Comment on "HAC Corrections for Strongly Autocorrelated Time Series" by Ulrich K. Müller

Yixiao Sun
Department of Economics, UC San Diego

May 20, 2014

1 On the Nearly-optimal Test

Müller applies the theory of optimal statistical testing to heteroskedasticity and autocorrelation robust (HAR) inference in the presence of strong autocorrelation. As a starting point, Müller uses Le Cam's idea on the limits of experiments ingeniously and converts a more complicated finite sample testing problem into an asymptotically equivalent and simpler testing problem. The main barrier to optimal testing is that both the null hypothesis and alternative hypothesis are composite, even after the asymptotic reduction based on Le Cam's idea. So the Neyman-Pearson lemma does not directly apply.

To reduce the dimension of the alternative hypothesis space, it is standard practice to employ a weighting function and take a weighted average of the probability distributions under the alternative hypothesis. See, for example, Cox and Hinkley (1974, pp. 102). The weighting function should reflect a user's belief about the likelihood of different parameter values and the associated cost of false acceptance under the alternative. Selecting the weighting function is as difficult as selecting one point out of many possible parameter values. A test that is designed to be optimal against a point alternative may not be optimal for other alternatives. The nearly-optimality of Müller's test should be interpreted with this point in mind.

There are a number of ways to reduce the dimension of the null hypothesis space, including the invariance arguments and the conditioning argument on sufficient statistics. See, for example, Cox and Hinkley (1974, Ch. 5). In fact, Müller uses scale invariance to remove one nuisance parameter. However, as in many other contexts, here the null cannot be reduced to a point by using the standard arguments. Müller follows Wald, Lehmann, Stein and other pioneers in statistical testing and construct the so-called least favorable distribution over the null parameter space and use it to average the probability distributions. This effectively reduces the composite null into a simple one. However, the least favorable distribution has to be found numerically, which can be a formidable task. This is perhaps one of the reasons that the theory of optimal testing has not been widely used in statistics and econometrics. A contribution of Müller's paper is to find an approximate least favorable distribution and construct a test that is nearly optimal.

The reason to employ the least favorable distribution is that we want to control the level of the test for each point in the parameter space under the null. While we are content with the average power under the alternative, we are not satisfied with the control of the average level under the null. In fact, the requirement on size control is even stronger: the

null rejection probability has to be controlled uniformly over the parameter space under the null. There is some contradiction here, which arises from the classical dogma that puts size control before power maximization. The requirement that the null rejection probability has to be controlled for each possible parameter value under the null, no matter how unlikely *a priori* a given value is, is overly conservative. The test designed under this principle can suffer from a severe power loss. In fact, in the simulation study, Müller's test often has a lower (size-adjusted) power than some commonly used tests. I am sympathetic with the argument that the power loss is the cost one has to pay to achieve the size accuracy as size-adjustment is not empirically feasible. However, one can always design a test with accurate size but no power. Ultimately, there is a trade-off between size control and power improvement. Using the least favorable distribution does not necessarily strike the optimal trade-off. As a comprise, one may want to control the average level/size of the test over a few empirically relevant regions in the parameter space.

There is some convincing simulation evidence that Müller's test is much more accurate than existing tests for some data generating processes (DGP). These are the DGP's where the finite sample testing problem can be approximated very well by the asymptotically equivalent testing problem. However, there is not much research on the quality of the approximation. If the approximation error is large, Müller's test, which is nearly optimal for the asymptotically equivalent problem, may not be optimal for the original problem. For example, when the error process in the location model is an AR(1) plus noise, Müller's simulation results show that his test can still over-reject considerably. As another example, if the error process follows the AR(2) model $u_t = 1.90u_{t-1} - 0.95u_{t-2} + e_t$ where $e_t \sim iid$ N(0,1), then Müller's test (and many other tests) suffers from under-rejection. In this example, the modulus of the larger root is about 0.97, which is close to one. However, the spectral density does not resemble that of an AR(1) process. It does not have a peak at the origin. Instead, there is a speak near the origin. As a result, the quality of the AR(1) approximation is low. If the periodograms used in the variance estimator include the peak, then the variance estimator will be biased upward, leading to a smaller test statistic and under-rejection.

2 Near-unity Fixed-smoothing Asymptotic Approximation

An attractive feature of Müller's test is that the scenarios under which it is optimal or nearly optimal are given explicitly. However, practitioners may find it unattractive because of the computation cost, the unfamiliar form of the test statistic, and its applicability to models beyond the simple Gaussian location model. An alternative approach to deal with strong autocorrelation is to derive a new approximation for the conventional test statistic that captures the strong autocorrelation. This has been recently developed in Sun (2014b).

To provide the context for further discussion, I give a brief summary of Sun (2014b) here. Consider a p-dimensional time series y_t of the form:

$$y_t = \theta + e_t, t = 1, 2, \ldots, T \tag{1}$$

where $y_t = (y_{1t}, \ldots, y_{pt})'$, $\theta = (\theta_1, \ldots, \theta_p)'$, and $e_t = (e_{1t}, \ldots, e_{pt})'$ is a zero mean process. We are interested in testing the null $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta \neq \theta_0$. The OLS estimator of θ is the average of $\{y_t\}$, i.e., $\hat{\theta} = \bar{y} := T^{-1} \sum_{t=1}^{T} y_t$. The F test version of

the Wald statistic based on the OLS estimator is given by

$$F_T = (\hat{\theta} - \theta_0)' \hat{\Omega}^{-1} (\hat{\theta} - \theta_0)/p$$

where $\hat{\Omega}$ is an estimator of the approximate variance of $(\hat{\theta} - \theta_0)$. When $p = 1$, we can construct the t-statistic $t_T = (\hat{\theta} - \theta_0)/\hat{\Omega}^{1/2}$.

A very general class of variance estimators is the class of quadratic variance estimators, which takes the following form:

$$\hat{\Omega} = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h \left(\frac{t}{T}, \frac{s}{T} \right) \hat{e}_t \hat{e}_s' \tag{2}$$

where $\hat{e}_t = e_t - \bar{e}$ for $\bar{e} = T^{-1} \sum_{t=1}^{T} e_t$ and $Q_h(r,s)$ is a weighting function that depends on the smoothing parameter h. When $Q_h(r,s) = k((r-s)/b)$ for some kernel function $k(\cdot)$ and smoothing parameter b, $\hat{\Omega}$ is the commonly used kernel variance estimator. When $Q_h(r,s) = K^{-1} \sum_{j=1}^{K} \phi_j(r) \phi_j(s)$ for some basis functions $\{\phi_j(r)\}$ on $\mathbb{L}^2[0,1]$ satisfying $\int_0^1 \phi_j(r)\, dr = 0$ and smoothing parameter K, we obtain the so-called series variance estimator. This estimator has a long history. It can be regarded as a multiple-window estimator with window function $\phi_k(t/T)$. See Thomson (1982). It also belongs to the class of filter-bank estimators and $\hat{\Omega}$ is a simple average of the individual filter-bank estimators. For more discussions along this line, see Chapter 5 of Stoica and Moses (2005). Recently there has been some renewed interest in this type of variance estimators, see Phillips (2005), Sun (2006, 2011, 2013), and Müller (2007).

Define

$$Q_{T,h}^*(r,s) = Q_h(r,s) - \frac{1}{T} \sum_{\tau_1=1}^{T} Q_h\left(\frac{\tau_1}{T}, s\right) - \frac{1}{T} \sum_{\tau_2=1}^{T} Q_h\left(r, \frac{\tau_2}{T}\right) + \frac{1}{T} \sum_{\tau_1=1}^{T} \sum_{\tau_2=1}^{T} Q_h\left(\frac{\tau_1}{T}, \frac{\tau_2}{T}\right),$$

then

$$\hat{\Omega} = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_{T,h}^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t e_s'. \tag{3}$$

The Wald statistic is then equal to

$$F_T = \left(\sum_{t=1}^{T} e_t \right)' \left[\sum_{t=1}^{T} \sum_{s=1}^{T} Q_{Th}^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t e_s' \right]^{-1} \left(\sum_{t=1}^{T} e_t \right) /p.$$

Similarly, the t-statistic becomes

$$t_T = \frac{\sum_{t=1}^{T} e_t}{\left[\sum_{t=1}^{T} \sum_{s=1}^{T} Q_{Th}^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t e_s \right]^{1/2}}.$$

The question is how to approximate the sampling distributions of F_T and t_T. If $\{e_t\}$ is stationary and $T^{-1/2} \sum_{t=1}^{[Tr]} e_t$ converges weakly to a Brownian motion process, then under some conditions on Q_h, it can be shown that, for a fixed h :

$$F_T \to^d F_\infty(h) := W_p(1)' \left[\int_0^1 \int_0^1 Q_h^*(r,s)\, dW_p(r)\, dW_p'(s) \right]^{-1} W_p(1)/p, \tag{4}$$

$$t_T \to^d t_\infty(h) := \frac{W_p(1)}{\sqrt{\int_0^1 \int_0^1 Q_h^*(r,s)\, dW_p(r)\, dW_p'(s)}}, \tag{5}$$

where $W_p(r)$ is a $p \times 1$ vector of standard Wiener processes and

$$Q_h^*(r,s) = Q_h(r,s) - \int_0^1 Q_h(\tau_1, s)\, d\tau_1 - \int_0^1 Q_h(r,\tau_2)\, d\tau_2 + \int_0^1 \int_0^1 Q_h(\tau_1,\tau_2)\, d\tau_1 d\tau_2.$$

For easy reference, I refer to the above approximations as the stationary fixed-smoothing asymptotic approximations. They are more accurate than the chi-square approximation or the normal approximation. As pointed out by Müller's paper, these approximations are still not good enough when e_t is highly autocorrelated.

To model the high autocorrelation, we assume that e_t follows an AR(1) process of the form:

$$e_t = \rho_T e_{t-1} + u_t \text{ where } e_0 = O_p(1) \text{ and } \rho_T = 1 - \frac{c_m}{T}$$

for some sequence $\{c_m\}$. This is in the spirit similar to Müller's paper and many other papers in the literature. See, for example, Phillips, Magdalinous and Giraitis (2010). Under the assumption that

$$\frac{1}{\sqrt{T}} e_{[Tr]} \to \Lambda J_{c_m}(r)$$

for some matrix Λ, where $J_{c_m}(r)$ is the Ornstein-Uhlenbeck process defined by

$$dJ_{c_m}(r) = -c_m J_{c_m}(r)\, dr + dW_p(r)$$

with $J_{c_m}(0) = 0$, we can obtain the following near-unity fixed-smoothing asymptotics when c_m and h are fixed:

$$\begin{aligned} F_T &\to {}^d F_\infty(c_m, h) \\ &: = \left[\int_0^1 J_{c_m}(r)\, dr\right]' \left[\int_0^1 \int_0^1 Q_h^*(r,s)\, J_{c_m}(r)\, J_{c_m}'(s)\, dr ds\right]^{-1} \left[\int_0^1 J_{c_m}(r)\, dr\right] / p. \end{aligned}$$

If we further assume that $Q_h(r,s)$ is positive definite, then for fixed c_m and h :

$$t_T \to^d t_\infty(c_m, h) := \frac{\int_0^1 J_{c_m}(r)\, dr}{\left[\int_0^1 \int_0^1 Q_h^*(r,s)\, J_{c_m}(r)\, J_{c_m}'(s)\, dr ds\right]^{1/2}}.$$

If we let $c_m \to \infty$, then the near-unity fixed-smoothing asymptotic distributions $F_\infty(c_m, h)$ and $t_\infty(c_m, h)$ approach the stationary fixed-smoothing asymptotic distributions given in (4) and (5). On the other hand, if we let $c_m \to 0$, then $F_\infty(c_m, h)$ and $t_\infty(c_m, h)$ approach the unit-root fixed-smoothing asymptotic distributions, which are defined as $F_\infty(c_m, h)$ and $t_\infty(c_m, h)$ but with $J_{c_m}(r)$ replaced by $W_p(r)$. Depending on the value of c_m, the limiting distributions $F_\infty(c_m, h)$ and $t_\infty(c_m, h)$ provide a smooth transition from the usual stationary fixed-smoothing asymptotics to the unit-root fixed-smoothing asymptotics.

In my view, the chi-square/normal approximation, the stationary fixed-smoothing approximation and the near-unity fixed-smoothing approximation are just different approximations to the same test statistic constructed using the same variance estimator. It is a little misleading to talk about consistent and inconsistent variance estimators. The variance estimator is actually the same but we embed it on different asymptotic paths. When the fixed-smoothing asymptotics is used, we do not necessarily require that we fix the smoothing parameter h in finite samples. In fact, in empirical applications, the sample size T is

usually given beforehand and the smoothing parameter h needs to be determined using a priori information and/or information obtained from the data. Very often, the selected smoothing parameter h is larger for a larger sample size but is still small relative to the sample size. So the empirical situations appear to be more compatible with the conventional increasing-smoothing asymptotics. The beauty of the fixed-smoothing asymptotics is that fixed-smoothing critical values are still correct under the increasing-smoothing asymptotics. In fact, in a sequence of papers (e.g. Sun, 2014a), I have shown that the fixed-smoothing critical values are second order correct under the increasing-smoothing asymptotics. In contrast, increasing-smoothing critical values are not even first order correct under the fixed-smoothing asymptotics. Given this, the fixed-smoothing approximation can be regarded as more robust than the increasing-smoothing approximation.

The same comment applies to the local-to-unity parameter c_m. When we use the near-unity fixed-smoothing approximation, we do not have to literally fix c_m at a given value in finite samples. Whether we hold c_m fixed or let it increase with the sample size can be viewed as different asymptotic specifications to obtain approximations to the same finite sample distribution. In practice, we can estimate c_m even though a consistent estimator is not available. For a stationary AR(1) process with a fixed autoregressive coefficient, the estimator $\hat{c}_m$ derived from the OLS estimator of $\hat{\rho}_{T,m}$ necessarily converges to infinity in probability. The critical values from the near-unity fixed-smoothing asymptotic distribution are thus close to those from the stationary fixed-smoothing asymptotic distribution. So the near-unity fixed-smoothing approximation is still asymptotically valid. For this reason, we can say that the near-unity fixed-smoothing approximation is a more robust approximation. Compared to the chi-square or normal approximation, the near-unity fixed-smoothing approximation achieves double robustness — it is asymptotically valid regardless of the limiting behaviors of c_m and h.

3 Some Simulation Evidence

To implement the near-unity fixed-smoothing approximation, we need to pin down the value of c_m, which cannot be consistently estimated. However, a nontrivial and informative confidence interval (CI) can still be constructed. I propose to construct a CI for c_m and use the maximum of the critical values, each of which corresponds to one value of c_m in the CI. An argument based on the Bonferroni bound can be used to determine the confidence level of the CI and the significance level of the critical values. More specifically, for tests with nominal level α, we could employ the critical value defined by

$$CV = \sup_{c_m \in CI_{1-\alpha+\delta}} CV(c_m, 1-\delta)$$

where $\delta \leq \alpha$, $CI_{1-\alpha+\delta}$ is a lower confidence interval for c_m with nominal coverage probability $1-\alpha+\delta$, and $CV(c_m, 1-\delta)$ is the $100\,(1-\delta)\,\%$ quantile from the distribution $F_\infty\,(c_m, h)$ or $t_\infty\,(c_m, h)$. This idea of choosing critical values in the presence of unidentified nuisance parameters has been used in various settings in statistics and econometrics. See for example McCloskey (2012) and references therein.

One drawback of the approach based on the Bonferroni correction is that the critical value is often too large and the resulting test often under-rejects. There are sophisticated ways to improve on the Bonferroni method. As a convenient empirical strategy, here I

employ $CV = \sup_{c_m \in CI_{90\%}} CV(c_m, 95\%)$ for nominal 5% tests. I construct the CI for c_m using the method of Andrews (1993). Other methods such as Stock (1991) and Hansen (1999) can also be used. See Mikusheva (2014) and Phillips (2014) for recent contributions on this matter. Since $CV(c_m, 95\%)$ is decreasing in c_m, we only need to find the lower limit of $CI_{90\%}$ in order to compute CV. The computational cost is very low.

I consider a univariate Gaussian location model with AR(2) error $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t$, where $u_t \sim iidN(0,1)$. The sample size is 200. The initial value of the error process is set to be standard normal. I generate a time series of length 400 and drop the first 200 observations to minimize the initialization effect. This is similar to generating a time series with 200 observations but with the initial value drawn from its stationary distribution. I consider Müller's test, the KVB test, and the test based the series variance estimator with the basis functions: $\phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, $\phi_{2j}(x) = \sqrt{2}\sin(2j\pi x), j = 1, \ldots, K/2$. The values of $K = 12, 24, 48$ correspond to the values of $q = 12, 24, 48$ in Müller's paper. The number of simulation replications is 20,000.

Table 1 reports the null rejection probabilities for various two-sided 5% tests. It is clear that the tests based on the near-unity fixed-smoothing approximation are in general more accurate than those based on the usual stationary fixed-smoothing approximation. This is especially true when the process is highly autocorrelated. In term of size accuracy, Müller' test is slightly better than the near-unity fixed-smoothing test. The size accuracy of the latter test is actually quite satisfactory. As I mentioned before, the AR(2) process with $(\rho_1, \rho_2) = (1.9, -.95)$ posts a challenge to all tests considered.

Figure 1 plots the size-adjusted power against the noncentrality parameter δ^2 in the presence AR(1) errors. The figure is representative of other configurations. It is clear from the figure that the slightly better size control of Müller's tests is achieved at the cost of some power loss.

Table 1: Empirical null rejection probability of nominal 5% tests with T = 200 under AR(2) errors

	Stationary Fixed-Smoothing				Near-Unity Fixed-Smoothing				Nearly-Optimal Tests		
(ρ_1, ρ_2)	K_{12}	K_{24}	K_{48}	KVB	K_{12}	K_{24}	K_{48}	KVB	S_{12}	S_{24}	S_{48}
(0, 0)	0.048	0.047	0.048	0.047	0.045	0.042	0.041	0.042	0.045	0.047	0.047
(.7,0)	0.058	0.083	0.148	0.058	0.040	0.035	0.032	0.040	0.046	0.047	0.047
(.9,0)	0.133	0.248	0.393	0.084	0.036	0.033	0.033	0.038	0.048	0.047	0.050
(.95,0)	0.258	0.412	0.553	0.125	0.039	0.037	0.036	0.039	0.048	0.049	0.050
(0.99,0)	0.630	0.738	0.816	0.333	0.079	0.079	0.079	0.071	0.045	0.044	0.045
(1.9,-.95)	0.005	0.002	0.015	0.015	0.000	0.000	0.000	0.000	0.025	0.001	0.000
(.8,.1)	0.146	0.272	0.417	0.089	0.050	0.051	0.052	0.045	0.049	0.048	0.052

4 Conclusion

Müller's paper makes an important contribution to the literature on HAR inference. It has the potential for developing a standard of practice for HAR inference when the process

is strongly autocorrelated. The paper inspires us to think more about optimality issues in hypothesis testing. Unfortunately, uniformly optimal tests do not exist except in some special cases. This opens the door to a wide range of competing test procedures. In this discussion, I have outlined an alternative test, which is based on the standard test statistic but employs a new asymptotic approximation. The alternative test has satisfactory size but is not as accurate as Müller's test. However, Müller's test is often less powerful. The trade-off between size accuracy and power improvement is unavoidable. A prewhitening testing procedure with good size property may also be crafted. HAR testing is fundamentally a nonparametric problem. A good test requires some prior knowledge about the data generating process. In the present setting, the prior knowledge should include the range of the largest AR root and the neighborhood around origin in which the spectral density remains more or less flat. Equipped with this knowledge, a practitioner can select a testing procedure to minimize their loss function.

References

[1] Andrews, D.W.K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica,* 61(1), 139–165.

[2] Hansen, B.E. (1999): "The Grid Bootstrap and the Autoregressive Model." *Review of Economics and Statistics*, 81(4), 594–607.

[3] Cox, D.R. and D.V. Hinkley (1974): *Theoretical Statistics*, Chapman and Hall: New York.

[4] Elliott, G., U.K. Müller and M.W. Watson (2012): "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis." Working paper, Department of Economics, UC San Diego and Princeton University.

[5] Müller, U.K. (2007): "A Theory of Robust Long-Run Variance Estimation." *Journal of Econometrics* 141, 1331–1352.

[6] McCloskey, A. (2012): "Bonferroni-Based Size-Correction for Nonstandard Testing Problems." Working paper, Department of Economics, Brown University.

[7] Mikusheva, A. (2014): "Second Order Expansion of the t-statistic in AR(1) Models." *Econometric Theory,* forthcoming.

[8] Phillips, P.C.B. (2005): "HAC Estimation by Automated Regression." *Econometric Theory* 21, 116–142.

[9] Phillips, P.C.B. (2014): "On Confidence Intervals for Autoregressive Roots and Predictive Regression." *Econometrica,* forthcoming.

[10] Phillips, P.C.B., T. Magdalinos and L. Giraitis (2010): "Smoothing Local-to-moderate Unit Root Theory." *Journal of Econometrics* 158, 274–279.

[11] Stoica, P. and R. Moses (2005): *Spectral Analysis of Signals*, Pearson Prentice Hall.

[12] Stock, J. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series." *Journal of Monetary Economics* 28, 435–459.

[13] Sun, Y. (2006): "Best Quadratic Unbiased Estimators of Integrated Variance in the Presence of Market Microstructure Noise." Working paper, Department of Economics, UC San Diego.

[14] Sun, Y. (2011): "Autocorrelation Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter." *Journal of Econometrics* 164, 345–366.

[15] Sun, Y. (2013): "A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." *Econometrics Journal* 16, 1–26.

[16] Sun, Y. (2014a): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference." *Journal of Econometrics*, 178(3), 659–677

[17] Sun, Y. (2014b): "Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation." Working paper, Department of Economics, UC San Diego.

[18] Thomson, D. J. (1982): "Spectrum Estimation and Harmonic Analysis." *IEEE Proceedings*, 70, 1055–1096.

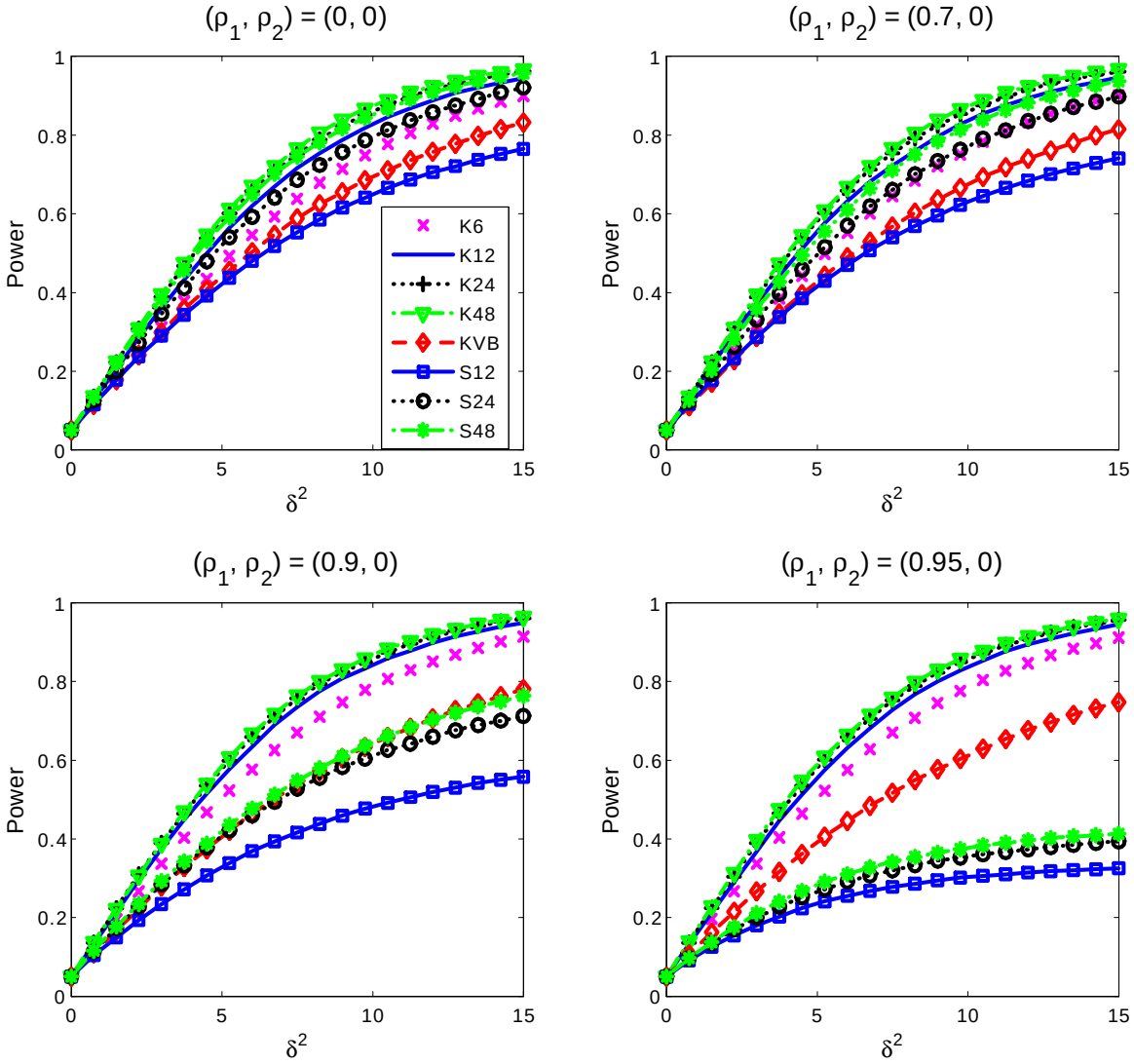

Figure 1: Size-adjusted Power of Different 5% Tests with sample size $T = 200$ ("K6", "K12", "K24" , "K48", and "KVB" are the Near-Unity Fixed-Smoothing tests while S12, S24 and S48 are Müller's tests)