SUPPLEMENT TO THE PAPER "TESTING EQUALITY OF SPECTRAL DENSITIES USING RANDOMIZATION TECHNIQUES"

CARSTEN JENTSCH AND MARKUS PAULY

Abstract. In this supplementary material we provide additional supporting simulations for the asymptotic test and all three randomization tests under consideration in a variety of examples.

1. Supplementary material

Additionally to the performance of the computational least-demanding randomization test φ_n^*, studied in the paper, we also illustrate the performance of the two other randomization tests $\varphi_{n,cent}^*$ and $\varphi_{n,stud}^*$, as proposed in Section 3. Moreover, we also show their performance for nominal sizes of α between 0 and 0.2 and additionally compare the power of the asymptotic test φ_n with all randomization tests.

Again we observe bivariate time series data ($\underline{X}_t = (X_{t,1}, X_{t,2})^T, t = 1, \ldots, n$) and we want to test the null hypothesis H_0 of equality of both corresponding one-dimensional spectral densities $f_1(\omega)$ and $f_2(\omega)$. In the setup of Section 1 of the paper, this means $q = 2$, $p = 1$ and $f_j(\omega) = \mathbf{F}_{jj}(\omega)$, $j = 1, 2$ and we test

H_0: $\{f_1(\omega) = f_2(\omega) \;\text{ for all }\; \omega \in [-\pi, \pi]\}$

against

H_1: $\{\exists\, A \subset \mathcal{B}([-\pi, \pi]) \text{ with } \lambda\!\!\lambda(A) > 0 : f_1(\omega) \neq f_2(\omega) \text{ for all } \omega \in A\}$.

In the following, we only consider data from vector moving average models

$$\underline{X}_t = \mathbf{B}\underline{e}_{t-1} + \underline{e}_t, \quad t \in \mathbb{Z} \tag{1.1}$$

and from vector autoregressive models

$$\underline{X}_t = \mathbf{A}\underline{X}_{t-1} + \underline{e}_t, \quad t \in \mathbb{Z} \tag{1.2}$$

of order one, respectively, where $\underline{e}_t \sim \mathcal{N}(0, \mathbf{\Sigma})$ is a bivariate normally distributed white noise process with covariance matrix $\mathbf{\Sigma}$. We consider these models for different choices of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Sigma}$ in the sequel. Recall that the (stationary) $VMA(1)$ and $VAR(1)$ processes in (1.1) and (1.2) possess spectral densities

$$\mathbf{f}_{VMA}(\omega) = \frac{1}{2\pi}\left(\mathbf{Id} + \mathbf{B}e^{-i\omega}\right)\mathbf{\Sigma}\overline{\left(\mathbf{Id} + \mathbf{B}e^{-i\omega}\right)}^T, \quad \omega \in [-\pi, \pi]$$

and

$$\mathbf{f}_{VAR}(\omega) = \frac{1}{2\pi}\left(\mathbf{Id} - \mathbf{A}e^{-i\omega}\right)^{-1}\mathbf{\Sigma}\overline{\left(\mathbf{Id} - \mathbf{A}e^{-i\omega}\right)^{-1}}^T, \quad \omega \in [-\pi, \pi],$$

respectively, where $\mathbf{Id}$ denotes the unit matrix.

1.1. Analysis of the size.

2000 *Mathematics Subject Classification.* 62M10, 62G10, 62G09.

Key words and phrases. multivariate time series; nonparametric tests; periodogram matrix; randomization tests; spectral density matrix.

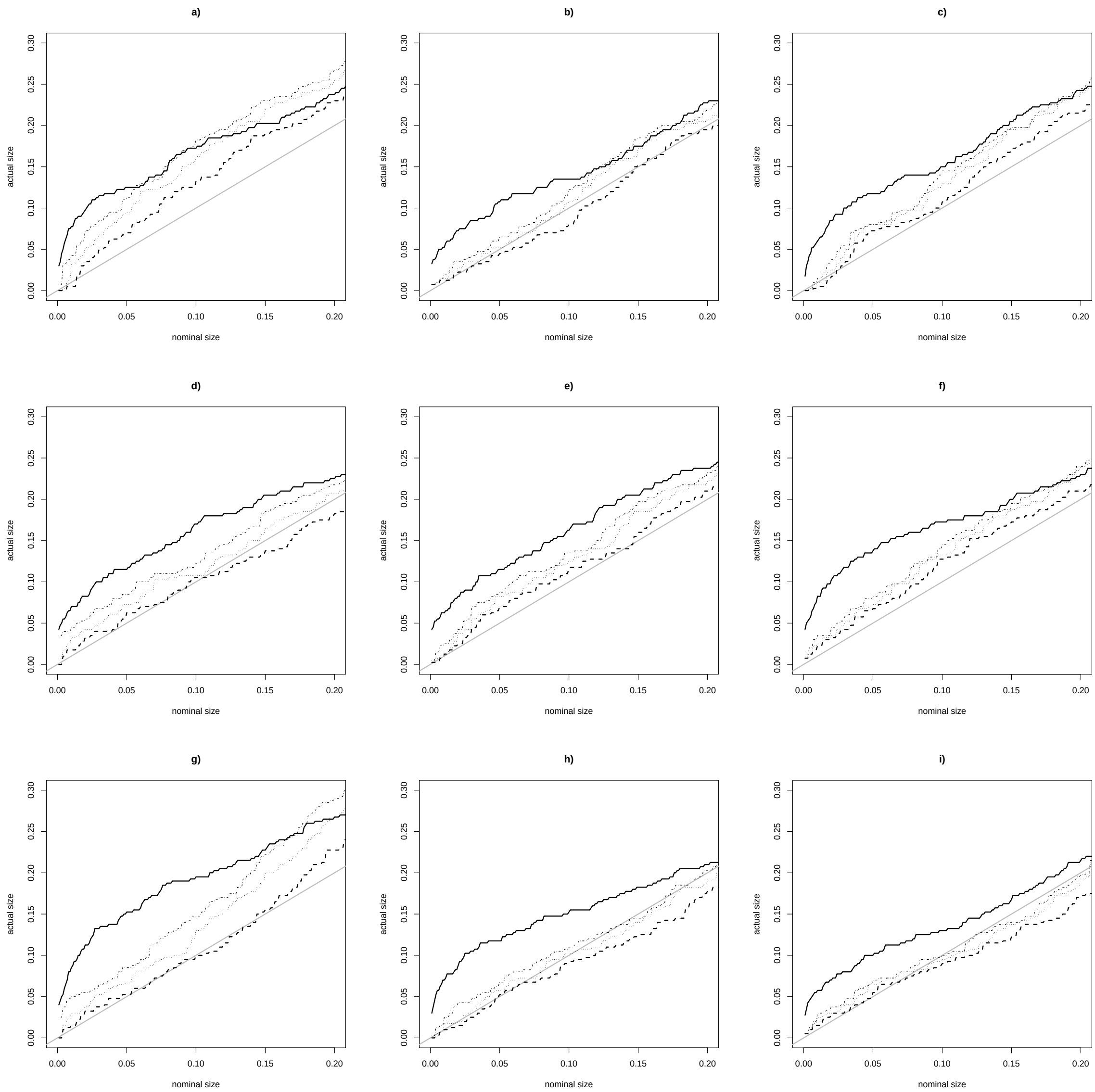

FIGURE 1. P-value plots for testing equality of two *independent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $MA_{H_0}(\mathbf{B}_i)$, $i = 1, 2, 3$ are used (top to bottom).

1.1.1. *Independent time series.*
To investigate the behavior of the test under the null in the independent case, we consider realizations from VMA and VAR models introduced in (1.1) and (1.2), respectively, where $\mathbf{B}$ and $\mathbf{A}$ are chosen as

$$\mathbf{A}_1 = \mathbf{B}_1 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}, \quad \mathbf{A}_2 = \mathbf{B}_2 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad \mathbf{A}_3 = \mathbf{B}_3 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$$

and $\mathbf{\Sigma} = \mathbf{\Sigma}_1 = diag(1, 1)$. The moving average model corresponding to $\mathbf{B}_i$ is denoted by $MA_{H_0}(\mathbf{B}_i)$ and the autoregressive model corresponding to $\mathbf{A}_i$ by $AR_{H_0}(\mathbf{B}_i)$.

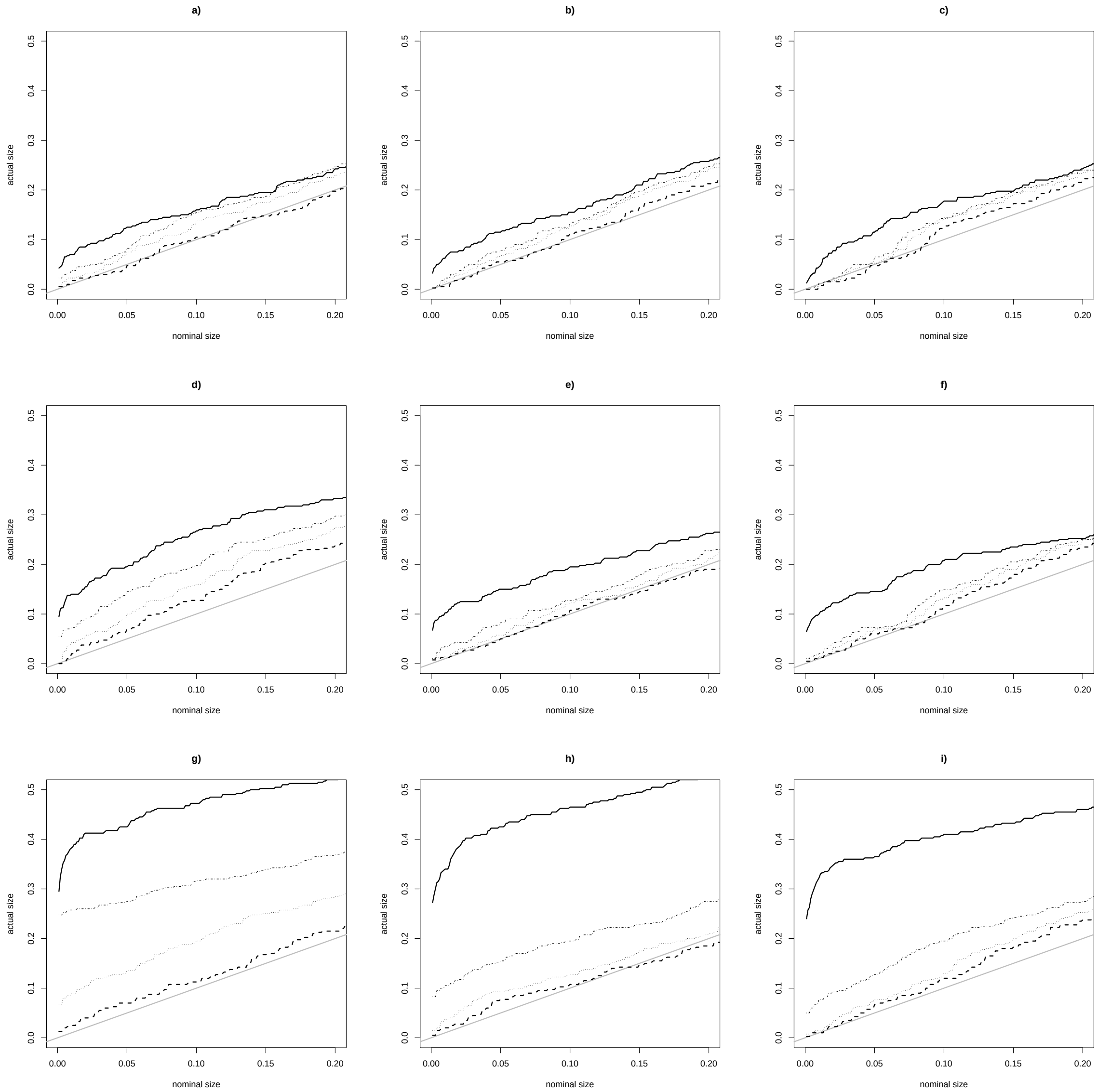

FIGURE 2. P-value plots for testing equality of two *independent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $AR_{H_0}(\mathbf{A}_i)$, $i = 1, 2, 3$ are used (top to bottom).

Observe that due to the diagonal shape of all involved matrices $\mathbf{\Sigma}_1$, $\mathbf{B}_i$ and $\mathbf{A}_i$, $i = 1, 2, 3$, we are dealing with two *independent* univariate time series.

For all models under consideration in this simulation section, we have generated $T = 400$ time series. For evaluation of the test statistic, we have chosen the bandwidth $h = 0.4$ for $n = 50$, $h = 0.3$ for $n = 100$, $h = 0.2$ for $n = 200$ and the Bartlett-Priestley kernel, see Priestley (1981, p448). Note that $A_K = \frac{6}{5}$ and $B_K = \frac{2672\pi}{385}$ for this particular kernel function. For each time series, the test φ_n has been executed with critical values from normal approximation as discussed in Section 2 and the randomization tests φ_n^*, $\varphi_{n,cent}^*$ and $\varphi_{n,stud}^*$ as discussed in Section 3 of the

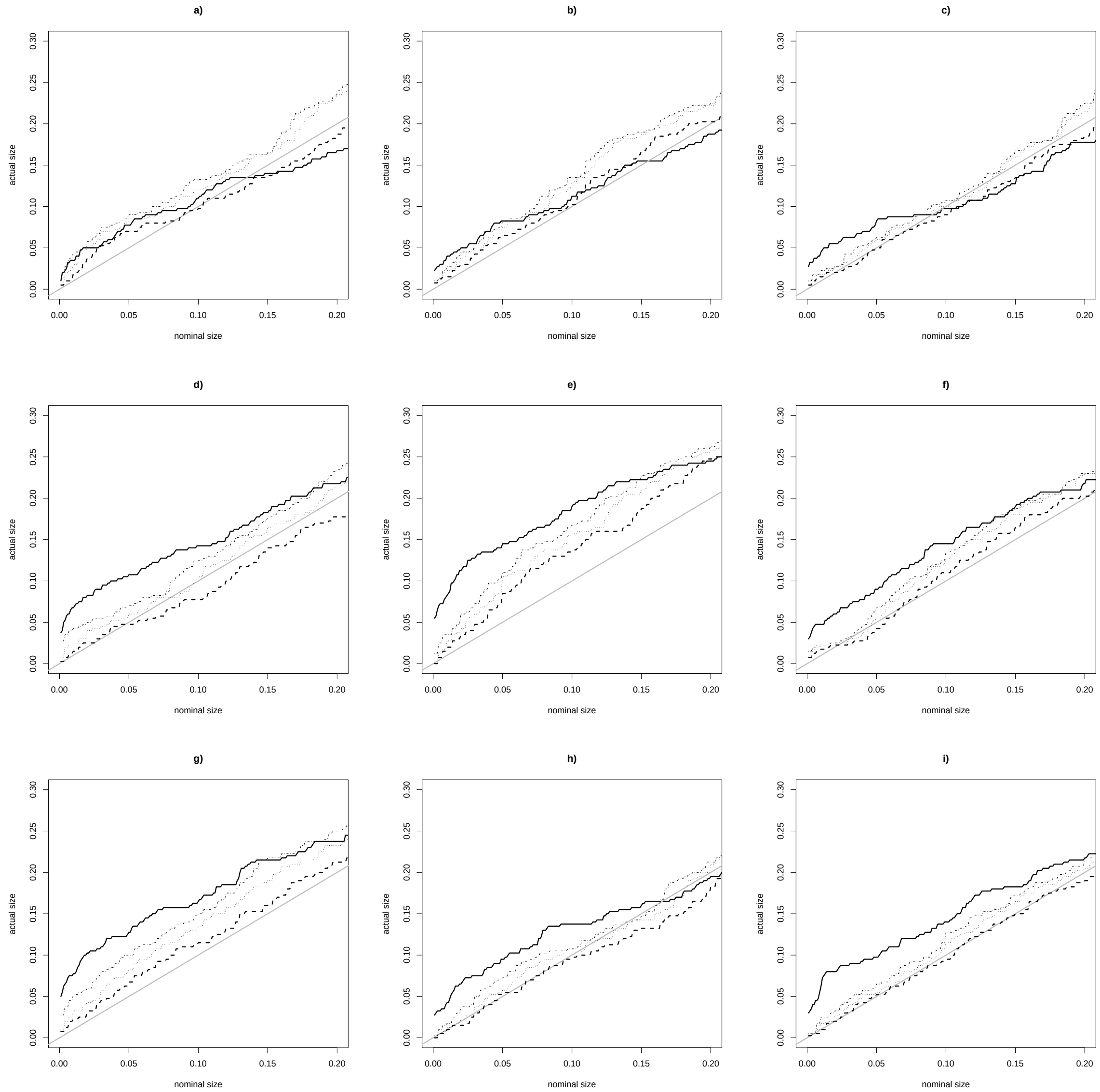

FIGURE 3. P-value plots for testing equality of two *dependent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $MA_{H_0}(\mathbf{B}_i)$, $i = 4, 5, 6$ are used (top to bottom).

paper, where $B = 300$ randomization Monte Carlo replicates have been used. The corresponding results are displayed in Figures 1-2 using p-value plots.

1.1.2. *Dependent time series.*

To investigate the behavior of the test under the null in the dependent case, we consider again realizations from the models (1.1) and (1.2), but now the matrix $\mathbf{B}$ is chosen from

$$\mathbf{B}_4 = \begin{pmatrix} 0.1 & 0.5 \\ 0.5 & 0.1 \end{pmatrix}, \quad \mathbf{B}_5 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \quad \mathbf{B}_6 = \begin{pmatrix} 0.9 & 0.5 \\ 0.5 & 0.9 \end{pmatrix}, \tag{1.3}$$

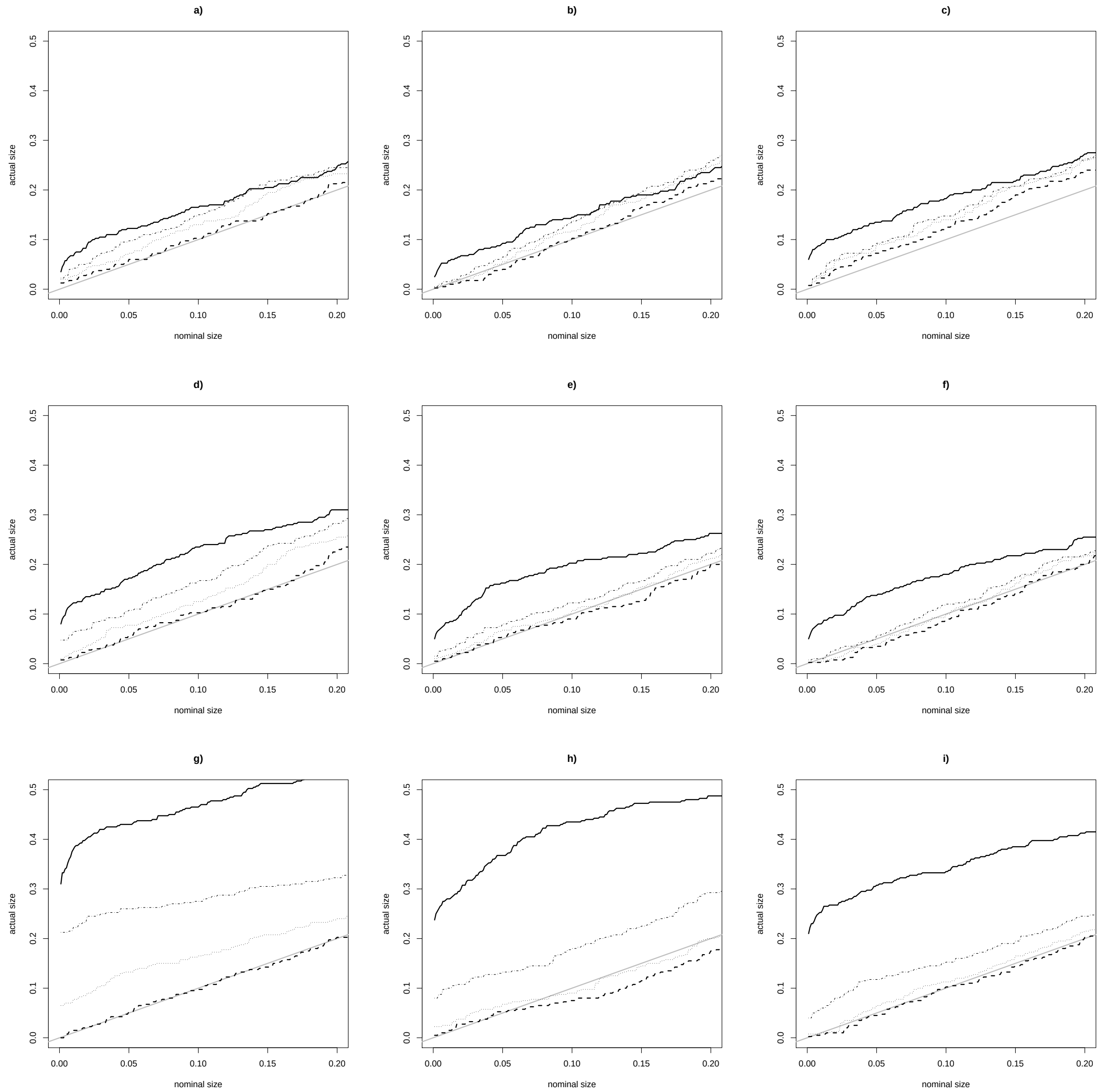

FIGURE 4. P-value plots for testing equality of two *dependent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $AR_{H_0}(\mathbf{A}_i)$, $i = 4, 5, 6$ are used (top to bottom).

denoted by $MA_{H_0}(\mathbf{B}_i)$, $i = 4, 5, 6$, and the matrix $\mathbf{A}$ from

$$\mathbf{A}_4 = \begin{pmatrix} 0.1 & 0.5 \\ 0.5 & 0.1 \end{pmatrix}, \quad \mathbf{A}_5 = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}, \quad \mathbf{A}_6 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}, \tag{1.4}$$

denoted by $AR_{H_0}(\mathbf{A}_i)$, $i = 4, 5, 6$, and

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Observe that due to the shape of $\mathbf{\Sigma}_2$, $\mathbf{B}_i$ and $\mathbf{A}_i$, $i = 4, 5, 6$, we are dealing with two *dependent* time series whose marginal spectral densities are equal. The results are shown in Figures 3-4.

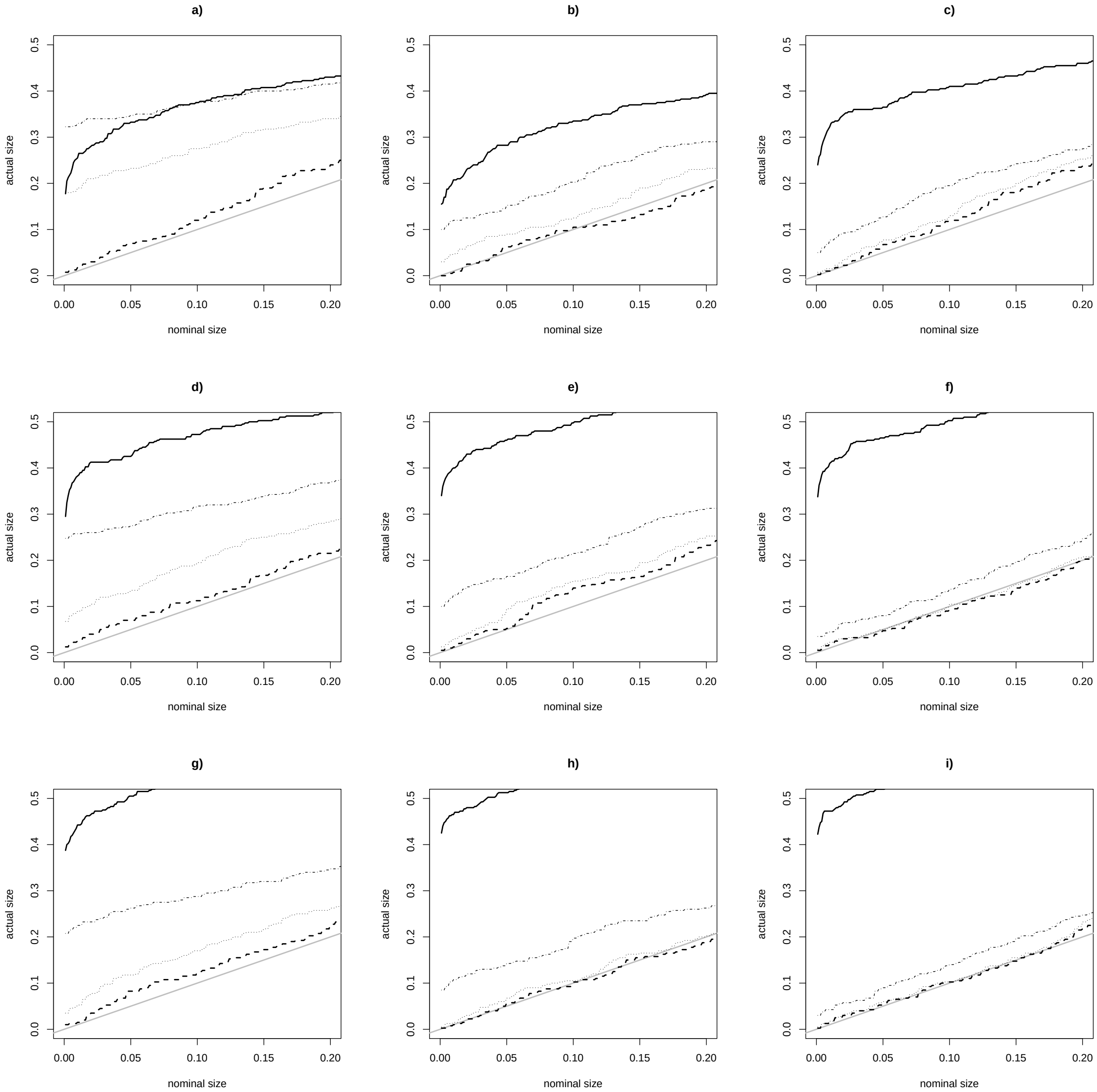

FIGURE 5. P-value plots for testing equality of two *independent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted) in model $AR_{H_0}(\mathbf{A}_3)$. Sample size $n \in \{50, 100, 200\}$ increases (left to right) and bandwidth $h \in \{0.2, 0.4, 0.6\}$ increases (top to bottom).

It is also worth noting that setting the off-diagonal elements in the second and third coefficient matrix in (1.4) equal to 0.5 would result in non-stationary (explosive) autoregressive models, which are not within the scope of this paper.

1.1.3. *Sensitivity of the test with respect to bandwidth choice under* H_0.
As already anticipated in the paper we will also analyze the size for different choices of bandwidths. Therefore we have repeated the simulations described above for model $AR_{H_0}(\mathbf{A}_3)$ and $MA_{H_0}(\mathbf{B}_5)$ for different bandwidth choices $h \in \{0.2, 0.4, 0.6\}$ to check how sensitive the tests behave with respect to its choice. The results are displayed in Figures 5-6. Note that in model

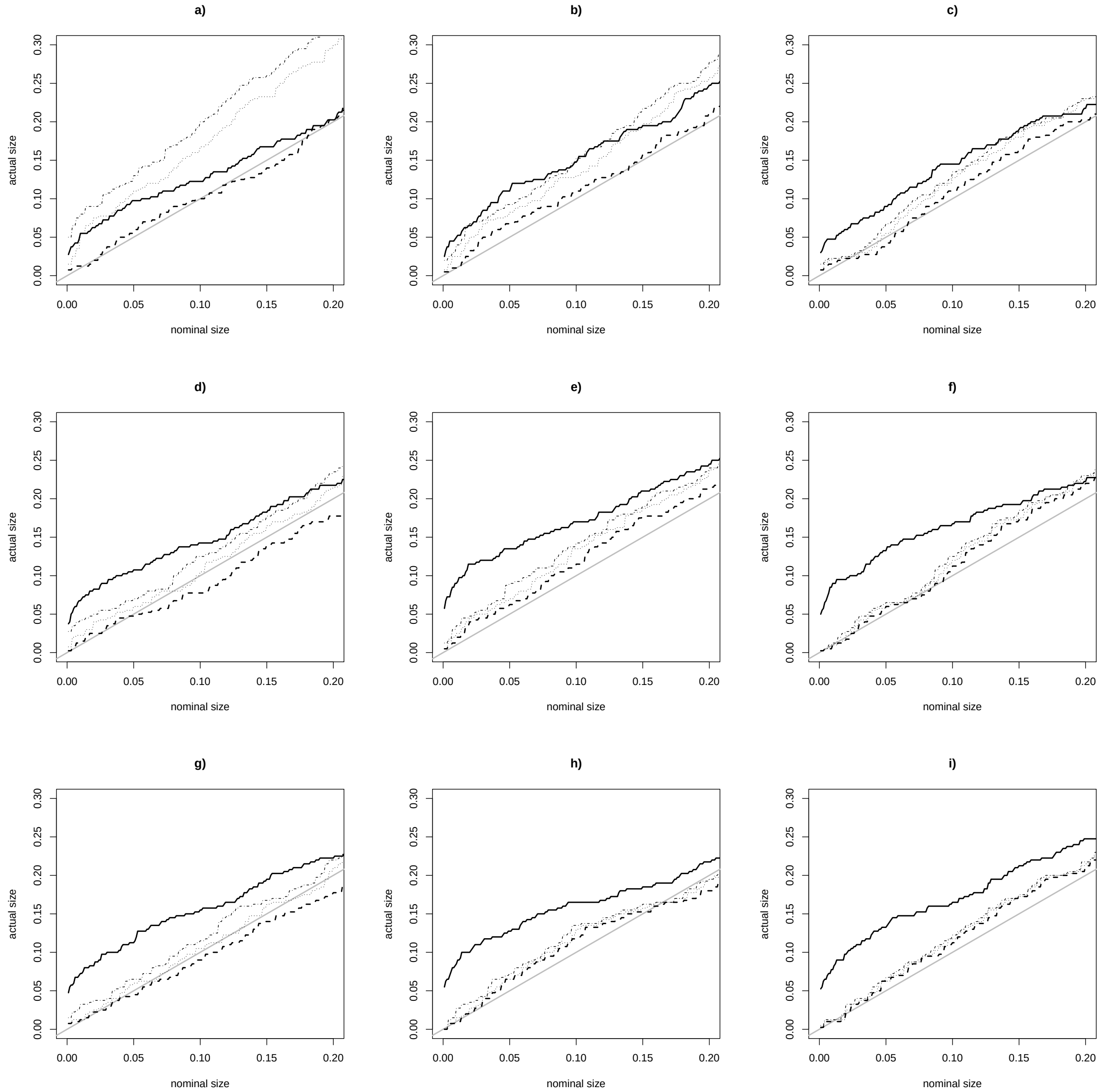

FIGURE 6. P-value plots for testing equality of two *independent* spectral densities. Actual size vs. nominal size is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted) in model $MA_{H_0}(\mathbf{B}_3)$. Sample size $n \in \{50, 100, 200\}$ increases (left to right) and bandwidth $h \in \{0.2, 0.4, 0.6\}$ increases (top to bottom).

$AR_{H_0}(\mathbf{A}_3)$, the tests have to compare two independent time series that have the same unimodal spectral density. In model $MA_{H_0}(\mathbf{B}_5)$, the tests have to compare two dependent time series that have the same spectral density with a more flat shape. Remark that we have chosen these models to cover two very distinct cases.

1.2. **Analysis of the power.**

To illustrate the behavior of the tests under the alternative, that is under inequality of both spectral densities, we consider two models that generate independent and dependent time series,

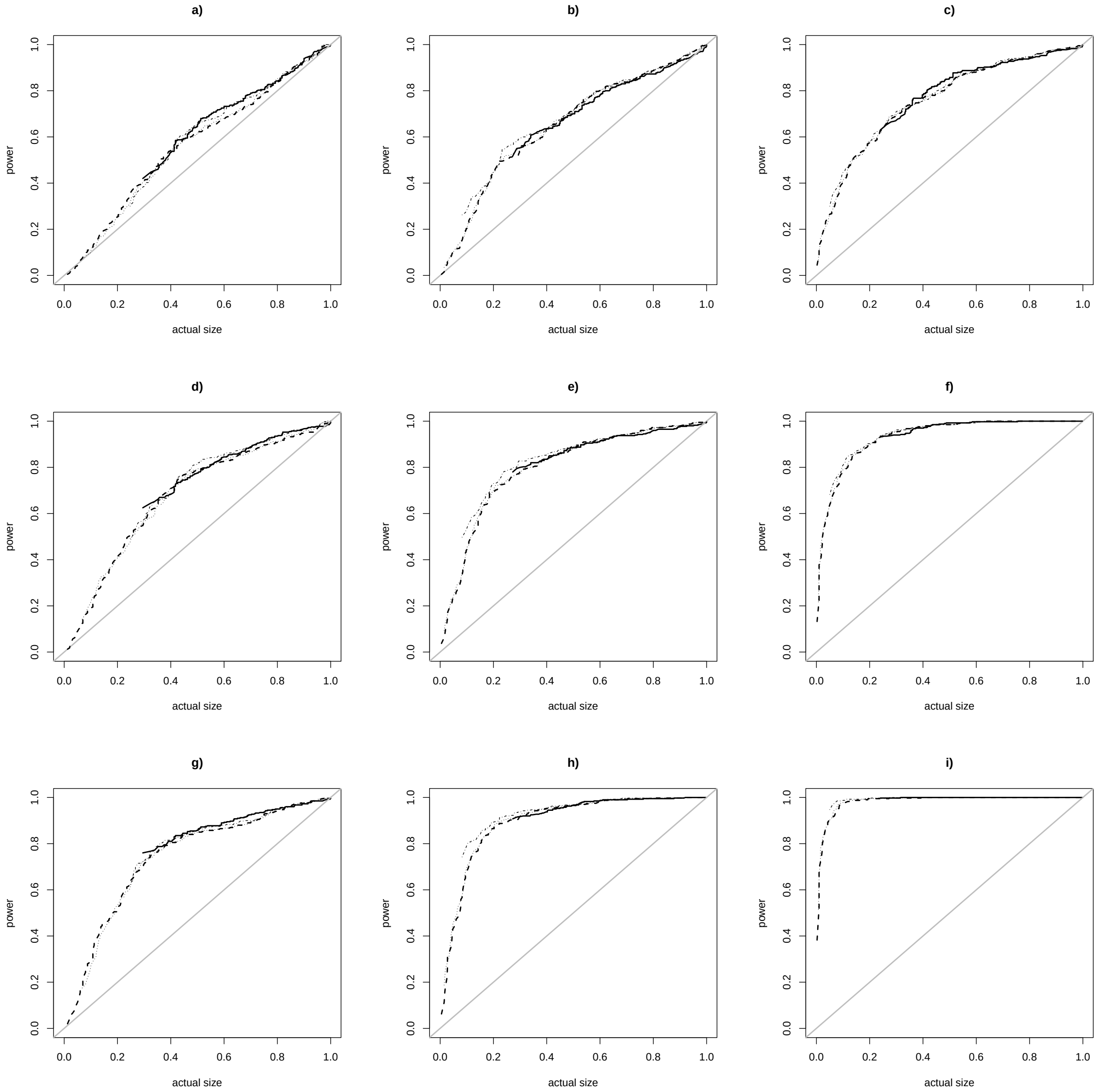

Figure 7. *Achieved* size power curves for testing equality of two *independent* spectral densities. Actual size from Figure 1 vs. power is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $AR_{H_1}(\mathbf{A}_i)$ $i = 7, 8, 9$ are used (top to bottom).

respectively. First, we consider realizations from the autoregressive model in (1.2). Here, $\mathbf{A}$ is chosen from

$$\mathbf{A}_7 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.8 \end{pmatrix}, \quad \mathbf{A}_8 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}, \quad \mathbf{A}_9 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.6 \end{pmatrix}$$

and $\mathbf{\Sigma} = \mathbf{\Sigma}_1$. The corresponding models are denoted by $AR_{H_1}(\mathbf{A}_i)$, $i = 7, 8, 9$, and due to the diagonal shape, we are dealing with two independent time series. In the second case, we generate

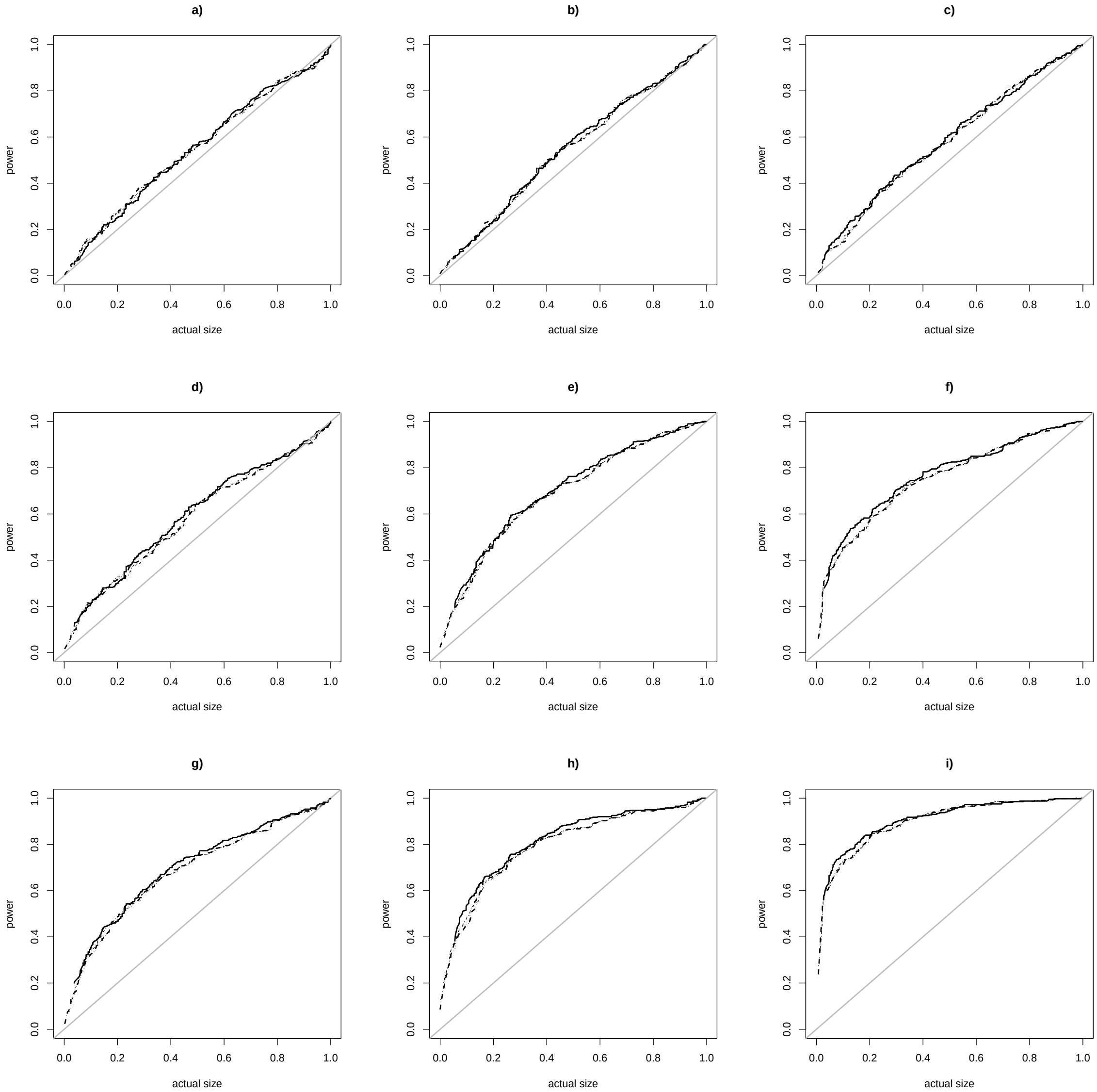

FIGURE 8. *Achieved* size power curves for testing equality of two *dependent* spectral densities. Actual size from Figure 1 vs. power is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted). Sample size $n \in \{50, 100, 200\}$ increases (left to right) and models $MA_{H_1}(\mathbf{B}_i)$ $i = 7, 8, 9$ are used (top to bottom).

realizations from the moving average model in (1.1) with $\mathbf{B}$ chosen from

$$\mathbf{B}_7 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.6 \end{pmatrix}, \quad \mathbf{B}_8 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}, \quad \mathbf{B}_9 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.8 \end{pmatrix}$$

and $\mathbf{\Sigma} = \mathbf{\Sigma}_2$ denoted by $MA_{H_1}(\mathbf{B}_i)$, $i = 7, 8, 9$. Due to non-diagonal shape, we are dealing with two dependent time series in this case.

As shown in Figures 1 - 6, some tests tend to over-reject the null hypothesis systematically. Therefore, it seems to be unfair to compare just the usual size power curves (nominal size vs.

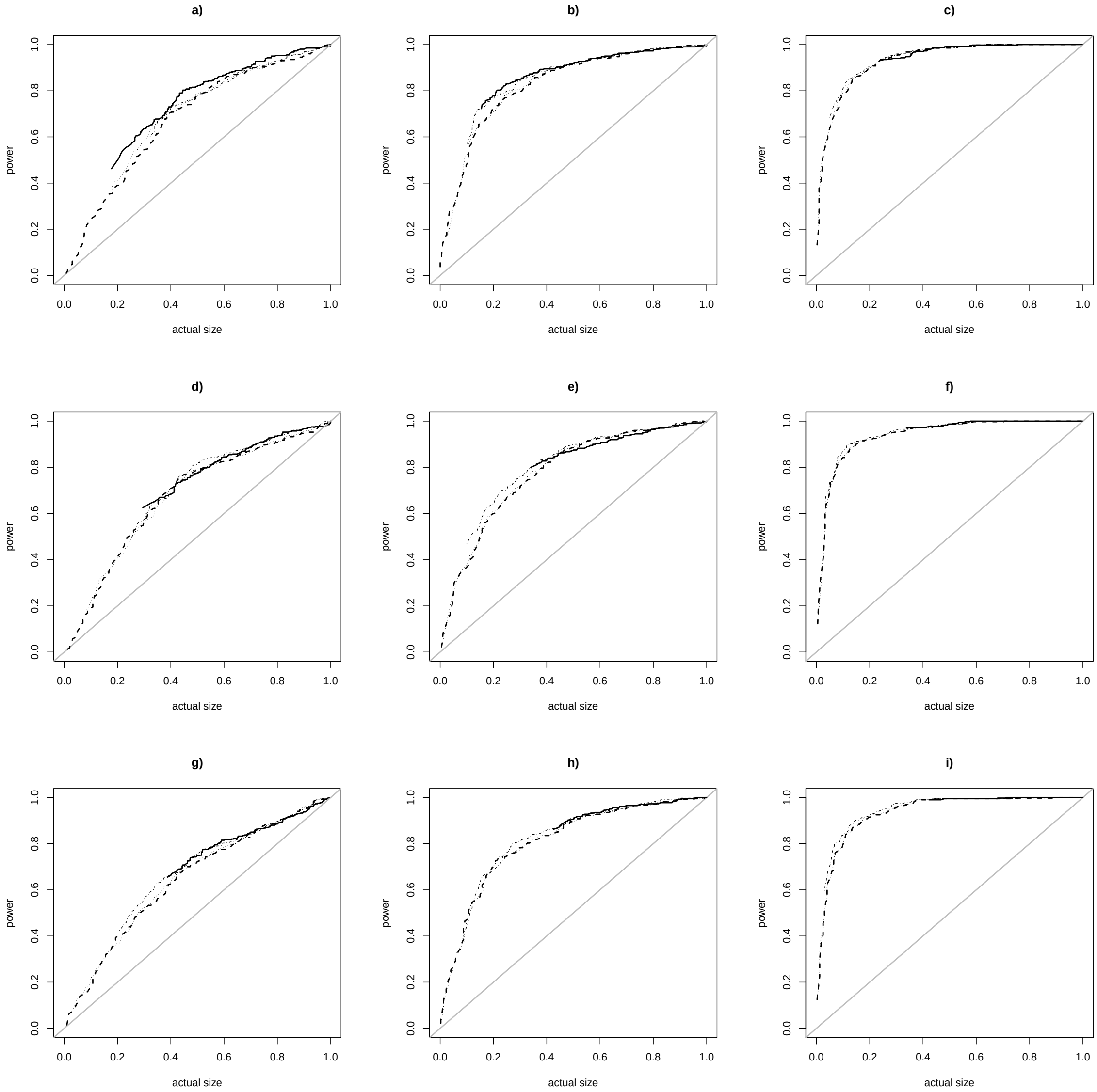

FIGURE 9. *Achieved* size power curves for testing equality of two *dependent* spectral densities. Actual size from model $AR_{H_0}(\mathbf{A}_3)$ displayed in Figure 5 vs. power is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted) in model $AR_{H_1}(\mathbf{A}_8)$. Sample size $n \in \{50, 100, 200\}$ increases (left to right) and bandwidth $h \in \{0.2, 0.4, 0.6\}$ increases (top to bottom)

power) to judge their performances under the alternative. For this reason, we present *achieved* size power curves that use actual sizes obtained for models $AR_{H_0}(\mathbf{A}_3)$ and $MA_{H_0}(\mathbf{B}_5)$ as displayed in panels g)-i) of Figure 2 and in d)-f) of Figure 3, respectively, instead of the nominal sizes. A comparison of these plots is shown in Figures 7 and 8 for $AR_{H_1}(\mathbf{A}_i)$ (independent case) and $MA_{H_1}(\mathbf{A}_i)$ (dependent case), $i = 7, 8, 9$, respectively. When studying these plots it seems that there are some missing values, especially for the unconditional test φ_n. However, these can be explained by the bad actual sizes of some tests, e.g. the actual size of φ_n in model $AR_{H_0}(\mathbf{A}_3)$ with a sample size of $n = 50$ is always greater than 0.29, see panel g) of Figure 2. This explains e.g. the 'missing values' in panel g) of Figure 7 for actual sizes less than 0.29

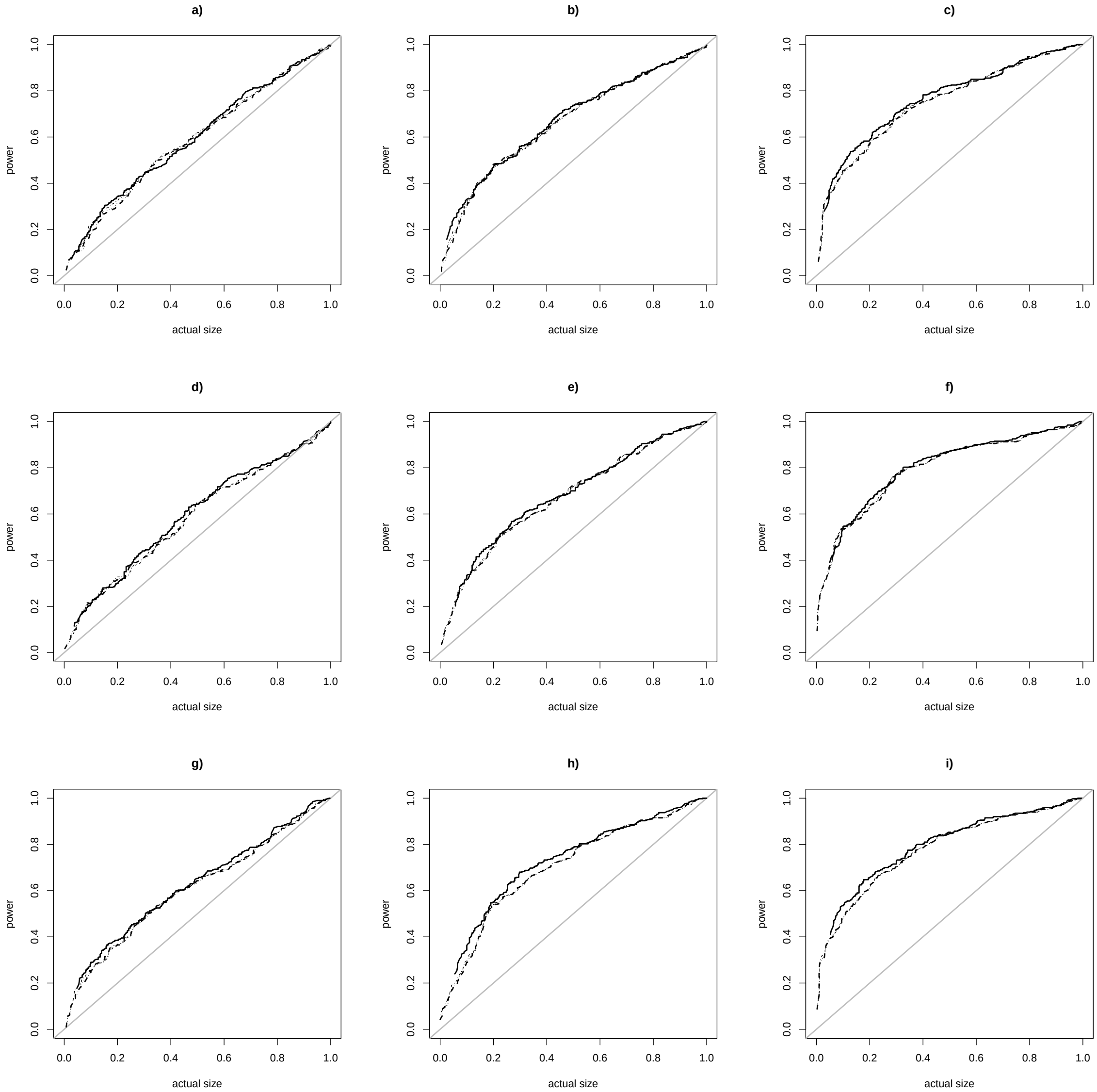

Figure 10. *Achieved* size power curves for testing equality of two *dependent* spectral densities. Actual size from model $MA_{H_0}(\mathbf{B}_3)$ displayed in Figure 6 vs. power is shown for φ_n (solid), φ_n^* (dashed), $\varphi_{n,cent}^*$ (dotted) and $\varphi_{n,stud}^*$ (dashed and dotted) in model $MA_{H_1}(\mathbf{B}_8)$. Sample size $n \in \{50, 100, 200\}$ increases (left to right) and bandwidth $h \in \{0.2, 0.4, 0.6\}$ increases (top to bottom)

1.2.1. *Sensitivity of the test with respect to bandwidth choice under* H_1.

To conclude with the analysis of the power, we have repeated the simulations described above for models $AR_{H_1}(\mathbf{A}_8)$ and $MA_{H_1}(\mathbf{B}_8)$ for different bandwidths $h \in \{0.2, 0.4, 0.6\}$ to check how sensitive the tests behave with respect to its choice. Note that in model $AR_{H_1}(\mathbf{A}_8)$, the tests have to compare two independent time series that have different unimodal spectral densities. In model $MA_{H_1}(\mathbf{B}_8)$, the tests have to compare two dependent time series that have different spectral densities with a more flat shape.

1.3. **Discussion.**

From Figures 1 - 6, it can be seen that the asymptotic test φ_n has difficulties in keeping the prescribed level and tends to overrejects the null systematically for all small ($n = 50$) and moderate ($n = 100$) sample sizes in all situations. Its performance is not even desirable for larger sample sizes ($n = 200$). Especially for the most critical autoregressive models $AR_{H_0}(\mathbf{A}_3)$ (independent case) and $AR_{H_0}(\mathbf{A}_6)$ (dependent case), where the corresponding spectral densities have non-flat shapes and are rather difficult to estimate, its null approximation is extremely poor and the performance is unacceptable. Nevertheless, this poor performance is not surprising since the slow convergence speed of L_2-type statistics is already known, see for instance Paparoditis (2000).

In comparison to that, all randomization tests φ_n^*, $\varphi_{n,cent}^*$ and $\varphi_{n,stud}^*$ perform better than φ_n. In particular the computationally least-demanding version φ_n^* holds the prescribed level very satisfactorily in all considered situations. However, the other randomization tests $\varphi_{n,cent}^*$ and $\varphi_{n,stud}^*$ tend to be a little bit more liberal than φ_n^* for smaller sample sizes, where $\varphi_{n,cent}^*$ has in general a better output than $\varphi_{n,stud}^*$. Actually both do not perform very well for the small sample size of $n = 50$ particularly for the critical models $AR_{H_0}(\mathbf{A}_3)$ and $AR_{H_0}(\mathbf{A}_6)$ in comparison to φ_n^*. Nevertheless, with increasing sample size ($n = 100$, 200) their performance gets much better; especially the centered version $\varphi_{n,cent}^*$ has comparable results to φ_n^*. An explanation for this observations is given by the slow convergence speed of the mean and variance estimators, where spectral density estimators are involved. Observe that no additional quantities as μ_0 or τ_0 have to be estimated nonparametrically for φ_n^*.

Finally, Figures 5 and 6 show that the bandwidth selection only has a slight effect on the behavior of the randomization tests, where this choice is more crucial for φ_n. Again particularly φ_n^* does not seem to react sensitively to the bandwidth choice in the small sample case. For larger sample sizes all randomization procedures do not seem to be very sensitive to variations of h. This suggests that the randomization technique can be applied without being unsure about setting any tuning parameters, which is a big advantage in comparison to other resampling methods, which usually depend on even more tuning parameters. Nevertheless, one can still apply a computational-more demanding data-driven bandwidth selection method as has been done in the simulation study of the paper.

Figures 7 and 8 show the power behavior of all tests. Note that we only show the fair comparison, where we compare the power of the tests to their actual size. It can be seen from these figures that there is actually no big difference in the power behavior between all four considered tests. Moreover, when studying the panels with increasing sample sizes (from left to right) the theoretically proved consistency results from the paper can be confirmed by the simulations. As already mentioned above note that the late start of the curves of some tests (especially of φ_n in Figure 8) is reasoned by their poor actual size performances in these situations as shown in Figures 2 and 3.

Finally, Figures 9 and 10 illustrate that the bandwidth selection again only has a slight effect on the power behavior of all randomization tests.

Additionally to the simulation experience of the paper we may summarize the current experience as follows:

- The randomization technique makes sure that the corresponding tests keep the prescribed level for small sample sizes quite well.

- All considered randomization tests have the big advantage that their performances do not depend on the choice of any additional tuning parameter except for the bandwidth and its choice is not as crucial as for the asymptotic test.
- From all tests the randomization test φ_n^* keeps the prescribed level almost always the best, is computationally least-demanding than the other randomization tests and it can be recommended in particular for small samples and linear time series.
- The performance of φ_n^* becomes even more excellent if one compares its behavior for the very small sample size of $n = 50$ with the poor performance of the unconditional test.
- Furthermore, the power performance of all tests (measured as power in comparison to actual size) is not really distinguishable and, as usual under consistency, improves for all tests with increasing sample size, see also the extensive simulation results in the paper.
- Finally, to sum up, the randomization procedure helps to hold the prescribed level under the null more satisfactorily and does not forfeit power under the alternative in comparison to the unconditional case

Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany
E-mail address: `cjentsch@mail.uni-mannheim.de (address for correspondence)`

Institute of Mathematics, University of Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany
E-mail address: `markus.pauly@uni-duesseldorf.de`