chang (1997) for applications of this type of multivariate gamma distributions in multiple hypothesis testing). Therefore and especially in view of the immense popularity of φ^{LSU} we conducted an extensive simulation study of FWER and FDR control of multiple tests suitable under MTP_2 (or PRDS) in the case that the vector of test statistics follows a multivariate chi-square distribution in the sense of Definition 2. Specifically, we investigated the shortcut test φ^{Hommel} for control of the FWER and the linear step-up test φ^{LSU} for control of the FDR and considered the following correlation structures among the variates $(Z_{k,\ell^*} : 1 \leq k \leq m)$ for any given $1 \leq \ell^* \leq \max\{\nu_k : 1 \leq k \leq m\}$. (Since only the coefficients of determination enter the correlation structure of the resulting chi-square variates, we restricted our attention to positive correlation coefficients among the $Z_{k,\ell}$.)

1. Autoregressive, AR(1): $\rho_{ij} = \rho^{|i-j|}$, $\rho \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$.
2. Compound symmetry (CS): $\rho_{ij} = \rho + (1-\rho)\mathbf{1}_{\{i=j\}}$, $\rho \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$.
3. Toeplitz: $\rho_{ij} = \rho_{|i-j|+1}$, with $\rho_1 \equiv 1$ and $\rho_2, ..., \rho_{m^*}$ randomly drawn from the interval $[0.1, 0.9]$.
4. Unstructured (UN): The ρ_{ij} are elements of a normalized realization of a Wishart-distributed random matrix with m degrees of freedom and diagonal expectation the elements of which were randomly drawn from $[0.1, 0.9]^m$.

In all four cases, we have $\rho_{ij} = \mathrm{Cov}(Z_{i,\ell^*}, Z_{j,\ell^*})$, $1 \leq i, j \leq m^*$, where $m^* = |\{1 \leq k \leq m : \nu_k \geq \ell^*\}|$. The marginal degrees of freedom $(\nu_k : 1 \leq k \leq m)$ have been drawn randomly from the set $\{1, 2, \ldots, 100\}$ for every simulation setup. In this, we chose decreasing sampling probabilities of the form $\gamma/(\nu + 1)$, $1 \leq \nu \leq 100$, where γ denotes the norming constant, because we were most interested in the small-scale behavior of φ^{Hommel} and φ^{LSU} under dependency. For the number of marginal test statistics, we considered $m \in \{2, 5, 10, 50, 100\}$ and for the number of true hypotheses the respective values of m_0 provided in Tables 1 - 4. For all false hypotheses, we set the corresponding p-values to zero, because the resulting so-called "Dirac-uniform configurations" are assumed to be least favorable for φ^{Hommel} and φ^{LSU}, see, for instance, Finner et al. (2009) and Blanchard et al. (2011). For every simulation setup, we performed $M = 1,000$ Monte Carlo repetitions of the respective multiple test procedures and estimated the FWER or FDR, respectively, by relative frequencies or means, respectively. We present our results in Tables 1 - 4 in the appendix.

Remark 1. *For carrying out these large-scale simulation studies efficiently, we made use of the simulation platform provided by the μTOSS software for multiple hypothesis testing, see Blanchard et al. (2010).*

To summarize our findings, φ^{Hommel} behaved remarkably well over the entire range of simulation setups. Only in a few cases, it violated the target FWER level slightly, but one has to keep in mind that Dirac-uniform configurations correspond to extreme deviations from the null hypotheses which are not expected to be encountered in practical applications. In line with the results by Benjamini and Yekutieli (2001) and Sarkar (2002), φ^{LSU} controlled the FDR well at level $m_0\alpha/m$ (compare with the bound reported at the end of Section 2). One could try to diminish the resulting conservativity for small values of m_0 either by pre-estimating m_0 and plugging the estimated value $\hat{m}_0$ into the nominal level, i. e., replacing α by $m\alpha/\hat{m}_0$, or by employing other sets of critical values. For instance, Finner et al. (2009) and Finner et al. (2012) developed non-linear critical values aiming at full exhaustion of the FDR level for any value of m_0 under Dirac-uniform configurations. However, both strategies are up to now only guaranteed to work well under the assumption of independent p-values and it would need deeper investigations of their validity under positive dependence. Here, we can at least report that we have no indications that φ^{LSU} may not keep the FDR level under our framework, militating in favour of applying this test for FDR control under the framework that we will consider in Section 3.

Example 1 (Communicated to the first author by Klaus Straßburger)**.** *Let us emphasize here that the observed control of FWER and FDR is a specific property of positively dependent test statistics. To give a counterexample, consider $m = 2$ and two normally distributed test statistics T_1 and T_2, where $T_i \sim \mathcal{N}(\mu_i, 1)$, $i = 1, 2$, and $\rho(T_1, T_2) = -1$. Let $H_i : \{\mu_i \leq 0\}$ and, consequently, $K_i : \{\mu_i > 0\}$, $i = 1, 2$, and notice that $T_2 = -T_1$ almost surely under $\mu_1 = \mu_2 = 0$, with corresponding probability measure $\mathbb{P}_{(0,0)}$. A single-step multiple test at local level $\alpha_{loc.}$ for this problem is given by $\varphi = (\varphi_1, \varphi_2)$ with $\varphi_i = \mathbf{1}_{[\Phi^{-1}(1-\alpha_{loc.}),\infty)}(T_i)$, $i = 1, 2$, where Φ denotes the cdf of the standard normal distribution.*

Now, in order to control the FWER at level α with φ, we have to choose $\alpha_{loc.} = \alpha/2$, because

$$\begin{aligned} FWER_{(0,0)}(\varphi) &= \mathbb{P}_{(0,0)}\left(T_1 \geq \Phi^{-1}(1-\alpha_{loc.}) \vee T_2 \geq \Phi^{-1}(1-\alpha_{loc.})\right) \\ &= \mathbb{P}_{(0,0)}\left(T_1 \geq \Phi^{-1}(1-\alpha_{loc.})\right) + \mathbb{P}_{(0,0)}\left(T_1 \leq -\Phi^{-1}(1-\alpha_{loc.})\right) = 2\alpha_{loc.}. \end{aligned}$$

Remark 2. *A different way to tackle the aforementioned problem of lacking higher-order dependency properties is not to rely on the asymptotic* $\mathbf{Q} \sim \chi^2(m, \vec{\nu}, R)$ *(where* R *is unspecified), but to approximate the finite-sample distribution of test statistics, for example by means of appropriate resampling schemes. Resampling-based SD tests for FWER control have been worked out by Troendle (1995) and Romano and Wolf (2005a,b). Resampling-based FDR control can be achieved by applying the methods by Yekutieli and Benjamini (1999), Troendle (2000), or Romano et al. (2008), among others. We will return to resampling-based multiple testing in the context of DFMs in Section 4.*

3. MULTIPLE TESTING IN DYNAMIC FACTOR MODELS

In order to maintain a self-contained presentation, we first briefly summarize some essential techniques and results discussed in previous literature.

Lemma 2. *The spectral density matrix* $S_{\mathbf{X}}$ *(say) of the observable process* $\mathbf{X}$ *can be decomposed as*

$$S_{\mathbf{X}}(\omega) = \tilde{\Lambda}(\omega) S_{\mathbf{f}}(\omega) \tilde{\Lambda}(\omega)' + S_{\varepsilon}(\omega), \quad -\pi \leq \omega \leq \pi, \tag{3.1}$$

where $\tilde{\Lambda}(\omega) = \sum_{s=-\infty}^{\infty} \Lambda(s) \exp(-i\omega s)$ *and the prime stands for transposition and conjugation.*

Proof. The assertion follows immediately by plugging the representation

$$\Gamma_{\mathbf{X}}(u) = \mathbb{E}[\mathbf{X}(t)\mathbf{X}(t+u)^\top] = \sum_{s=-\infty}^{\infty} \Lambda(s) \sum_{v=-\infty}^{\infty} \Gamma_{\mathbf{f}}(u+s-v)\Lambda(v)^\top + \Gamma_{\varepsilon}(u)$$

for the autocovariance function of $\mathbf{X}$ into the formula

$$S_{\mathbf{X}}(\omega) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \Gamma_{\mathbf{X}}(u) \exp(-i\omega u).$$

□

The identifiability conditions mentioned in Section 1 can be plainly phrased by postulating that the representation in (3.1) is unique (up to scaling). All further methods in this section rely on the assumption of an identified model and on asymptotic considerations as $T \to \infty$. To this end, we utilize a localization technique which is due to Hannan (1970, 1973); see also Geweke and Singleton (1981). We consider a scaled version of the empirical (finite) Fourier transform of $\mathbf{X}$. Evaluated at harmonic frequencies, it is given by

$$\tilde{\mathbf{X}}(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^{T} \mathbf{X}(t) \exp(it\omega_j), \text{ where } \omega_j = 2\pi j/T, \ -T/2 < j \leq \lfloor T/2 \rfloor.$$

For asymptotic inference with respect to T, we impose the following additional assumptions.

Assumption 1. *There exist* B *disjoint frequency bands* $\Omega_1, \ldots, \Omega_B$*, such that* $S_{\mathbf{X}}$ *can be assumed approximately constant and different from zero within each of these bands. Let* $\omega^{(b)} \notin \{0, \pi\}$ *denote the center of the band* Ω_b*,* $1 \leq b \leq B$*.*

Notice that Geweke and Singleton (1981) have made an assumption similar to Assumption 1. As in Hannan (1970) and in their work, we will denote by $n_b = n_b(T)$ a number of harmonic frequencies $(\omega_{j,b})_{1\leq j\leq n_b}$ of the form $2\pi j_u/T$ which are as near as possible to $\omega^{(b)}$, $1 \leq b \leq B$. In this, the integers $j_u, 1 \leq u \leq n_b$, in $\omega_{j,b} = 2\pi j_u/T$ are chosen in successive order of closeness to the center. To derive a weak convergence result for $(\tilde{\mathbf{X}}(\omega_{j,b}))_j$ one of the following two additional assumptions, which are due to Hannan (1970, 1973), is needed.

Assumption 2. *The process* $\mathbf{X}$ *is a generalized linear process of the form*

$$\mathbf{X}(t) = \sum_{j=-\infty}^{\infty} A(j)\epsilon(t-j), \tag{3.2}$$

where the process $(\epsilon_t)_t$ *is independent and identically distributed (i.i.d.) white noise and* $A(j) \in \mathbb{R}^{p\times p}$ *fulfills* $\sum_j \|A(j)\|^2 < \infty$.

Assumption 3. *The best linear predictor of* $\mathbf{X}(t)$ *is the best predictor of* $\mathbf{X}(t)$*, both in the least squares sense, given the past of the process.*

Notice that under Assumption 3 we can also represent $\mathbf{X}$ as a linear process of the form

$$\mathbf{X}(t) = \sum_{j=0}^{\infty} A(j)e(t-j), \tag{3.3}$$

where $A(j) \in \mathbb{R}^{p\times p}$ and the process $(e_t)_t$ is uncorrelated white noise, see Hannan (1970). The representations of $\mathbf{X}$ in (3.2) and (3.3) justify the term "white noise factor score model" (WNFS) which has been used, for instance, by Nesselroade et al. (2009).

Throughout the remainder, we denote convergence in distribution by $\xrightarrow{\mathcal{D}}$.

Theorem 1. *Suppose that Assumption 1 and one of the following two conditions hold true:*

(a) Assumption 2 is fufilled.

(b) Assumption 3 holds and the $A(j)$ *of the representation* (3.3) *fulfill*

$$\sum_{j=0}^{\infty} \|A(j)\| < \infty. \tag{3.4}$$

Then we have weak convergence

$$((\tilde{\mathbf{X}}(\omega_{j,b}))_{1\leq j\leq n_b}, 0_{\mathbb{N}}) \xrightarrow{\mathcal{D}} (Z_{j,b})_{j\in\mathbb{N}}, \quad \min(n_b(T), T) \to \infty, \tag{3.5}$$

where the left-hand side of (3.5) *denotes the natural embedding of* $(\tilde{\mathbf{X}}(\omega_{j,b}))_{1\leq j\leq n_b}$ *into* $(\mathbb{R}^p)^{\mathbb{N}}$ *and* $(Z_{j,b})_{j\in\mathbb{N}}$ *is a sequence of independent random vectors, each of which follows a complex normal distribution with mean zero and covariance matrix* $S_{\mathbf{X}}(\omega^{(b)})$.

Proof. Following (Billingsley, 1968, p. 29 f.), it suffices to show convergence of finite-dimensional margins. Recall that the indices $j_u, 1 \leq u \leq n_b$, are chosen in successive order of closeness of $\omega_{j,b} = 2\pi j_u/T$ to the center $\omega^{(b)}$. Hence, under Assumptions 1 and 2, this convergence follows from Theorem 4.13 in Hannan (1970) together with the continuous mapping theorem. In the other case, the convergence in (3.5) is a consequence of Theorem 3 in Hannan (1973), again applied together with the continuous mapping theorem. □

Remark 3.

1. *It is well known that* (3.4) *entails ergodicity of* $\mathbf{X}$.

2. *Actually, Theorem 1 holds under slightly weaker conditions; see Hannan (1973) for details. Moreover, Peligrad and Wu (2010) have recently studied the weak convergence of the finite Fourier transform* $\tilde{\mathbf{X}}$ *under different assumptions.*

3. *While* (3.3) *or* (3.2) *may appear structurally simpler than* (1.1), *notice that the involved coefficient matrices* $A(j)$ *have (potentially much) higher dimensionality than* $\Lambda(s)$ *in* (1.1).

4. *In practice, it seems that the bands* Ω_b *as well as the numbers* n_b *have to be chosen adaptively. To avoid frequencies at the boundary of* Ω_b, *choosing* $n_b = o(T)$ *seems appropriate.*

Let the parameter vector ϑ_b contain all $d = 2pk + k^2 + p$ distinct parameters in $\tilde{\Lambda}(\omega^{(b)})$, $S_{\mathbf{f}}(\omega^{(b)})$ and $S_{\varepsilon}(\omega^{(b)})$, where each of the (in general) complex elements in $\tilde{\Lambda}(\omega^{(b)})$ and $S_{\mathbf{f}}(\omega^{(b)})$ is represented by a pair of real components in ϑ_b, corresponding to its real part and its imaginary part. The full model dimension is consequently equal to Bd. For convenience and in view of Lemma 2, we write with slight abuse of notation $\vartheta_b = \mathrm{vech}(S_{\mathbf{X}}(\omega^{(b)}))$, and $\mathrm{ivech}(\vartheta_b) = S_{\mathbf{X}}(\omega^{(b)})$. The above results motivate to study the (local) likelihood function of the parameter ϑ_b for a given realization $\mathbf{X} = \mathbf{x}$ of the process (from which we calculate $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}$). In frequency band Ω_b, it is given by

$$\ell_b(\vartheta_b, \mathbf{x}) = \pi^{-p \times n_b} \, |\mathrm{ivech}(\vartheta_b)|^{-n_b} \, \exp\left(- \sum_{j=1}^{n_b} \tilde{\mathbf{x}}(\omega_{j,b})' \, \mathrm{ivech}(\vartheta_b)^{-1} \, \tilde{\mathbf{x}}(\omega_{j,b}) \right);$$

see Goodman (1963). Optimization of the B local (log-) likelihood functions requires to solve a system of d non-linear (in the parameters contained in ϑ_b) equations of the form

$$\begin{aligned} 2S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}\tilde{\Lambda}S_{\mathbf{f}} &= 0, \\ 2\tilde{\Lambda}'S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}\tilde{\Lambda} &= 0, \\ diag(S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}) &= 0, \end{aligned}$$

where we dropped the argument $\omega^{(b)}$ in $S_{\mathbf{X}}$, $\tilde{\Lambda}$, and $S_{\mathbf{f}}$, and introduced

$$S = (n_b)^{-1} \sum_{j=1}^{n_b} \tilde{\mathbf{x}}(\omega_{j,b})\tilde{\mathbf{x}}(\omega_{j,b})'.$$

To this end, the algorithm originally developed by Jöreskog (1969) for static factor models can be used (where formally covariance matrices are replaced by spectral density matrices, cf. Geweke and Singleton (1981), and complex numbers are represented by two-dimensional vectors in each optimization step). The algorithm delivers not only the numerical value of the maximum likelihood estimator (MLE) $\hat{\vartheta}_b$, but additionally an estimate of the covariance matrix V_b (say) of $\sqrt{n_b}\hat{\vartheta}_b$. In view of Theorem 1 and standard results from likelihood theory (cf., e. g., Section 12.4 in Lehmann and Romano, 2005) concerning asymptotic normality of MLEs, it appears reasonable to assume that

$$\sqrt{n_b}(\hat{\vartheta}_b - \vartheta_b) \xrightarrow{\mathcal{D}} T_b \sim \mathcal{N}_d(0, V_b), \quad 1 \le b \le B \tag{3.6}$$

as $\min(n_b(T), T) \to \infty$, where the multivariate normal limit random vectors T_b are independent for $1 \le b \le B$, and that $\hat{V}_b$ is a consistent estimator of V_b, which we will assume throughout the remainder. This, in connection with the fact that the vectors $\hat{\vartheta}_b$, $1 \le b \le B$, are asymptotically jointly uncorrelated with each other, is very helpful for testing linear (point) hypotheses. Such hypotheses are of the form $H : C\vartheta = \xi$ with a contrast matrix $C \in \mathbb{R}^{r \times Bd}$, $\xi \in \mathbb{R}^r$ and ϑ consisting of all elements of all the vectors ϑ_b. Geweke and Singleton (1981) proposed the usage of Wald statistics in this context. The Wald statistic for testing H is given by

$$W = N(C\hat{\vartheta} - \xi)^\top (C\hat{V}C^\top)^+ (C\hat{\vartheta} - \xi), \tag{3.7}$$

where $N = \sum_{b=1}^{B} n_b$, $\hat{V}$ is the block matrix built up from the band-specific matrices $N\hat{V}_b/n_b$, $1 \le b \le B$, and A^+ denotes the Moore-Penrose pseudo inverse of a matrix A.

Theorem 2. *Under the above assumptions, W is asymptotically χ^2-distributed with rank(C) degrees of freedom under the null hypothesis H, provided that V is positive definite and $N/n_b \le K < \infty$ for all $1 \le b \le B$.*

Proof. The assertion follows from basic central limit theorems for quadratic forms; see, for example, Theorem 9.2.2 in Rao and Mitra (1971) or Theorem 3.1. in Pauly et al. (2012). □

In the remainder of this section, we return to the two exemplary simultaneous statistical inference problems outlined in Problems 1 and 2 and demonstrate that they can be formalized by families of linear hypotheses regarding (components of) ϑ which in turn can be tested employing the statistical framework that we have considered in Section 2.

Lemma 3 (Problem 1 revisited.)**.** *In the notational framework of Section 2, we have $m = p$, $I = \{1, \ldots, p\}$ and for all $i \in I$ we can consider the linear hypothesis $H_i : C_{Dunnett}\,\mathbf{s}_{\varepsilon_i} = 0$. The contrast matrix $C_{Dunnett}$ is the "multiple comparisons with a control" contrast matrix with $B-1$ rows and B columns, where in each row j the first entry equals $+1$, the $(j+1)$-th entry equals -1 and all other entries are equal to zero. The vector $\mathbf{s}_{\varepsilon_i} \in \mathbb{R}^B$ consists of the values of the spectral density matrix S_ε corresponding to the i-th noise component, evaluated at the B centers $(\omega^{(b)} : 1 \le b \le B)$ of the chosen frequency bins. Denoting the subvector of $\hat{\vartheta}$ that corresponds to $\mathbf{s}_{\varepsilon_i}$ by $\hat{\mathbf{s}}_{\varepsilon_i}$, the i-th Wald statistic is given by*

$$W_i = (C_{Dunnett}\,\hat{\mathbf{s}}_{\varepsilon_i})^\top \left[C_{Dunnett}\hat{V}_{\varepsilon_i}C_{Dunnett}^\top\right]^+ (C_{Dunnett}\,\hat{\mathbf{s}}_{\varepsilon_i}),$$

where $\hat{V}_{\varepsilon_i} = diag(\hat{\sigma}^2_{\varepsilon_i}(\omega^{(b)}) : 1 \le b \le B)$. Then, under H_i, W_i asymptotically follows a χ^2-distribution with $B-1$ degrees of freedom if the corresponding limit matrix V_{ε_i} is assumed to be positive definite. Considering the vector $\mathbf{W} = (W_1, \ldots, W_p)^\top$ of all p Wald statistics corresponding to the p specific factors in the model, we finally have $\mathbf{W} \overset{asympt.}{\sim} \chi^2(p, (B-1, \ldots, B-1)^\top, R)$ under the intersection H_0 of the p hypotheses $H_1, \ldots, H_p$, with some correlation matrix R. This allows to employ the multiple tests considered in Section 2 for solving Problem 1.

Lemma 4 (Problem 2 revisited.)**.** *As done by Geweke and Singleton (1981), we formalize the hypothesis that common factor j has a purely instantaneous effect on $\mathbf{X}_i$, $1 \le j \le k$, $1 \le i \le p$, in the spectral domain by*

$$H_{ij} : |\tilde{\Lambda}_{ij}|^2 \text{ is constant across the } B \text{ frequency bands.}$$

In an analogous manner to the derivations in Lemma 3, the contrast matrix $C_{Dunnett}$ can be used as the basis to construct a Wald statistic W_{ij}. The vector $\mathbf{W} = (W_{ij} : 1 \leq i \leq p, 1 \leq j \leq k)$ then asymptotically follows a multivariate chi-square distribution with $B-1$ degrees of freedom in each marginal under the corresponding null hypotheses and we can proceed as mentioned in Lemma 3.

Many other problems of practical relevance can be formalized analogously by making use of linear contrasts and thus, our framework applies to them, too. Furthermore, the hypotheses of interest may also refer to different subsets of $\{1, \ldots, B\}$. In such a case, the marginal degrees of freedom for the test statistics are not balanced, as considered in the general Definition 2 and in our simulations reported in Section 2.1.

4. FINITE-SAMPLE BOOTSTRAP APPROXIMATON

It is well known that the convergence of Wald-type statistics to their asymptotic χ^2-distribution is rather slow, see Pauly et al. (2012) and references therein. To address this problem and to make use of the actual dependency structure of $\mathbf{W}$ in the multiple test procedure, we propose a model-based bootstrap approximation of the finite-sample distribution of W in (3.7), given by the following algorithm.

1. Given the data $\mathbf{X} = \mathbf{x}$, calculate in each band Ω_b the quantities $\hat{\vartheta}_b$ and $\hat{V}_b$.
2. For all $1 \leq b \leq B$, generate (pseudo) random numbers which behave like realizations of independent random variables $Z^*_{1,b}, \ldots, Z^*_{n_b,b} \overset{i.i.d.}{\sim} \mathcal{N}_d(\hat{\vartheta}_b, \hat{V}_b)$.
3. For all $1 \leq b \leq B$, calculate $\hat{\vartheta}^*_b = n_b^{-1} \sum_{j=1}^{n_b} Z^*_{j,b}$ and $\hat{V}^*_b = n_b^{-1} \sum_{j=1}^{n_b} (Z^*_{j,b} - \hat{\vartheta}^*_b)(Z^*_{j,b} - \hat{\vartheta}^*_b)^\top$.
4. Calculate $W^* = N(\hat{\vartheta}^* - \hat{\vartheta})^\top C^\top (C\hat{V}^* C^\top)^+ C(\hat{\vartheta}^* - \hat{\vartheta})$, where $\hat{\vartheta}^*$ and $\hat{V}^*$ are constructed in analogy to $\hat{\vartheta}$ and $\hat{V}$.
5. Repeat steps 2. - 4. M times to obtain M pseudo replicates of W^* and approximate the distribution of W by the empirical distribution of these pseudo replicates.

The heuristic justification for this algorithm is as follows. Due to Theorem 1 and the complemental Jöreskog algorithm, it is appropriate to approximate the empirical Fourier transforms in the band Ω_b by $Z^*_{1,b}, \ldots, Z^*_{n_b,b}$. Moreover, to capture the structure of W, we build the MLEs $\hat{\vartheta}^*$ and $\hat{V}^*$ of the mean and the covariance matrix, respectively, also in this resampling model. Furthermore, for finite sample sizes it seems more suitable to approximate the distribution of the quadratic form W by a statistic of the same structure. Throughout the remainder, we denote convergence in probability by $\xrightarrow{p}$.

Theorem 3. *Under the assumptions of Theorem 2, it holds*

$$\sup_{w \in \mathbb{R}} |Prob(W^* \leq w|\mathbf{X}) - Prob(W \leq w|H)| \xrightarrow{p} 0, \tag{4.1}$$

where $Prob(W^* \leq \cdot|\mathbf{X})$ denotes the conditional cumulative distribution function (cdf) of W^* given $\mathbf{X}$ and $Prob(W \leq \cdot|H)$ the cdf of W under $H : C\vartheta = \xi$.

Proof. Throughout ρ_k stands for a distance that metrizes weak convergence on $\mathbb{R}^k, k \in \mathbb{N}$, for example the Prohorov distance. Moreover, for a random variable T we denote by $\mathcal{L}(T)$ and $\mathcal{L}(T|\mathbf{X})$ the distribution and conditional distribution of T given $\mathbf{X}$, respectively. Note, that we have by assumption convergences in probability of the conditional mean and variance of $Z^*_{1,b}$, i. e.,

$$\mathrm{E}(Z^*_{1,b}|\mathbf{X}) = \hat{\vartheta}_b \xrightarrow{p} \vartheta_b \quad \text{and} \quad \mathrm{Var}(Z^*_{1,b}|\mathbf{X}) = \hat{V}_b \xrightarrow{p} V_b.$$

Moreover, for each fixed $1 \le b \le B$ and fixed data $\mathbf{X}$, the sequence of random vectors $(Z^*_{j,b})_j$ is row-wise i.i.d. with $\limsup E(\|Z^*_{1,b}\|^4|\mathbf{X}) < \infty$ almost surely. Hence an application of Lyapunov's multivariate Central Limit Theorem together with Slutzky's Theorem implies conditional convergence in distribution given the data $\mathbf{X}$ in the sense that

$$\rho_d\Big(\mathcal{L}(\sqrt{n_b}(\hat{\vartheta}^*_b - \hat{\vartheta}_b)|\mathbf{X}), \mathcal{L}(T_b)\Big) \xrightarrow{p} 0$$

for all $1 \le b \le B$, where $\mathcal{L}(T_b) = \mathcal{N}_d(0, V_b)$. Note that, as usual for resampling mechanisms, the weak convergence originates from the randomness of the bootstrap procedure given $\mathbf{X}$, whereas the convergence in probability arises from the sample $\mathbf{X}$. We can now proceed similarly to the proof of Theorem 3.1. in Pauly et al. (2012). Since the random vectors $\sqrt{n_b}(\hat{\vartheta}^*_b - \hat{\vartheta}_b)$ are also independent within $1 \le b \le B$ given the data, the appearing multivariate normal limit vectors $T_b, 1 \le b \le B$, are independent as well. Together with the continuous mapping theorem this shows that the conditional distribution of $\sqrt{N}C(\hat{\vartheta}^* - \hat{\vartheta})$ given $\mathbf{X}$ converges weakly to a multivariate normal distribution with mean zero and covariance matrix $CVC^\top$ in probability:

$$\rho_r\Big(\mathcal{L}(\sqrt{N}C(\hat{\vartheta}^* - \hat{\vartheta})|\mathbf{X}), \mathcal{N}_r(0, CVC^\top)\Big) \xrightarrow{p} 0.$$

Furthermore, the weak law of large numbers for triangular arrays implies $\hat{V}^*_b - \hat{V}_b \xrightarrow{p} 0$. Since all $V_b, 1 \le b \le B$, are positive definite, we finally have $det(\hat{V}_b) > 0$ almost surely and therefore also $det(\hat{V}^*_b) > 0$ finally almost surely. This, together with the continuous mapping theorem, implies convergence in probability of the Moore-Penrose inverses, i. e.,

$$(C\hat{V}^*C^\top)^+ \xrightarrow{p} (CVC^\top)^+.$$

Thus another application of the continuous mapping theorem together with Theorem 9.2.2 in Rao and Mitra (1971) shows conditional weak convergence of W^* given $\mathbf{X}$ to $\mathcal{L}(W|H)$, the distribution of W under $H: C\vartheta = \xi$, in probability, i.e. $\rho_1(\mathcal{L}(W^*|\mathbf{X}), \mathcal{L}(W|H)) \xrightarrow{p} 0$. The final result is then a consequence of Helly Bray's Theorem and Polya's Uniform Convergence Theorem, since the cdf of W is continuous. $\square$

Remark 4.

1. *Notice that the conditional distribution of W^* always approximates the null distribution of W, even if H does not hold true.*

2. *In view of applications to multiple test problems involving a vector $\mathbf{W} = (W_1, \ldots, W_m)^\top$ as in Problem 1 $(m = p)$ and Problem 2 $(m = pk)$, our resampling approach can be applied as follows. The vector $\mathbf{W}$ can be written as a continuous function g (say) of $C\hat{\vartheta} - \xi$ and $\hat{V}$. Note that the proof of Theorem 3 shows that $C(\hat{\vartheta}^* - \hat{\vartheta})$ always approximates the distribution*

of $C(\hat{\vartheta} - \vartheta)$ and $\hat{V}^* - \hat{V}$ converges to zero in probability. Thus, we can approximate the distribution of $\mathbf{W} = g(C\hat{\vartheta} - \xi, \hat{V})$ under H_0 by $\mathbf{W}^* = g(C(\hat{\vartheta}^* - \hat{\vartheta}), \hat{V}^*)$. Slutzky's Theorem, together with the continuous mapping theorem, ensures that an analogous result to Theorem 3 applies for $\mathbf{W}^*$. This immediately implies that multiple test procedures for weak FWER control can be calibrated by the conditional distribution of $\mathbf{W}^*$. For strong control of the FWER and for FDR control, the resampling approach is valid under the so-called subset pivotality condition (SPC) introduced by Westfall and Young (1993). Validity of the SPC heavily relies on the structure of the function g. For Problems 1 and 2, the SPC is fulfilled, because every W_i depends on mutually different coordinates of $\hat{\vartheta}$.

5. CONCLUDING REMARKS AND OUTLOOK

First of all, we would like to mention that the multiple testing results with respect to FWER control achieved in Sections 2 and 3 also imply (approximate) simultaneous confidence regions for the parameters of model (1.1) by virtue of the extended correspondence theorem, see Section 4.1 of Finner (1994). In such cases (in which focus is on FWER control), a promising alternative method for constructing a multiple test procedure is to deduce the limiting joint distribution of the vector $(Q_1, \ldots, Q_m)^\top$ (say) of likelihood ratio statistics. For instance, one may follow the derivations by Katayama (2008) for the case of likelihood ratio statistics stemming from models with independent and identically distributed observations. Once this limiting joint distribution is obtained, simultaneous test procedures like the ones developed by Hothorn et al. (2008) are applicable.

Second, it may be interesting to assess the variance of the FDP in dynamic factor models, too. Among others, Finner et al. (2007) and Blanchard et al. (2011) have shown that this variance can be large in models with dependent test statistics and have consequently questioned if it is appropriate only to control the first moment of the FDP, because this does not imply a type I error control guarantee for the actual experiment at hand. A maybe more convincing concept in such cases is given by control of the false discovery exceedance, see Farcomeni (2009) for a good survey.

A topic relevant for economic applications is a numerical comparison of the asymptotic multiple tests discussed in Section 2 and the bootstrap-based method derived in Section 4. We will provide such a comparison in a companion paper. Furthermore, one may ask to which extent the results in the present paper can be transferred to more complicated models where factor loadings are modeled as a function of covariates like in Park et al. (2009). To this end, stochastic process techniques way beyond the scope of our setup are required. A first step may be the consideration of parametric models in which conditioning on the design matrix will lead to our framework.

Finally, another relevant multiple test problem in DFMs is to test for cross-sectional correlations between specific factors. While the respective test problems can be formalized by linear contrasts in analogy to Lemmas 3 and 4, they can not straightforwardly be addressed under our likelihood-based framework, because the computation of the MLE by means of the system of normal equations discussed in Section 3 heavily relies on the general assumption of cross-sectionally uncorrelated error terms. Addressing this multiple test problem is therefore devoted to future research.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous referees for their detailed reading and very helpful comments which led to a considerable improvement of the paper.

APPENDIX

Table 1: Simulated FWER control of φ^{Hommel} under AR(1) and compound symmetry structure, respectively. The target FWER level was set to 5% in all simulations.

m	ρ	m_0	$\widehat{\text{FWER}}_{AR(1),\rho}(\varphi^{\text{Hommel}})$	$\widehat{\text{FWER}}_{CS,\rho}(\varphi^{\text{Hommel}})$
2	0.1	1	0.052	0.045
2	0.1	2	0.052	0.057
2	0.25	1	0.06	0.064
2	0.25	2	0.049	0.049
2	0.5	1	0.035	0.056
2	0.5	2	0.055	0.043
2	0.75	1	0.056	0.043
2	0.75	2	0.052	0.049
2	0.9	1	0.051	0.048
2	0.9	2	0.054	0.042
5	0.1	1	0.05	0.053
5	0.1	3	0.047	0.046
5	0.1	5	0.042	0.043
5	0.25	1	0.047	0.031
5	0.25	3	0.057	0.055
5	0.25	5	0.057	0.047
5	0.5	1	0.051	0.043
5	0.5	3	0.052	0.038
5	0.5	5	0.05	0.048
5	0.75	1	0.049	0.054
5	0.75	3	0.055	0.04
5	0.75	5	0.049	0.041
5	0.9	1	0.053	0.045
5	0.9	3	0.043	0.045
5	0.9	5	0.044	0.035
10	0.1	1	0.044	0.054
10	0.1	4	0.06	0.049
10	0.1	7	0.047	0.059
10	0.1	10	0.06	0.057
10	0.25	1	0.048	0.046
10	0.25	4	0.061	0.035

m	ρ	m_0	$\widehat{\text{FWER}}_{AR(1),\rho}(\varphi^{\text{Hommel}})$	$\widehat{\text{FWER}}_{CS,\rho}(\varphi^{\text{Hommel}})$
10	0.25	7	0.056	0.045
10	0.25	10	0.057	0.041
10	0.5	1	0.042	0.053
10	0.5	4	0.047	0.059
10	0.5	7	0.049	0.04
10	0.5	10	0.055	0.062
10	0.75	1	0.048	0.056
10	0.75	4	0.051	0.038
10	0.75	7	0.036	0.049
10	0.75	10	0.031	0.044
10	0.9	1	0.049	0.053
10	0.9	4	0.04	0.038
10	0.9	7	0.041	0.036
10	0.9	10	0.036	0.026
50	0.1	1	0.044	0.061
50	0.1	10	0.036	0.055
50	0.1	25	0.051	0.055
50	0.1	40	0.055	0.043
50	0.1	50	0.042	0.041
50	0.25	1	0.048	0.047
50	0.25	10	0.05	0.062
50	0.25	25	0.03	0.052
50	0.25	40	0.04	0.052
50	0.25	50	0.041	0.052
50	0.5	1	0.047	0.05
50	0.5	10	0.046	0.045
50	0.5	25	0.047	0.058
50	0.5	40	0.047	0.046
50	0.5	50	0.052	0.039
50	0.75	1	0.055	0.055
50	0.75	10	0.055	0.028
50	0.75	25	0.041	0.029
50	0.75	40	0.04	0.044
50	0.75	50	0.039	0.029
50	0.9	1	0.05	0.059
50	0.9	10	0.038	0.03
50	0.9	25	0.037	0.017
50	0.9	40	0.044	0.022
50	0.9	50	0.028	0.024

m	ρ	m_0	$\widehat{\mathrm{FWER}}_{AR(1),\rho}(\varphi^{\mathrm{Hommel}})$	$\widehat{\mathrm{FWER}}_{CS,\rho}(\varphi^{\mathrm{Hommel}})$
100	0.1	1	0.056	0.05
100	0.1	10	0.038	0.055
100	0.1	25	0.046	0.056
100	0.1	50	0.06	0.053
100	0.1	75	0.049	0.047
100	0.1	90	0.06	0.051
100	0.1	100	0.057	0.05
100	0.25	1	0.047	0.057
100	0.25	10	0.055	0.047
100	0.25	25	0.054	0.044
100	0.25	50	0.048	0.045
100	0.25	75	0.041	0.051
100	0.25	90	0.044	0.052
100	0.25	100	0.054	0.044
100	0.5	1	0.047	0.046
100	0.5	10	0.053	0.04
100	0.5	25	0.048	0.04
100	0.5	50	0.056	0.052
100	0.5	75	0.043	0.045
100	0.5	90	0.047	0.033
100	0.5	100	0.042	0.049
100	0.75	1	0.046	0.052
100	0.75	10	0.039	0.039
100	0.75	25	0.044	0.034
100	0.75	50	0.046	0.03
100	0.75	75	0.047	0.024
100	0.75	90	0.048	0.026
100	0.75	100	0.043	0.028
100	0.9	1	0.051	0.05
100	0.9	10	0.045	0.038
100	0.9	25	0.033	0.02
100	0.9	50	0.042	0.008
100	0.9	75	0.046	0.017
100	0.9	90	0.04	0.012
100	0.9	100	0.045	0.016

Table 2: Simulated FWER control of φ^{Hommel} under Toeplitz structure and for unstructured correlation matrices, respectively. The target FWER level was set to 5% in all simulations.

m	m_0	$\widehat{\text{FWER}}_{\text{Toeplitz}}(\varphi^{\text{Hommel}})$	$\widehat{\text{FWER}}_{UN}(\varphi^{\text{Hommel}})$
2	1	0.043	0.052
2	2	0.049	0.052
5	1	0.052	0.057
5	3	0.048	0.041
5	5	0.044	0.037
10	1	0.048	0.05
10	4	0.057	0.04
10	7	0.048	0.046
10	10	0.045	0.043
50	1	0.046	0.043
50	10	0.069	0.043
50	25	0.048	0.044
50	40	0.047	0.036
50	50	0.045	0.054
100	1	0.044	0.047
100	10	0.044	0.054
100	25	0.05	0.048
100	50	0.055	0.054
100	75	0.044	0.055
100	90	0.055	0.038
100	100	0.047	0.055

Table 3: Simulated FDR control of φ^{LSU} under AR(1) and compound symmetry structure, respectively. The target FDR level was set to $\alpha = 5\%$ in all simulations.

m	ρ	m_0	$m_0\alpha/m$	$\widehat{\text{FDR}}_{AR(1),\rho}(\varphi^{\text{LSU}})$	$\widehat{\text{FDR}}_{CS,\rho}(\varphi^{\text{LSU}})$
2	0.1	1	0.025	0.026	0.0225
2	0.1	2	0.05	0.052	0.057
2	0.25	1	0.025	0.03	0.032
2	0.25	2	0.05	0.049	0.049
2	0.5	1	0.025	0.0175	0.028
2	0.5	2	0.05	0.055	0.043
2	0.75	1	0.025	0.028	0.0215
2	0.75	2	0.05	0.052	0.049
2	0.9	1	0.025	0.026	0.024
2	0.9	2	0.05	0.054	0.042
5	0.1	1	0.01	0.01	0.0106
5	0.1	3	0.03	0.028	0.0275
5	0.1	5	0.05	0.043	0.043
5	0.25	1	0.01	0.0094	0.0062
5	0.25	3	0.03	0.033	0.030
5	0.25	5	0.05	0.058	0.05
5	0.5	1	0.01	0.0102	0.0086
5	0.5	3	0.03	0.0308	0.025
5	0.5	5	0.05	0.051	0.049
5	0.75	1	0.01	0.0098	0.0108
5	0.75	3	0.03	0.034	0.030
5	0.75	5	0.05	0.052	0.041
5	0.9	1	0.01	0.0106	0.009
5	0.9	3	0.03	0.0302	0.026
5	0.9	5	0.05	0.048	0.038
10	0.1	1	0.005	0.0044	0.0054
10	0.1	4	0.02	0.0201	0.023
10	0.1	7	0.035	0.032	0.037
10	0.1	10	0.05	0.061	0.058
10	0.25	1	0.005	0.0048	0.0046
10	0.25	4	0.02	0.0201	0.020
10	0.25	7	0.035	0.0375	0.0336
10	0.25	10	0.05	0.057	0.043
10	0.5	1	0.005	0.0042	0.0053
10	0.5	4	0.02	0.022	0.022
10	0.5	7	0.035	0.033	0.029
10	0.5	10	0.05	0.055	0.068

m	ρ	m_0	$m_0\alpha/m$	$\widehat{\mathrm{FDR}}_{AR(1),\rho}(\varphi^{\mathrm{LSU}})$	$\widehat{\mathrm{FDR}}_{CS,\rho}(\varphi^{\mathrm{LSU}})$
10	0.75	1	0.005	0.0048	0.0056
10	0.75	4	0.02	0.021	0.019
10	0.75	7	0.035	0.032	0.038
10	0.75	10	0.05	0.034	0.045
10	0.9	1	0.005	0.0049	0.0053
10	0.9	4	0.02	0.017	0.017
10	0.9	7	0.035	0.035	0.033
10	0.9	10	0.05	0.037	0.03
50	0.1	1	0.001	0.00088	0.00122
50	0.1	10	0.01	0.0093	0.010
50	0.1	25	0.025	0.025	0.025
50	0.1	40	0.04	0.043	0.041
50	0.1	50	0.05	0.042	0.042
50	0.25	1	0.001	0.00096	0.00094
50	0.25	10	0.01	0.0094	0.0099
50	0.25	25	0.025	0.023	0.025
50	0.25	40	0.04	0.037	0.040
50	0.25	50	0.05	0.042	0.053
50	0.5	1	0.001	0.00094	0.001
50	0.5	10	0.01	0.0101	0.010
50	0.5	25	0.025	0.024	0.024
50	0.5	40	0.04	0.042	0.037
50	0.5	50	0.05	0.054	0.04
50	0.75	1	0.001	0.0011	0.0011
50	0.75	10	0.01	0.011	0.0096
50	0.75	25	0.025	0.026	0.021
50	0.75	40	0.04	0.040	0.040
50	0.75	50	0.05	0.04	0.034
50	0.9	1	0.001	0.001	0.0012
50	0.9	10	0.01	0.0097	0.0086
50	0.9	25	0.025	0.024	0.020
50	0.9	40	0.04	0.040	0.039
50	0.9	50	0.05	0.034	0.032
100	0.1	1	0.0005	0.00056	0.00050
100	0.1	10	0.005	0.0045	0.0049
100	0.1	25	0.0125	0.012	0.012
100	0.1	50	0.025	0.026	0.025
100	0.1	75	0.0375	0.037	0.035
100	0.1	90	0.045	0.044	0.046

m	ρ	m_0	$m_0\alpha/m$	$\widehat{\mathrm{FDR}}_{AR(1),\rho}(\varphi^{\mathrm{LSU}})$	$\widehat{\mathrm{FDR}}_{CS,\rho}(\varphi^{\mathrm{LSU}})$
100	0.1	100	0.05	0.058	0.05
100	0.25	1	0.0005	0.00047	0.00057
100	0.25	10	0.005	0.0049	0.0051
100	0.25	25	0.0125	0.013	0.013
100	0.25	50	0.025	0.025	0.026
100	0.25	75	0.0375	0.036	0.038
100	0.25	90	0.045	0.044	0.044
100	0.25	100	0.05	0.055	0.047
100	0.5	1	0.0005	0.00047	0.00046
100	0.5	10	0.005	0.0051	0.0044
100	0.5	25	0.0125	0.013	0.013
100	0.5	50	0.025	0.025	0.027
100	0.5	75	0.0375	0.036	0.038
100	0.5	90	0.045	0.045	0.038
100	0.5	100	0.05	0.045	0.054
100	0.75	1	0.0005	0.00046	0.00052
100	0.75	10	0.005	0.0047	0.0046
100	0.75	25	0.0125	0.012	0.012
100	0.75	50	0.025	0.024	0.023
100	0.75	75	0.0375	0.039	0.034
100	0.75	90	0.045	0.044	0.035
100	0.75	100	0.05	0.044	0.035
100	0.9	1	0.0005	0.00051	0.00050
100	0.9	10	0.005	0.0050	0.0050
100	0.9	25	0.0125	0.012	0.012
100	0.9	50	0.025	0.026	0.020
100	0.9	75	0.0375	0.039	0.033
100	0.9	90	0.045	0.042	0.032
100	0.9	100	0.05	0.048	0.022

Table 4: Simulated FDR control of φ^{LSU} under Toeplitz structure and for unstructured correlation matrices, respectively. The target FDR level was set to $\alpha = 5\%$ in all simulations.

m	m_0	$m_0\alpha/m$	$\widehat{\text{FDR}}_{\text{Toeplitz}}(\varphi^{\text{LSU}})$	$\widehat{\text{FDR}}_{UN}(\varphi^{\text{LSU}})$
2	1	0.025	0.0215	0.026
2	2	0.05	0.049	0.052
5	1	0.01	0.0104	0.011
5	3	0.03	0.034	0.033
5	5	0.05	0.045	0.037
10	1	0.005	0.0048	0.005
10	4	0.02	0.022	0.019
10	7	0.035	0.035	0.033
10	10	0.05	0.046	0.045
50	1	0.001	0.00092	0.00086
50	10	0.01	0.011	0.0096
50	25	0.025	0.025	0.023
50	40	0.04	0.037	0.038
50	50	0.05	0.047	0.057
100	1	0.0005	0.00044	0.00047
100	10	0.005	0.0047	0.0053
100	25	0.0125	0.012	0.012
100	50	0.025	0.025	0.026
100	75	0.0375	0.034	0.037
100	90	0.045	0.044	0.044
100	100	0.05	0.049	0.057

References

Benjamini, Y. and Y. Hochberg (1995): "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *J. R. Stat. Soc., Ser. B*, 57, 289–300.

Benjamini, Y. and D. Yekutieli (2001): "The control of the false discovery rate in multiple testing under dependency." *Ann. Stat.*, 29, 1165–1188.

Billingsley, P. (1968): *Convergence of probability measures.*, New York-London-Sydney-Toronto: John Wiley and Sons, Inc.

Blanchard, G., T. Dickhaus, N. Hack, F. Konietschke, K. Rohmeyer, J. Rosenblatt, M. Scheer, and W. Werft (2010): "μTOSS - Multiple hypothesis testing in an open software system." *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 11, 12–19.

Blanchard, G., T. Dickhaus, E. Roquain, and F. Villers (2011): "On least favorable configurations for step-up-down tests." *In revision, preprint available from arXiv:1108.5262v1.*

Breitung, J. and S. Eickmeier (2005): *Dynamic factor models. Discussion Paper Series 1: Economic Studies 38/2005*, Deutsche Bundesbank.

Dickhaus, T. and J. Stange (2012): "Multiple point hypothesis test problems and effective numbers of tests," SFB 649 Discussion Paper 2012-041, Sonderforschungsbereich 649, Humboldt-Universität zu Berlin, Germany, available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-041.pdf.

Farcomeni, A. (2009): "Generalized augmentation to control false discovery exceedance in multiple testing." *Scand. J. Stat.*, 36, 501–517.

Finner, H. (1994): *Testing Multiple Hypotheses: General Theory, Specific Problems, and Relationships to Other Multiple Decision Procedures*, Habilitationsschrift. Fachbereich IV, Universität Trier.

Finner, H., T. Dickhaus, and M. Roters (2007): "Dependency and false discovery rate: Asymptotics." *Ann. Stat.,*, 35, 1432–1455.

Finner, H., T. Dickhaus, and M. Roters (2009): "On the false discovery rate and an asymptotically optimal rejection curve." *Ann. Stat.*, 37, 596–618.

Finner, H., V. Gontscharuk, and T. Dickhaus (2012): "False Discovery Rate Control of Step-Up-Down Tests with Special Emphasis on the Asymptotically Optimal Rejection Curve." *Scandinavian Journal of Statistics*, 39, 382–397.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic-factor model: Identification and estimation." *The Review of Economics and Statistics*, 82, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2009): "Opening the black box: Structural factor models with large cross sections." *Econometric Theory*, 25, 1319–1347.

Geweke, J. F. and K. J. Singleton (1981): "Maximum likelihood "confirmatory" factor analysis of economic time series." *Int. Econ. Rev.*, 22, 37–54.

Goodman, N. (1963): "Statistical analysis based on a certain multivariate complex Gaussian distribution. (An introduction.)." *Ann. Math. Stat.*, 34, 152–177.

Hallin, M. and M. Lippi (2013): "Factor models in high-dimensional time series - A time-domain approach," *Stochastic Processes and their Applications*, 123, 2678–2695.

Hannan, E. J. (1970): *Multiple time series*, John Wiley and Sons, Inc., New York-London-Sydney.

Hannan, E. J. (1973): "Central limit theorems for time series regression." *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 26, 157–170.

Holm, S. (1979): "A simple sequentially rejective multiple test procedure." *Scand. J. Stat., Theory Appl.*, 6, 65–70.

Hommel, G. (1988): "A stagewise rejective multiple test procedure based on a modified Bonferroni test." *Biometrika*, 75, 383–386.

Hothorn, T., F. Bretz, and P. Westfall (2008): "Simultaneous inference in general parametric models," *Biom. J.*, 50, 346–363.

Jensen, D. R. (1970): "A generalization of the multivariate Rayleigh distribution." *Sankhyā, Series A*, 32, 193–208.

Jöreskog, K. G. (1969): "A general approach to confirmatory maximum likelihood factor analysis." *Psychometrika*, 34, 183–202.

Karlin, S. and Y. Rinott (1980): "Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions." *J. Multivariate Anal.*, 10, 467–498.

Katayama, N. (2008): *Portmanteau Likelihood Ratio Tests for Morel Selection. Discussion Paper Series 2008-1*, Faculty of Economics, Kyushu University.

Lehmann, E. L. and J. P. Romano (2005): *Testing statistical hypotheses. 3rd ed.*, Springer Texts in Statistics. New York, NY: Springer.

Marcus, R., E. Peritz, and K. Gabriel (1976): "On closed testing procedures with special reference to ordered analysis of variance." *Biometrika*, 63, 655–660.

Nesselroade, J. R., J. J. McArdle, S. H. Aggen, and J. M. Meyers (2009): "Dynamic Factor Analysis Models for Representing Process in Multivariate Time-Series," D.S. Moskowitz, Scott L. Hershberger (ed.), Modeling Intraindividual Variability With Repeated Measures Data:Methods and Applications. Lawrence Erlbaum Associates, Inc., Chapter 9.

Park, B. U., E. Mammen, W. Härdle, and S. Borak (2009): "Time series modelling with semiparametric factor dynamics." *J. Am. Stat. Assoc.*, 104, 284–298.

Pauly, M., E. Brunner, and F. Konietschke (2012): "Asymptotic Permutation Tests in General Factorial Designs." Technical report, Heinrich Heine University Düsseldorf, available at http://www.math.uni-duesseldorf.de/˜stoch/Pauly/TechReport.pdf.

Peligrad, M. and W. B. Wu (2010): "Central limit theorem for Fourier transforms of stationary processes." *Ann. Probab.*, 38, 2009–2022.

Peña, D. and G. E. Box (1987): "Identifying a simplifying structure in time series." *J. Am. Stat. Assoc.*, 82, 836–843.

Rao, C. R. and S. K. Mitra (1971): *Generalized inverse of matrices and its applications*, John Wiley & Sons, Inc., New York-London-Sydney.

Robinson, P. (1991): "Automatic frequency domain inference on semiparametric and nonparametric models." *Econometrica*, 59, 1329–1363.

Romano, J. P., A. M. Shaikh, and M. Wolf (2008): "Control of the false discovery rate under dependence using the bootstrap and subsampling." *Test*, 17, 417–442.

Romano, J. P. and M. Wolf (2005a): "Exact and approximate stepdown methods for multiple hypothesis testing." *J. Am. Stat. Assoc.*, 100, 94–108.

Romano, J. P. and M. Wolf (2005b): “Stepwise multiple testing as formalized data snooping.” *Econometrica*, 73, 1237–1282.

Sarkar, S. K. (1998): “Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture.” *Ann. Stat.*, 26, 494–504.

Sarkar, S. K. (2002): “Some results on false discovery rate in stepwise multiple testing procedures.” *Ann. Stat.*, 30, 239–257.

Sarkar, S. K. and C.-K. Chang (1997): “The Simes method for multiple hypothesis testing with positively dependent test statistics.” *J. Am. Stat. Assoc.*, 92, 1601–1608.

Simes, R. (1986): “An improved Bonferroni procedure for multiple tests of significance.” *Biometrika*, 73, 751–754.

Sonnemann, E. (2008): “General solutions to multiple testing problems. Translation of ”Sonnemann, E. (1982). Allgemeine Lösungen multipler Testprobleme. EDV in Medizin und Biologie 13(4), 120-128”,” *Biom. J.*, 50, 641–656.

Tamhane, A. C., W. Liu, and C. W. Dunnett (1998): “A generalized step-up-down multiple test procedure.” *Can. J. Stat.*, 26, 353–363.

Timm, N. H. (2002): *Applied multivariate analysis.*, New York, NY: Springer.

Troendle, J. F. (1995): “A stepwise resampling method of multiple hypothesis testing.” *J. Am. Stat. Assoc.*, 90, 370–378.

Troendle, J. F. (2000): “Stepwise normal theory multiple test procedures controlling the false discovery rate.” *J. Stat. Plann. Inference*, 84, 139–158.

van Bömmel, A., S. Song, P. Majer, P. N. C. Mohr, H. R. Heekeren, and W. K. Härdle (2013): “Risk Patterns and Correlated Brain Activities. Multidimensional statistical analysis of fMRI data in economic decision making study,” *Psychometrika*, forthcoming.

Westfall, P. H. and S. S. Young (1993): *Resampling-based multiple testing: examples and methods for p-value adjustment.*, Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley, New York.

Yekutieli, D. and Y. Benjamini (1999): “Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics.” *J. Stat. Plann. Inference*, 82, 171–196.