Accurately Sized Test Statistics with Misspecified Conditional Homoskedasticity

Douglas G. Steigerwald *†
Economics Department
University of California, Santa Barbara
E-mail: doug@econ.ucsb.edu

Jack Erb
Economics Department
University of California, Santa Barbara
E-mail: erb@econ.ucsb.edu

July 2007

Abstract

We study the problem of obtaining accurately sized test statistics in finite samples for linear regression models where the error dependence is of unknown form. With an unknown dependence structure there is traditionally a trade-off between the maximum lag over which the correlation is estimated (the bandwidth) and the decision to introduce conditional heteroskedasticity. In consequence, the correlation at far lags is generally omitted and the resultant inflation of the empirical size of test statistics has long been recognized. To allow for correlation at far lags we study test statistics constructed under the possibly misspecified assumption of conditional homoskedasticity. To improve the accuracy of the test statistics, we employ the second-order asymptotic refinement in Rothenberg (1988) to determine critical values. We find substantial size improvements resulting from the

*Corresponding author

†We thank the members of the Econometrics Lab at Santa Barbara for helpful comments.

second-order theory across a wide range of specifications, including substantial conditional heteroskedasticity. We also find that the size gains result in only moderate increases in the length of the associated confidence interval, which yields an increase in size-adjusted power. Finally, we note that the proposed test statistics do not require that the researcher specify the bandwidth or the kernel.

Key Words: Test size, confidence interval estimation, heteroskedasticity, autocorrelation

Subject Classification: C1, C13, C14

1 Introduction

When forming test statistics for coefficients in linear regression models, it has become widely accepted to use the Newey-West covariance estimator to account for error serial correlation. The appeal of the Newey-West method (introduced in 1987) is that it allows for conditional heteroskedasticity, although at the cost of only admitting serial correlation at near lags. The inability to account for serial correlation at far lags leads to test statistics with empirical sizes that far exceed nominal sizes. To address this problem, Kiefer and Vogelsang (2005) refine the asymptotic theory to more accurately model the admission of serial correlation at far lags. As Kiefer and Vogelsang show that the resultant non-Gaussian critical values increase with the admitted lag length, the desire to accommodate both conditional heteroskedasticity and correlation at far lags carries its own cost of considerably lengthening confidence intervals. In an effort to reduce this cost, we reexamine the relative merits of allowing for conditional heteroskedasticity and for serial correlation at far lags. To do so, we compare the performance of test statistics that allow for conditional heteroskedasticity with test statistics constructed under the (possibly) misspecified assumption of conditional homoskedasticity.

Driven by the desire to allow for general dependence in economic time series, White and Domowitz (1984) develop a consistent standard error estimator under conditional heteroskedasticity. The key condition is that the maximum lag over which the serial correlation is estimated, the bandwidth, is an asymptotically negligible fraction of the sample size. Although the

White-Domowitz estimator is consistent, it is not guaranteed to be positive definite. In response, Newey and West demonstrate that the introduction of a kernel that downweights correlations as the lag length grows, ensures that the consistent standard error estimator is also positive semi-definite.

With every solution there comes another problem. While the Newey-West estimator is consistent and positive semi-definite, the estimated standard errors are often too small. As Andrews (1991) demonstrates, if the errors exhibit substantive temporal dependence, then test statistics formed from the Newey-West standard error estimates have empirical size far in excess of nominal size (test statistics that reject too often). To reduce the nominal size, Andrews and Monahan (1992) propose a two-step method, in which the first step consists of prewhitening the residuals by fitting a low-dimension process (such as a VAR(1)) to capture serial correlation at far lags. In the second step, the conditional heteroskedasticity is estimated at near lags.

Prewhitening the residuals prior to estimating conditional heteroskedasticity at near lags goes part way to resolving the problem of high nominal size. In an effort to make further improvements, Kiefer and Vogelsang suggest forgoing the first step prewhitening and estimating the conditional heteroskedasticity directly at both near and far lags. When including correlation at far lags, it is no longer tenable to assume that the bandwidth is an asymptotically negligible fraction of the sample size. In consequence, the (first-order) asymptotic distribution of resultant test statistics is not Gaussian. The alternative asymptotic distribution delivers simulated critical values that are considerably larger than their Gaussian counterparts. If only serial correlation at near lags is admitted, then the refined asymptotic critical values deliver size improvements in line with the improvements obtained by prewhitening. If serial correlation at all lags is admitted (note that the theory does not deliver an optimal bandwidth), then there are substantial further reductions in empirical size.

A key insight in previous research is the need to account for serial correlation at far lags to obtain more accurate coverage probabilities. Current methods to account for correlation at far lags are either completely general, as in Kiefer and Vogelsang, or specific, as in Andrews and Monahan. While Kiefer and Vogelsang allow for conditional heteroskedasticity of unknown form, the cost is longer confidence intervals resulting in loss of power for associated test statistics. The low-dimension parametric method of Andrews and Monahan produces confidence intervals of more moderate length, but still suffers from high empirical size. It is therefore of interest to study methods that lie between the two, to adjudge the trade-off between size and power.

In contrast to Kiefer and Vogelsang, we propose broadening the first step of Andrews-Monahan by fitting a high-dimension process to capture serial correlation at all lags, while forgoing the second-step conditional heteroskedasticity estimation. The assumption underlying the method is that standard errors can be well approximated by a conditionally homoskedastic covariance matrix that is band diagonal. The band diagonals are not restricted to be related through a low-dimension process. To obtain size improvements we too rely on asymptotic refinements, namely the second-order theory of Rothenberg (1988). The second-order theory yields critical values that adjust to incorporate the behavior of the regressors. As Rothenberg establishes, the bandwidth need not be an asymptotically negligible fraction of the sample size, so all correlation lags are included under conditional homoskedasticity. Further, the estimator is positive semi-definite by construction without need of a kernel.

We study the size and size-adjusted power of test statistics constructed under the three methods. We focus not only on hypothesis tests of a single parameter, but also on tests of multiple parameters to determine the impact of off-diagonal elements of the estimated covariance matrix. In Section 2 we present the quantities of interest and the models to be simulated. The conditional heteroskedasticity specifications allow us to investigate an additional observation in Rothenberg: Namely, that the degree of correlation between the regressor under test and the conditional heteroskedasticity plays a key role in the empirical size. We examine the range of models typically used to assess the performance of confidence intervals under conditional heteroskedasticity. Results from the simulations are contained in Section 3. We provide an empirical application to the estimation techniques in Section 4. A brief summary of developments in standard error estimation is contained in Appendix A.

2 Covariance Estimators

To determine the finite sample size and size-adjusted power of hypothesis tests constructed under the (potentially misspecified) assumption of conditional homoskedasticity we employ a simulation model. Our simulation model is

$$Y_t = X_t'\beta + U_t \quad t = 1, \ldots, n, \tag{1}$$

where X_t contains a constant and four regressors (to allow for comparison with the findings in both Andrews as well as Andrews and Monahan). Conditional heteroskedasticity is introduced through a scale parameter that depends equally on each of the varying regressors

$$U_t = |X_t'\zeta| \times \tilde{U}_t,$$

where $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$ and $\left\{\tilde{U}_t\right\}$ is a sequence of possibly dependent random variables defined below. This specification of conditional heteroskedasticity is also employed by Andrews (as well as Andrews and Monahan) to demonstrate the superior performance of estimators that incorporate conditional heteroskedasticity over the more traditional parametric covariance estimators.

The relative magnitude of conditional heteroskedasticity present in the model is controlled through the degree of serial correlation in the regressors and error. To capture serial correlation, the regressors and error are generated for each $t = 1, \ldots, n$ (and each $k = 2, \ldots, 5$) as

$$\begin{aligned} \tilde{U}_t &= \rho_U \tilde{U}_{t-1} + \varepsilon_t, \\ X_{kt} &= \rho_X X_{kt-1} + \eta_{kt}. \end{aligned}$$

We also consider serial correlation of limited duration, under which conditional heteroskedasticity plays a correspondingly larger role, through the moving-average specification

$$\begin{aligned} \tilde{U}_t &= \varepsilon_t + \theta_U \varepsilon_{t-1}, \\ X_{kt} &= \eta_{kt} + \theta_X \eta_{kt-1}. \end{aligned}$$

The underlying errors, ε_t and η_{kt}, are mutually independent $N(0,1)$ random variables.[1] The serial correlation parameters (ρ_U, ρ_X) and (θ_U, θ_X) take values in the set $\Lambda = \{0, .1, .3, .5, .7, .9\}$.

Rothenberg derives the second-order asymptotic distribution of test statistics on coefficients of a linear regression under the assumption of conditional homoskedasticity. He finds (on page 1011) that as the correlation

[1] We set ε_0 and η_{k0} equal to 0 in the simulations and discard the first 50 observations to remove any influence from initial values.

between the regressor under test and the scale parameter increases, the inflation of the empirical size of the test statistics increases. To determine the impact of this correlation, we employ a second conditional heteroskedasticity specification in which the scale parameter depends on only one of the varying regressors. To isolate the impact noted by Rothenberg, we consider two values for the parameter ζ, specifically $\zeta = (0, 1, 0, 0, 0)'$ and $\zeta = (0, 0, 1, 0, 0)'$, while testing the hypothesis that the first varying regressor is equal to zero.[2,3] The first specification provides the highest level of correlation between the regressor under test and the error scale, while in the second specification the correlation between the regressor and scale is zero.

We focus not only on hypothesis tests of a single coefficient, but also on tests of multiple coefficients. Our hypothesis tests of multiple coefficients are designed to assess the effect of including off-diagonal elements of the covariance matrix. To capture this effect, we consider the single restriction imposed by the hypothesis $H_0 : \beta_2 - \beta_3 = 0$. In the multiple coefficient tests we consider values of the heteroskedasticity parameter, ζ, equal to $\left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$, $(0, 1, 0, 0, 0)'$, and $(0, 0, 0, 1, 0)'$.

To construct test statistics for hypotheses concerning β, an estimator of the (conditional) variance of the ordinary least-squares estimator, B, is needed. The variance of the (ordinary) least-squares estimator B, conditional on $X = (X_1, \ldots, X_n)'$, is

$$
\begin{aligned}
& Var\left(n^{\frac{1}{2}}(B-\beta)\,|X\right) \\
= & \left(n^{-1}\sum_{t=1}^{n} X_t X_t'\right)^{-1} n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n} E\left(U_s X_s U_t X_t' | X\right)\left(n^{-1}\sum_{t=1}^{n} X_t X_t'\right)^{-1}.
\end{aligned}
$$

The key component for estimation is $J = n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n} E\left(U_s X_s U_t X_t'|X\right)$.

We consider five estimators of J. The first is the classic OLS variance

[2] The values of ζ have been chosen to ensure that the unconditional variance of U_t is the same in all specifications. Andrews and Monahan also study this specification, albeit without reference to the findings in Rothenberg.

[3] To see that this model brings conditional heteroskedasticity, consider the case in which $\tilde{U}_t$ follows an MA(1) process. Then

$$
E\left(U_t U_{t-1}\right) = E\left[\left(|X_{2t}|\,\varepsilon_t + \theta_U\,|X_{2t}|\,\varepsilon_{t-1}\right)\left(|X_{2t-1}|\,\varepsilon_{t-1} + \theta_U\,|X_{2t-1}|\,\varepsilon_{t-2}\right)\right],
$$

and the covariance conditional on X is $\theta_U\,|X_{2t}|\,|X_{2t-1}|$.

estimator, which is consistent if the errors are i.i.d., and is

$$\hat{J}_{iid} = \left(\frac{1}{n-5} \sum_{t=1}^{n} \hat{U}_t^2 \right) \left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right),$$

where $\left\{ \hat{U}_t \right\}_{t=1}^{n}$ is the OLS residual vector.[4] The second estimator allows for serial correlation, but no conditional heterogeneity, and is consistent under the assumption that the error is generated by a parametric (AR(1)) process

$$\hat{J}_{par} = \left(\frac{1}{n-5} \sum_{t=1}^{n} \hat{U}_t^2 \right) \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \hat{\rho}^{|t-s|} X_t X_s' \right),$$

where $\hat{\rho}$ is obtained from regression of residuals on lagged residuals. The third estimator is the heteroskedasticity-autocorrelation consistent estimator introduced by Newey and West

$$\hat{J}_{hac} = \frac{1}{n-5} \left[\sum_{t=1}^{n} \hat{U}_t^2 X_t X_t' + \sum_{j=1}^{m} \left(1 - \frac{j}{m+1} \right) \sum_{t=j+1}^{n} \hat{U}_t \hat{U}_{t-j} \left(X_t X_{t-j}' + X_{t-j} X_t' \right) \right].$$

The value of m determines the maximum lag length at which the conditional heteroskedasticity is estimated. As m controls the number of far lags that enter the estimator, sample-based selection of m is very important in controlling the size of test statistics. The value of the bandwidth is allowed to vary across simulations and is chosen according to the automatic selection procedure developed by Andrews.

We also focus on a variant of the heteroskedasticity-autocorrelation consistent estimator, discussed in Andrews and Monahan. To reduce the bias in $\hat{J}_{hac}$, Andrews and Monahan advocate separating the variance estimation into three steps. First, estimate the temporal correlation (at far lags) by fitting a vector autoregression to $\left\{ \hat{V}_t \right\}$ (where $\hat{V}_t = \hat{U}_t X_t$), which yields the prewhitened residuals $\left\{ \tilde{V}_t \right\}$. (In our implementation, we fit a vector autoregression of order 1.) Second, construct the variance estimator with the prewhitened residuals

$$\tilde{J}_{pw} = \frac{1}{n-5} \left[\sum_{t=1}^{n} \tilde{V}_t \tilde{V}_t' + \sum_{j=1}^{m} \left(1 - \frac{j}{m+1} \right) \sum_{t=j+1}^{n} \left(\tilde{V}_t \tilde{V}_{t-j}' + \tilde{V}_{t-j} \tilde{V}_t' \right) \right].$$

[4] We use $n-5$, rather than n, as the divisor because the degrees-of-freedom calculation is likely to be used when $n = 50$, as it does in the simulations.

The last step is to recolor the estimator $\tilde{J}_{pw}$ to obtain the pre-whitened variance estimator, $\hat{J}_{pw}$, according to

$$\hat{J}_{pw} = \hat{C}\tilde{J}_{pw}\hat{C}', \qquad \text{where} \qquad \hat{C} = \left(I - \sum_{s=1}^{p} \hat{A}_s\right)^{-1}.$$

Here, $\left\{\hat{A}_s\right\}_{s=1}^{p}$ are the estimated coefficient matrices from a p^{th}-order vector autoregression of V_t.[5]

With the reduction in correlation brought about by the first step prewhitening, there is less need to select a large value of m.[6] Andrews and Monahan find that use of the prewhitened residuals reduces the (downward) bias of $\hat{J}_{hac}$ although the variance of $\hat{J}_{pw}$ exceeds the variance of $\hat{J}_{hac}$. The downward bias in the estimated standard errors is reduced to such an extent that, despite a loss of precision in estimating the standard errors, the coverage probabilities of confidence intervals are increased.

Constructing accurately sized tests with each of the above estimators of J remains a problem. Because the parametric prewhitening in $\hat{J}_{pw}$ does improve the size of test statistics, it may be the case that by increasing the richness of the first step, in which temporal correlation is accounted for, it is possible to forego the second step, in which conditional heteroskedasticity is accounted for. Rather than assume a low-dimension parametric model for temporal correlation, one could assume that the errors are generated by a conditionally homoskedastic (stationary stochastic) process with nothing further known about the autocorrelation function, under which

$$n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n} E\left(U_s X_s U_t X_t' | X\right) = n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n} \delta_{|t-s|} X_s X_t',$$

where $\delta_{|t-s|} = E\left(U_s U_t\right)$ depends only on $|t-s|$. The fourth estimator of J, which is consistent if the errors are conditionally homoskedastic, is

$$\hat{J}_{cho} = \frac{1}{n-5}\sum_{s=1}^{n}\sum_{t=1}^{n} \hat{\delta}_{|t-s|} X_s X_t',$$

[5] To ensure the matrix $I - \sum_s A_s$ is not too close to singularity, we restrict the eigenvalues of $\sum_s \hat{A}_s$ to be no larger than 0.97 in absolute value. See Andrews and Monahan (1992, pg. 957) for the details.

[6] This estimator, which includes pre-whitening, retains many of the asymptotic properties of $\hat{J}_{hac}$ including the rate of convergence.

where (for $t > s$) $\hat{\delta}_{|t-s|} = \frac{1}{n}\sum_{t=s+1}^{n}\hat{U}_t\hat{U}_{t-s}$. (If $s > t$, simply switch the values of t and s in the formula.) For this estimator we present two sets of critical values, those corresponding to the standard Gaussian limit theory and those corresponding to the second-order asymptotic refinement in Rothenberg.[7]

While all of the estimators of J lead to Gaussian limit distributions in testing situations, extensions to the asymptotic theory are available for two of the estimators. For $\hat{J}_{hac}$, Kiefer and Vogelsang have developed an alternative limit theory based on the assumption that the fraction of lags that appear in the estimator, $\frac{m}{n}$, is not asymptotically negligible.[8] The critical values that arise from the alternative limit theory can be considerably larger than the standard Gaussian critical values. As these critical values depend on m, we report results for two sets of critical values. The first uses the Andrews automatic bandwidth procedure to compute the bandwidth and the second sets the bandwidth equal to the sample size ($m = n$).[9]

For $\hat{J}_{cho}$, Rothenberg provides critical values based on a higher-order asymptotic refinement. If cv denotes the critical value from the first-order Gaussian approximation, then Rothenberg's second-order theory delivers the adjusted critical value

$$cv^R = cv\left(1 + \frac{1}{n}f\left(X, \hat{U}\right)\right).$$[10]

His asymptotic refinements indicate that Gaussian critical values should generally be increased (as f is generally greater than 0), although the precise form of his covariance estimator differs slightly from $\hat{J}_{cho}$.[11] Because the adjusted critical value is a function of $\left(X, \hat{U}\right)$, the adjusted critical value is correlated with the estimated standard error. If this correlation is negative,

[7] Our finite-sample results are designed to guide researchers with moderate sample sizes, in which size inflation is known to be a problem. While $\hat{J}_{cho}$ is not a consistent estimator of J under conditional heteroskedasticity, a consistent estimator is easily obtained by switching from $\hat{J}_{cho}$ to $\hat{J}_{hac}$ with first-order Gaussian critical values for large sample sizes (say $n > 5,000$).

[8] A similar limit theory is developed by Phillips, Sun and Jin (2006).

[9] The asymptotic theory in Kiefer and Vogelsang assumes that the bandwidth is selected without use of the automatic procedure of Andrews.

[10] The precise form of $f()$ is detailed in Appendix B.

[11] Rothenberg considers covariance estimators of the form $\frac{1}{n-j}\sum\hat{U}_t\hat{U}_{t-j}$, where j corrects for the number of observations lost due to the lag length. To ensure a positive semi-definite estimator we replace the factor $\frac{1}{n-j}$ with $\frac{1}{n}$.

then the critical value adjusts to the magnitude of the standard error and lessens the length of estimated confidence intervals. Such an adjustment feature can lead to a test statistic with large gains in size at the cost of only small declines in size-adjusted power.

3 Simulation Results

For the simulations, we construct Y_t according to (1) with $\beta = \mathbf{0}$. If we let c be a 5×1 vector of constants that selects the parameters under test, we construct the test statistic for the hypothesis $H_0 : c'\beta = 0$ according to

$$t = [c' V_B c]^{-1/2} \cdot \sqrt{n}\, c' B,$$

where B is the OLS estimate of β and V_B is the 5×5 sample analog of $Var\left(n^{\frac{1}{2}} (B - \beta) |X\right)$. ($V_B$ is constructed for each of the variance estimators in Section 2.) The selected critical values are for a two-sided test with five percent nominal size. The sample size is $n = 50$ in all models to allow for direct comparison with the simulation results presented in recent papers, such as Kiefer and Vogelsang (2005) and Phillips, Sun and Jin (2006). Each experiment consists of 50,000 replications. While there are a variety of statistics that can be used to assess the finite-sample performance of the variance estimators, we follow the convention of more recent authors and focus our attention on the finite-sample size and size-adjusted power of test statistics.

It is well known that with any hypothesis test, there is a trade off between size and power. Because each test employs the OLS estimator as a point estimate, improvements in size will generally be accompanied by a corresponding decrease in power. However, each test varies in either the variance estimate, critical value, or both. Consequently, there is the possibility that a particular estimator may display more accuracy in terms of both improved test size and higher power against alternatives.

Similar to previous authors, we have chosen to study the performance of test statistics where the underlying models of serial correlation are parametric in nature. We note that this may place the $\hat{J}_{hac}$ and $\hat{J}_{cho}$ estimators at a slight disadvantage as they do not restrict the relationship among the covariances to a specific functional form.

3.1 Single Parameter Tests

We first study $H_0 : c'\beta = 0$, with $c = (0, 1, 0, 0, 0)'$, and so test whether the coefficient on the first non-constant regressor is significantly different from zero. This choice of c leads to a standard error estimate generated only from the diagonal elements of V_B. In the tables that follow, the test statistics are referenced by the covariance matrix estimator the test employs. Thus, *hac* denotes the t-statistic when V_B is constructed according to $\hat{J}_{hac}$ and evaluated with standard asymptotic critical values. Critical values from asymptotic refinements are indicated by superscripts, so cho^R is the t-statistic constructed with $\hat{J}_{cho}$ and evaluated with Rothenberg's second-order critical values. In similar fashion, hac^{KV} indicates use of $\hat{J}_{hac}$ with the Kiefer-Vogelsang asymptotic approximation to generate test critical values and a bandwidth determined by the Andrews automatic selection procedure. As the Kiefer-Vogelsang approximation allows the bandwidth to equal the sample size, we denote this statistic as hac^{KVn}.[12]

Table 1 reports the finite-sample empirical size of each test statistic when the AR(1) errors are overlaid with multiplicative heteroskedasticity entering from all four non-constant regressors, i.e. $\zeta = (0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})'$.[13] As with previous authors, we find that the traditional *hac* test performs quite poorly in terms of test size, especially when the dependence in the data is strong. Indeed, when $\rho_X = \rho_U = 0.9$, the *hac* test rejects the null hypothesis 38% of the time. This is quite unsettling considering the applied researcher will be making inference based on a nominal size of 5%.

As the test statistics all employ the same point estimate in the numerator, improvements in test size will be achieved by either increasing the standard error in the denominator or widening the test critical values. In column 2, we see that prewhitening residuals reduces test size by inflating the estimated standard errors, although over rejection of the null remains a problem. In the $\rho_X = \rho_U = 0.9$ case, the empirical size drops to 0.32, a 16% size gain. However, attempting to remove correlation at far lags by prewhitening when serial correlation in the data is weak can inflate test size even more than using the HAC estimator, which can be seen by comparing columns 1 and 2 when either ρ_X or ρ_U is less than 0.3. Column 3 shows that use of the Kiefer-Vogelsang asymptotic refinements reduces size inflation by widening the test

[12] Empirical sizes for tests constructed using $\hat{J}_{iid}$ and $\hat{J}_{par}$ are presented in Appendix C.

[13] As the serial correlation in the regressor mirrors the serial correlation in the error in each simulation model, the regressors are AR(1) in Table 1.

Table 1: **Empirical Size - Heteroskedastic AR(1) Errors** - $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$

		(1)	(2)	(3)	(4)	(5)	(6)
ρ_X	ρ_U	hac	pw	hac^{KV}	hac^{KVn}	cho	cho^R
0.9	0.9	0.3808	0.3235	0.3171	0.2332	0.3192	0.2367
	0.7	0.2834	0.2351	0.2327	0.1686	0.2371	0.1641
	0.5	0.2126	0.1803	0.1766	0.1322	0.1941	0.1306
	0.3	0.1655	0.1501	0.1391	0.1102	0.1668	0.1076
	0.1	0.1310	0.1308	0.1088	0.0934	0.1479	0.0906
	0.0	0.1213	0.1254	0.0987	0.0892	0.1403	0.0809
0.7	0.9	0.2859	0.2460	0.2393	0.1680	0.2481	0.1854
	0.7	0.2352	0.2025	0.1970	0.1414	0.2087	0.1508
	0.5	0.1933	0.1726	0.1645	0.1217	0.1844	0.1326
	0.3	0.1554	0.1469	0.1324	0.1034	0.1591	0.1136
	0.1	0.1327	0.1332	0.1108	0.0903	0.1440	0.1018
	0.0	0.1209	0.1260	0.1007	0.0886	0.1345	0.0929
0.5	0.9	0.2108	0.1880	0.1745	0.1208	0.1948	0.1493
	0.7	0.1836	0.1669	0.1557	0.1080	0.1711	0.1295
	0.5	0.1621	0.1519	0.1385	0.1019	0.1569	0.1204
	0.3	0.1407	0.1382	0.1214	0.0927	0.1434	0.1077
	0.1	0.1226	0.1275	0.1054	0.0831	0.1342	0.1008
	0.0	0.1159	0.1244	0.0994	0.0819	0.1285	0.0963
0.3	0.9	0.1592	0.1492	0.1355	0.0949	0.1579	0.1273
	0.7	0.1473	0.1419	0.1261	0.0924	0.1471	0.1153
	0.5	0.1367	0.1353	0.1194	0.0871	0.1409	0.1100
	0.3	0.1239	0.1277	0.1080	0.0819	0.1281	0.0994
	0.1	0.1167	0.1232	0.1024	0.0793	0.1282	0.0996
	0.0	0.1142	0.1240	0.0996	0.0783	0.1237	0.0967
0.1	0.9	0.1215	0.1244	0.1046	0.0749	0.1295	0.1073
	0.7	0.1187	0.1236	0.1024	0.0777	0.1285	0.1036
	0.5	0.1199	0.1277	0.1047	0.0796	0.1289	0.1031
	0.3	0.1138	0.1227	0.0995	0.0758	0.1239	0.0997
	0.1	0.1091	0.1180	0.0957	0.0745	0.1189	0.0937
	0.0	0.1094	0.1186	0.0954	0.0736	0.1195	0.0945
0.0	0.9	0.1105	0.1190	0.0948	0.0719	0.1225	0.1011
	0.7	0.1087	0.1181	0.0941	0.0728	0.1225	0.0995
	0.5	0.1089	0.1191	0.0949	0.0745	0.1203	0.0958
	0.3	0.1069	0.1178	0.0945	0.0747	0.1181	0.0949
	0.1	0.1078	0.1196	0.0947	0.0759	0.1200	0.0947
	0.0	0.1059	0.1172	0.0925	0.0729	0.1190	0.0939

critical values. The new limit theory leads to critical values that typically take on (absolute) values in the range of 2.0 to 4.81, depending on the chosen bandwidth for the HAC estimator. When the bandwidth is chosen according to the Andrews procedure, the size gains offered by the Kiefer-Vogelsang asymptotics are slightly larger than those achieved by prewhitening, but are typically comparable when the serial correlation is strong.

Similar to prewhitening, we find that $\hat{J}_{cho}$ delivers larger standard errors that improve the finite-sample size of the resulting test statistics, and similar to the Kiefer-Vogelsang asymptotics, we find that the second-order critical value refinement of Rothenberg further improves test size by increasing the critical values. Columns 5 and 6 of Table 1 present the empirical size of test statistics when $\hat{J}_{cho}$ is used in conjunction with either standard normal critical values or the second-order critical values of Rothenberg, respectively. Even under misspecification, the cho^R test delivers more substantial size reductions than both the *pw* and hac^{KV} tests. For $\rho_X = \rho_U = 0.9$, the empirical size of the cho^R test is 0.24. While the actual size remains significantly larger than the 5% nominal level, the cho^R test reduces size inflation by more than a third of the level of the *hac* test. Moreover, the improved accuracy of the cho^R test over the more conventional HAC tests continues to hold for very low levels of serial correlation. For $\rho_X = \rho_U = 0.1$, the second-order adjusted, homoskedastic estimator improves test size accuracy relative to the Newey-West estimator by about 14%, or a drop in raw size from 0.1091 to 0.0937.[14]

Recall that for a fixed value of ρ_X, the heteroskedastic component of the error becomes more pronounced as ρ_U decreases. That the conditionally homoskedastic estimator retains an advantage in test size is quite remarkable. The favorable performance of cho^R is due primarily to the adaptability of the critical value to the data generating process. The second-order theory for $\hat{J}_{cho}$ delivers critical values that, while typically larger than their Gaussian counterparts, adjust with the regressors and residuals in such a way that the critical value increases when the estimated standard error is small and decreases when the estimated standard error is large. This negative correlation between the standard error and critical value serves as a hedge in cases

[14] As the table makes clear, the conditionally homoskedastic variance estimator can only be recommended in conjunction with Rothenberg's second-order critical value adjustment when the data is heteroskedastic. While the *cho* test has better size than the *hac* test if the serial correlation is high, it rarely exhibits smaller size than the *pw* or hac^{KV} tests, and often performs more poorly than the *par* and *iid* tests.

Table 2: **Performance of Rothenberg's second-order-adjusted critical value under heteroskedasticity**

$\rho_X = 0.9$	(1)	(2)	(3)
ρ_U	*Mean*	*Variance*	*Corr. with* $\sqrt{c'V_Bc}$
0.9	2.37	0.04	-0.15
0.7	2.34	0.04	-0.22
0.5	2.32	0.05	-0.27
0.3	2.32	0.05	-0.32
0.1	2.33	0.08	-0.34
0.0	2.34	0.08	-0.37

where over-rejections of the null are most likely to occur. As can be seen from column 5 of Table 1, $\hat{J}_{cho}$ is indeed influenced by the heteroskedasticity as the *cho* test performs more poorly than the *hac* test when ρ_U drops below 0.3. However, when the $\hat{J}_{cho}$ estimator is used in conjunction with the second-order critical values, the cho^R test retains a size advantage over the *hac* and *pw* tests for all values of ρ_X and ρ_U, and a size advantage over hac^{KV} for all but the smallest values of ρ_X and ρ_U.[15]

Table 2 reports the mean and variance of the second-order critical value as well as its correlation with the estimated standard error when ρ_X is fixed at 0.9. While the mean of the critical value remains relatively constant as ρ_U decreases, the variation increases and the negative correlation with the estimated standard error becomes more pronounced. This critical value adjustment feature is the reason the conditionally homoskedastic estimator is able to maintain size improvements over the *hac*, hac^{KV}, and *pw* tests as the heteroskedasticity becomes more pronounced.

It is important to note that the asymptotic theory put forth by Kiefer and Vogelsang does not restrict the bandwidth to be small relative to the sample size in order for the testing procedure to be valid. When the bandwidth is set equal to the sample size, the considerable downward bias of the Newey-West estimator is offset by an adjusted critical value of 4.81, which is almost two and a half times the critical value of the standard normal distribution.

[15] Intuition would suggest that for a fixed value of ρ_U the performance of the *cho* and cho^R tests should deteriorate as ρ_X increases. However, ρ_X also adds to the overall pattern of serial correlation in the regressors as well as the errors. For this reason, the *cho* and cho^R tests show size gains over other HAC tests when ρ_U is large, even when ρ_X is large.

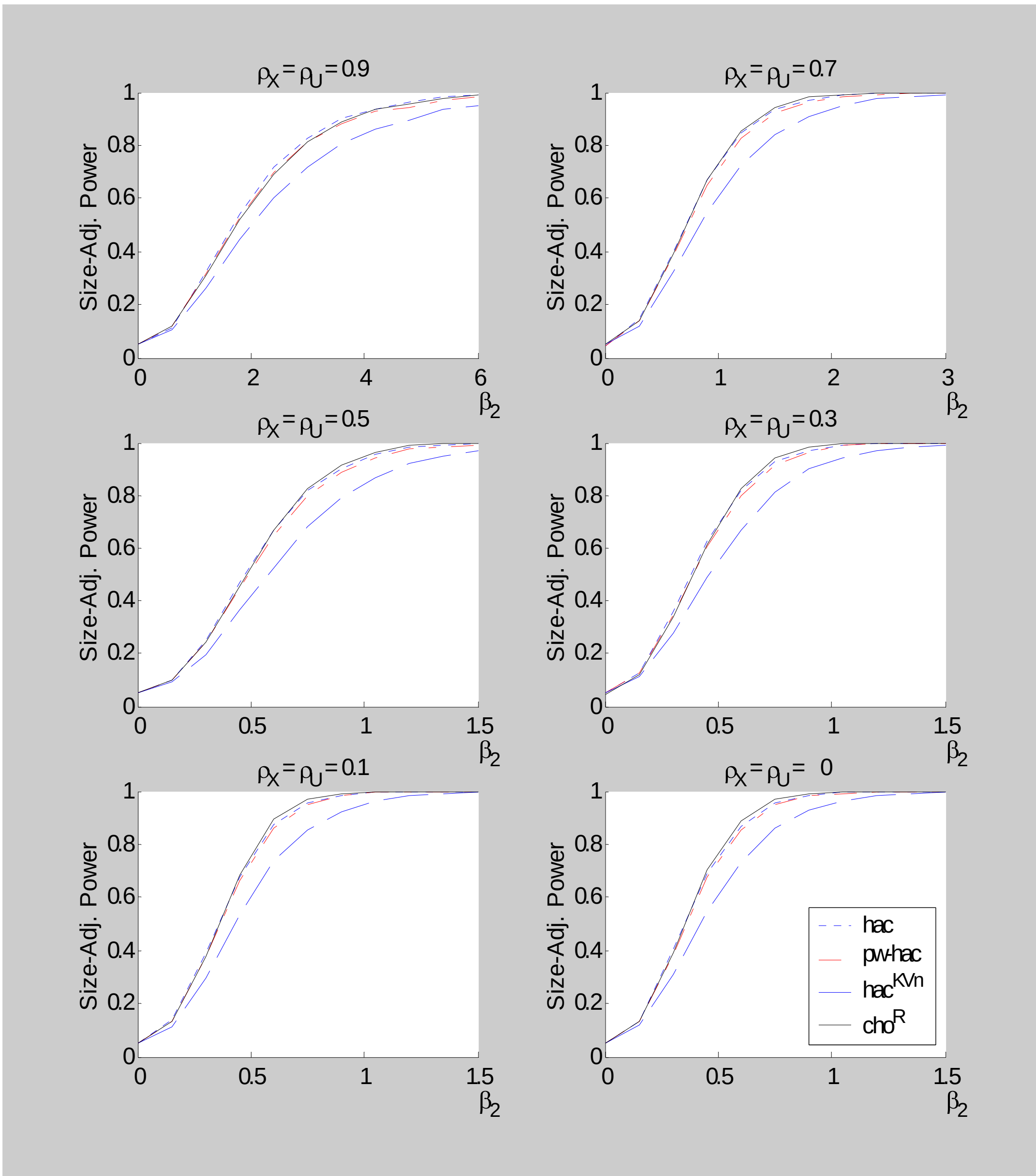
In comparison, the average critical value for the hac^{KV} test (as selected by the Andrews method) ranges from 2.04 when the temporal dependence is low to 2.26 when the dependence is high, and the average critical value for the cho^R test ranges from 2.11 when the dependence is low to 2.37 when the dependence is high. Column 4 in Table 1 shows that for more moderate levels of dependence, further reductions in test size are achieved with the Kiefer-Vogelsang critical values if the bandwidth is fixed and equal to the sample size, though the improvements are small. However, such drastic inflation of the critical value is sure to decrease the probability of rejecting the hypothesis for all values of β, and the slight size improvements of the hac^{KVn} test prove to be extremely costly in terms of test power.

Figure 1 plots the upper half of the size-adjusted power functions for the test statistics under the AR(1) specification presented in Table 1 when $\rho_X = \rho_U$. While size adjustment is not possible in actual testing situations, it is a useful measure in comparing the power of tests that do not have the same finite-sample size. We compute the size-adjusted power as the fraction of test rejections that arise when the true value of β_2 is different from 0. Specifically, we set $\beta = \psi \times (0, 1, 0, 0, 0)'$ where ψ is chosen as a set of eleven, equally spaced points from zero to some upper bound for which the estimated power for all tests is roughly one.[16] We simulate the power of test statistics using 10,000 replications for each value of ψ. The size-adjusted critical values are also computed via simulation methods using 50,000 replications. As each estimator gives rise to only one finite-sample distribution, the size-adjusted critical values and size-adjusted power curves for the hac and hac^{KV} tests are equivalent (as are those for the cho and cho^R tests).

The most striking feature of Figure 1 is the extent to which the power of the hac^{KVn} test lags the power of the other tests. When $\beta_2 = 2.4$ and the serial correlation parameters are equal to 0.9, the power of the hac test is 0.72, while the power of the cho^R test is 0.70. Recall the empirical sizes of the two tests were 0.38 and 0.24, respectively, indicating a 37% reduction in size is achieved with approximately a 3% decrease in power. However, the power of the hac^{KVn} test falls to 0.59 while the empirical size is 0.23. We see that the slight additional improvement in the size of the hac^{KVn} test comes at the cost of reducing the power by 16%. It is also clear from this figure that the cho^R test provides size-adjusted power which is quite similar

[16] For example, if the upper bound is set at 6, then the distance between each element of ψ is $\frac{6}{10}$, and $\psi = \{0,\ .6,\ 1.2,\ 1.8, \ldots, 6\}$.

Figure 1: **Heteroskedastic AR(1) Regressors and Errors** $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)' - H_0 : \beta_2 = 0$



to both the *hac* and *pw* tests, although there is some crossing.

In general, the power functions for other values of ρ_X and ρ_U are similar in shape and relative performance to those presented in Figure 1.[17] The HAC, pre-whitened, and conditionally homoskedastic estimators give rise to tests with very similar power, while the hac^{KVn} test exhibits power that is substantially lower across the range of alternatives.

Table 1 and Figure 1 illustrate that the performance of the hac^{KV} test depends heavily upon the choice of bandwidth in the estimation of the standard error. Andrews (among others) has also documented the ties between test performance and other user-choice parameters (including bandwidth, weighting kernel, and prewhitening model) for the *hac* and *pw* tests. For this reason, Andrews and others have developed data-driven "optimal" bandwidth selection procedures in an attempt to make the process more automatic. The Kiefer-Vogelsang asymptotics incorporate the choice of bandwidth directly into the testing problem by allowing the critical values to adjust with the bandwidth, effectively eliminating the need for the practitioner to find an "optimal" bandwidth. In practice, however, the researcher must still chose a bandwidth, and different choices may give rise to very different inferences on the parameter under test.[18]

To illustrate the problem that bandwidth selection could cause, Table 3 shows the probability that the hac^{KV} test rejects the null hypothesis for at least one bandwidth in a given replication. That is, for any given replication, we constructed a test statistic and critical value pair for each value of the bandwidth between 5 and $m = n = 50$.[19] If the null hypothesis was rejected for one or more of the test statistic/critical value pairs, the entire trial was considered as if it rejected the null hypothesis. We then repeated the process 50,000 times and found the fraction of replications which produced at least one rejection. The sizes of the hac^{KV} test in Table 3 are comparable in

[17] Note that the range of values along the β_2 axis differs for alternative values of the serial correlation parameters. All test are showing substantial increases in power as the correlation in the data falls.

[18] There is no data-dependent method of choosing an "optimal" bandwidth for the hac^{KV} test. Phillips, Sun and Jin propose a data dependent rule for their test that minimizes a weighted sum of type I and type II errors, which Kiefer and Vogelsang conjecture can be extended to their test. However, in place of selecting the bandwidth, the researcher is now left to choose the proper weights for the type I and II errors.

[19] To ensure the estimator accurately accounts for the serial correlation, we impose a minimum bandwidth of 5. Clearly, allowing for bandwidths less than 5 will further inflate the rejection probabilities.

Table 3: **Probability the hac^{KV} test rejects the null hypothesis for at least one value of the bandwidth**

AR(1) regressors and errors – $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$

	ρ_X					
ρ_U	0.9	0.7	0.5	0.3	0.1	0.0
0.9	0.3662	0.2788	0.2164	0.1724	0.1442	0.1336
0.7	0.2776	0.2310	0.2020	0.1654	0.1418	0.1300
0.5	0.2258	0.2030	0.1756	0.1482	0.1454	0.1256
0.3	0.1916	0.1752	0.1620	0.1448	0.1342	0.1174
0.1	0.1766	0.1542	0.1478	0.1344	0.1380	0.1284
0.0	0.1468	0.1458	0.1362	0.1250	0.1328	0.1412

magnitude to the sizes of the traditional *hac* test in column 1 of Table 1. Clearly, the failure to properly account for the pre-test estimation of the bandwidth results in further size inflation and negates the advantages of the hac^{KV} test over the more traditional tests.

The problem of nuisance parameters in the testing process highlights an important advantage of $\hat{J}_{cho}$ over other estimators. First, $\hat{J}_{cho}$ estimates the correlation at all lags, eliminating the necessity of choosing a particular bandwidth. Second, there is no weighting kernel or prewhitening filter involved in the estimation. And third, while the second-order critical value adjustment isn't trivial, it is completely data dependent and requires no choices by the user in implementation.

While the cho^R test performs favorably under the specification in Table 1 when compared to more traditional tests commonly used in the literature, it is also important to evaluate the test under varying degrees of serial correlation and heteroskedasticity.[20] In Table 4, we present the finite-sample empirical size when the data generating process for the regressors and errors is MA(1). Once again, we set $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$ so that the conditional heteroskedasticity enters multiplicatively through all four non-constant regressors. In this case, the temporal dependence is of limited duration and the heteroskedasticity plays a correspondingly larger role in the distribution of the errors than in the AR(1) specification. Consequently, one may expect

[20]Not surprisingly, when the errors truely are homoskedastic, the cho^R test outperforms all other robust tests in terms of finite-sample empirical size, as it exploits the homogeneity in the data. These results can be found in the Appendix C.

Table 4: **Empirical Size - Heteroskedastic MA(1) Errors** - $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$

θ_X	θ_U	(1) *hac*	(2) *pw*	(3) hac^{KV}	(4) hac^{KVn}	(5) *cho*	(6) cho^R
0.9	0.9	0.1514	0.1373	0.1285	0.0936	0.1451	0.1096
	0.7	0.1533	0.1425	0.1311	0.0947	0.1448	0.1115
	0.5	0.1465	0.1377	0.1257	0.0921	0.1439	0.1087
	0.3	0.1371	0.1348	0.1174	0.0880	0.1383	0.1041
	0.1	0.1260	0.1305	0.1074	0.0860	0.1345	0.0994
	0.0	0.1201	0.1259	0.1017	0.0824	0.1307	0.0965
0.7	0.9	0.1508	0.1394	0.1285	0.0930	0.1474	0.1138
	0.7	0.1487	0.1389	0.1273	0.0939	0.1447	0.1115
	0.5	0.1444	0.1374	0.1240	0.0907	0.1422	0.1082
	0.3	0.1342	0.1325	0.1160	0.0871	0.1398	0.1049
	0.1	0.1211	0.1262	0.1040	0.0817	0.1312	0.0979
	0.0	0.1163	0.1239	0.1000	0.0785	0.1269	0.0943
0.5	0.9	0.1427	0.1361	0.1234	0.0897	0.1410	0.1089
	0.7	0.1410	0.1342	0.1215	0.0882	0.1403	0.1089
	0.5	0.1337	0.1311	0.1149	0.0870	0.1363	0.1045
	0.3	0.1285	0.1300	0.1120	0.0860	0.1318	0.1005
	0.1	0.1228	0.1285	0.1060	0.0834	0.1301	0.0998
	0.0	0.1143	0.1233	0.0990	0.0772	0.1259	0.0955
0.3	0.9	0.1306	0.1306	0.1125	0.0827	0.1340	0.1044
	0.7	0.1277	0.1287	0.1113	0.0846	0.1314	0.1031
	0.5	0.1274	0.1289	0.1113	0.0823	0.1316	0.1029
	0.3	0.1226	0.1271	0.1070	0.0817	0.1304	0.1021
	0.1	0.1156	0.1233	0.1003	0.0766	0.1238	0.0965
	0.0	0.1105	0.1204	0.0963	0.0754	0.1206	0.0945
0.1	0.9	0.1162	0.1242	0.1015	0.0753	0.1248	0.1002
	0.7	0.1165	0.1244	0.1020	0.0770	0.1250	0.0997
	0.5	0.1154	0.1228	0.1007	0.0759	0.1248	0.0992
	0.3	0.1118	0.1212	0.0969	0.0750	0.1205	0.0954
	0.1	0.1120	0.1225	0.0988	0.0763	0.1217	0.0956
	0.0	0.1093	0.1200	0.0953	0.0747	0.1207	0.0948
0.0	0.9	0.1085	0.1196	0.0948	0.0738	0.1197	0.0964
	0.7	0.1076	0.1187	0.0940	0.0736	0.1214	0.0959
	0.5	0.1080	0.1201	0.0937	0.0729	0.1194	0.0956
	0.3	0.1058	0.1178	0.0926	0.0721	0.1169	0.0928
	0.1	0.1096	0.1193	0.0955	0.0730	0.1192	0.0949
	0.0	0.1101	0.1206	0.0961	0.0741	0.1208	0.0964

the cho^R test to be disadvantaged relative to more traditional tests since it neglects to model this heterogeneity. However, we find that the cho^R test enjoys advantages similar to the AR(1) case.

Table 4 shows that while the testing problem becomes easier across all estimators (i.e. the finite-sample size is nearer the 5% nominal level for all tests), the relative performance of the tests in the AR(1) specification remains intact in the MA(1) specification. The prewhitened estimator shows improvement over the traditional HAC estimator, and the Kiefer-Vogelsang critical values offer even further improvement in size. When $\theta_X = \theta_U = 0.9$, the *hac*, *pw*, and hac^{KV} tests have size of 0.15, 0.14, and 0.13, respectively. In comparison, the cho^R test still offers even greater improvements with a size of 0.11, or a 27% gain in test accuracy over the traditional Newey-West estimator. While the size gains are slightly smaller when the correlation is of limited duration, they are still substantial.

Figure 2 shows the size-adjusted power of the test statistics under the MA(1) specification when $\theta_X = \theta_U$. The results, as compared to the AR(1) model, remain largely unchanged as the *hac*, *pw*, and cho^R tests all show similar power, and the hac^{KVn} test shows significantly less power against alternatives.

Another way to alter the degree of heteroskedasticity in the data is to adjust the impact of individual regressors on the error scale. Recall that the error term is generated according to

$$U_t = |X_t'\zeta| \times \tilde{U}_t.$$

Rather than set $\zeta = (0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})'$, as in the previous specifications, we now set $\zeta = (0, 1, 0, 0, 0)'$ and allow the heteroskedasticity to arise from scaling the homoskedastic error term, $\tilde{U}_t$, by the absolute value of the first non-constant regressor. As we are testing the hypothesis $H_0 : \beta_2 = 0$, it may matter whether the heteroskedasticity is brought about by the first non-constant regressor or some other regressor in the model. For this reason, we also report the empirical size when the error is scaled by the absolute value of the second non-constant regressor, $\zeta = (0, 0, 1, 0, 0)'$.

Panel A of Table 5 shows the empirical size of tests when the heterogeneity arrises from the first non-constant regressor where, for brevity, we report only the case where $\rho_U = \rho_X$. While the cho^R test performs comparably to the other robust tests when the amount of serial correlation is high, it no longer exhibits a size advantage, and it becomes progressively disadvantaged as the

Figure 2: **Heteroskedastic MA(1) Regressors and Errors** $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)' - H_0 : \beta_2 = 0$

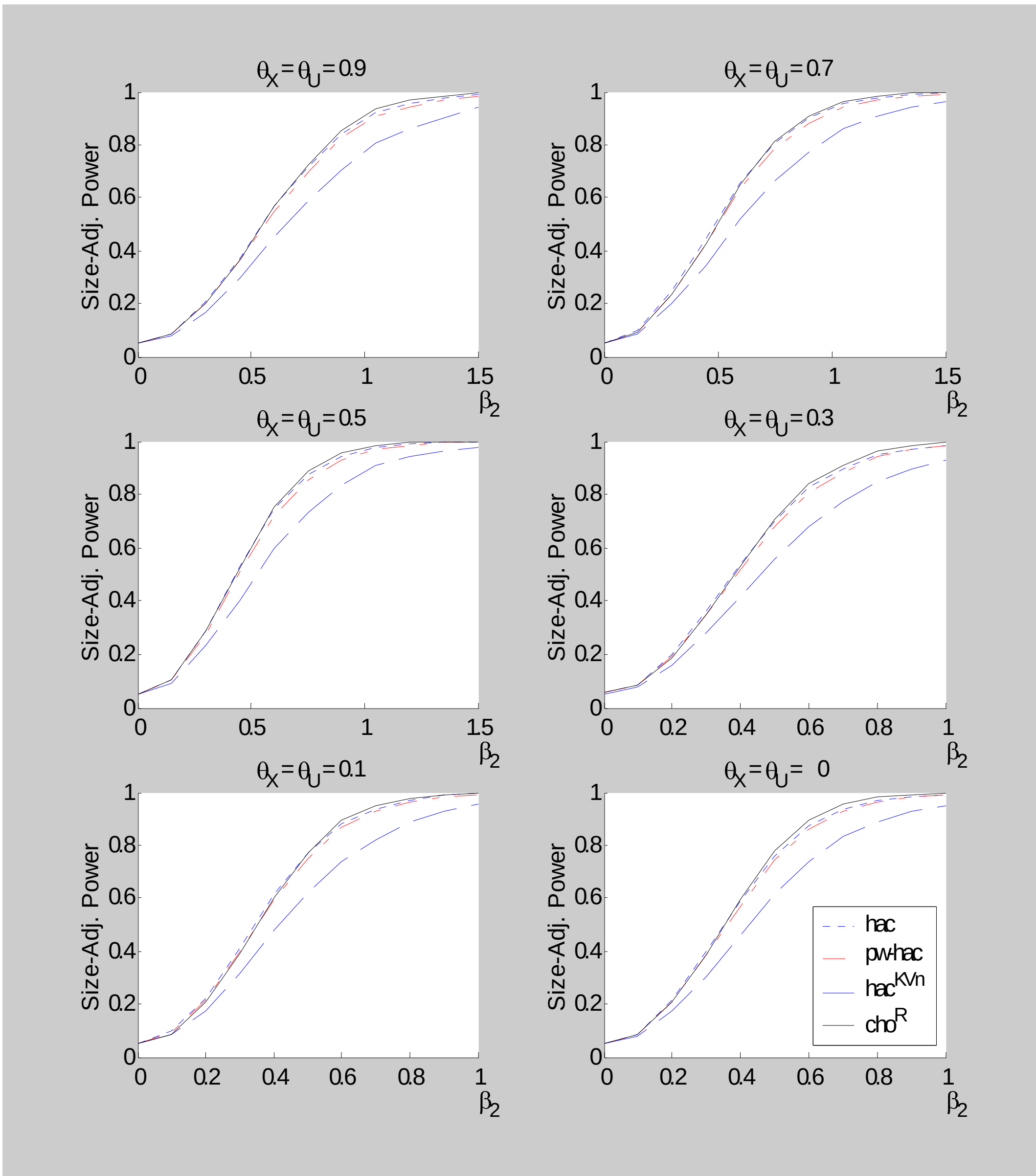


Table 5: **Empirical Size - Heteroskedastic AR(1) Errors**

$\rho_X = \rho_U$	(1) *hac*	(2) *pw*	(3) hac^{KV}	(4) hac^{KVn}	(5) *cho*	(6) cho^R
			Panel A: $\zeta = (0,1,0,0,0)'$			
0.9	0.4452	0.3959	0.3820	0.2781	0.4851	0.4023
0.7	0.2672	0.2313	0.2287	0.1625	0.3748	0.3090
0.5	0.1834	0.1708	0.1609	0.1142	0.3169	0.2674
0.3	0.1417	0.1440	0.1253	0.0935	0.2791	0.2399
0.1	0.1271	0.1369	0.1123	0.0839	0.2697	0.2321
0.0	0.1244	0.1335	0.1108	0.0829	0.2945	0.2312
			Panel B: $\zeta = (0,0,1,0,0)'$			
0.9	0.3191	0.2561	0.2566	0.1883	0.2136	0.1402
0.7	0.1925	0.1632	0.1580	0.1120	0.1168	0.0756
0.5	0.1319	0.1234	0.1104	0.0822	0.0824	0.0559
0.3	0.0994	0.1049	0.0844	0.0672	0.0670	0.0473
0.1	0.0886	0.0988	0.0761	0.0610	0.0606	0.0439
0.0	0.0866	0.0978	0.0748	0.0624	0.0607	0.0439

values of ρ_U and ρ_X fall toward zero. However, when testing the coefficient on the second non-constant regressor, the results change dramatically. Panel B reports the empirical size of the two-sided t-tests when $\zeta = (0,0,1,0,0)'$ and the heteroskedasticity arises from a regressor other than the regressor under test. In this case, the cho^R test offers a considerable size advantage over all other tests, even as the values of ρ_X and ρ_U approach zero. In fact, as the serial correlation parameters fall to zero, the size of the cho^R test falls below the 5% nominal level.

This table would appear to confirm the observation in Rothenberg that the degree of correlation between the regressor under test and the error variance has a substantial impact on inflating the empirical size of test statistics. While this is true for all tests under examination, the size distortion is especially pronounced for cho^R. It is also of note that the sizes of the tests presented in Panel B are closer to their nominal level than the corresponding sizes of the homoskedastic model presented in Table C3 in Appendix C. In practice, the form in which the heteroskedasticity enters the model appears to be of considerable importance when performing tests of hypotheses.

Table 6 reports similar results for the conditionally heteroskedastic, MA(1)

Table 6: **Empirical Size - Heteroskedastic MA(1) Errors**

$\theta_X = \theta_U$	(1) hac	(2) pw	(3) hac^{KV}	(4) hac^{KVn}	(5) cho	(6) cho^R
			Panel A: $\zeta = (0,1,0,0,0)'$			
0.9	0.1711	0.1552	0.1487	0.1092	0.3012	0.2554
0.7	0.1636	0.1511	0.1428	0.1016	0.2969	0.2524
0.5	0.1560	0.1491	0.1368	0.0997	0.2931	0.2498
0.3	0.1401	0.1427	0.1247	0.0910	0.2784	0.2403
0.1	0.1257	0.1353	0.1118	0.0836	0.2646	0.2309
0.0	0.1236	0.1339	0.1098	0.0833	0.2646	0.2295
			Panel B: $\zeta = (0,0,1,0,0)'$			
0.9	0.1237	0.1129	0.1023	0.0747	0.0768	0.0529
0.7	0.1213	0.1132	0.1004	0.0753	0.0755	0.0528
0.5	0.1125	0.1108	0.0951	0.0713	0.0721	0.0501
0.3	0.0997	0.1057	0.0847	0.0665	0.0661	0.0478
0.1	0.0896	0.0994	0.0761	0.0605	0.0613	0.0438
0.0	0.0877	0.0978	0.0752	0.0621	0.0603	0.0433

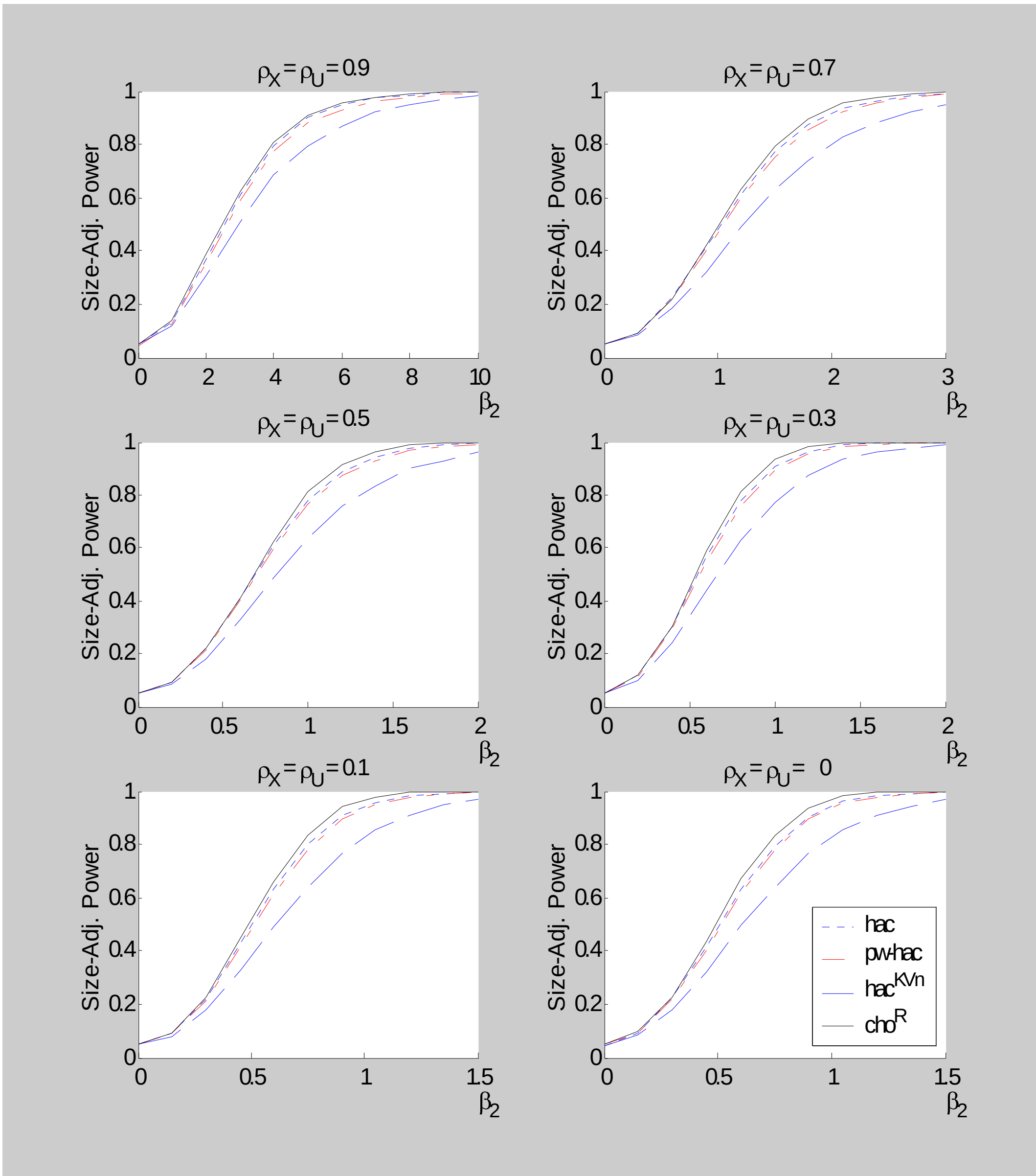
model. Once again, there appears to be little change in the conclusions from the AR(1) specification. The cho^R test performs rather poorly when the error is scaled by the regressor under test, but nearly achieves the 5% nominal size when the error is scaled by the second non-constant regressor.

Figures 3 and 4 plot the corresponding size-adjusted power curves for the AR(1) process with $\rho_X = \rho_U$ when $\zeta = (0,1,0,0,0)'$ and $\zeta = (0,0,1,0,0)'$, respectively. Likewise, Figures 5 and 6 plot the MA(1) power curves. Once again, we see cho^R performs comparably to hac and hac^{KV} in terms of size-adjusted power, and all three outperform hac^{KVn} by a considerable margin.

3.2 Multiple Parameter Tests

We now briefly turn our attention to testing hypotheses that involve multiple parameters. These multi-parameter tests incorporate at least one off-diagonal element of V_B in computing the standard error. Specifically, we examine the performance of tests under the null hypothesis $H_0 : \beta_2 - \beta_3 = 0$ (or $H_0 : c'\beta = 0$ for $c = (0,1,-1,0,0)'$) for each of the heteroskedastic specifications listed above. Tables 7 and 8 report the finite sample size for these

Figure 3: **Heteroskedastic AR(1) Regressors and Errors** $\zeta = (0, 1, 0, 0, 0)' - H_0 : \beta_2 = 0$



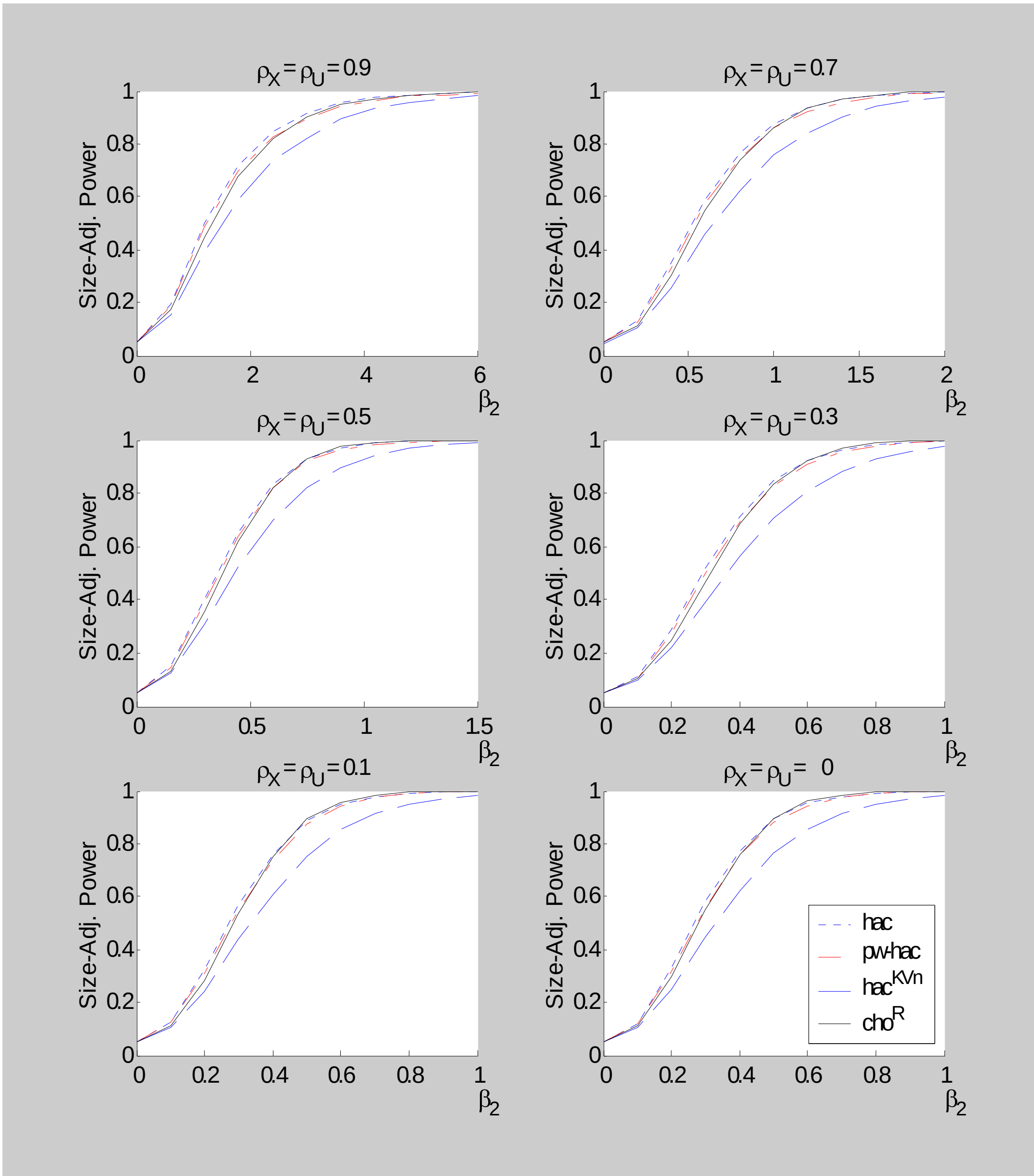

$\rho_X = \rho_U = 0.9$
$\rho_X = \rho_U = 0.7$
$\rho_X = \rho_U = 0.5$
$\rho_X = \rho_U = 0.3$
$\rho_X = \rho_U = 0.1$
$\rho_X = \rho_U = 0$
Size-Adj. Power
β_2
hac
pw-hac
hac^{KVn}
cho^{R}

Figure 4:

Figure 5: **Heteroskedastic MA(1) Regressors and Errors** $\zeta = (0,1,0,0,0)' - H_0 : \beta_2 = 0$

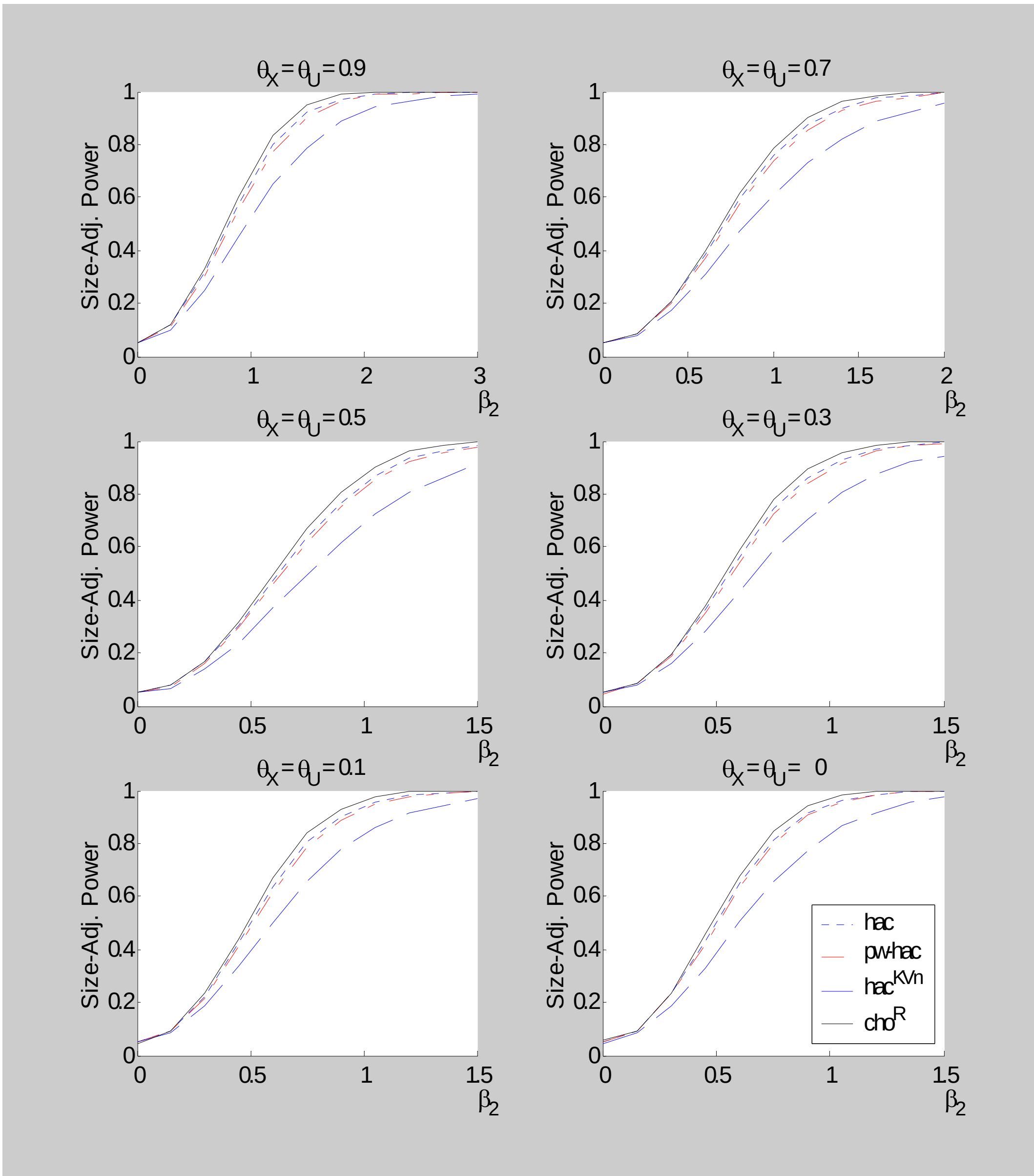
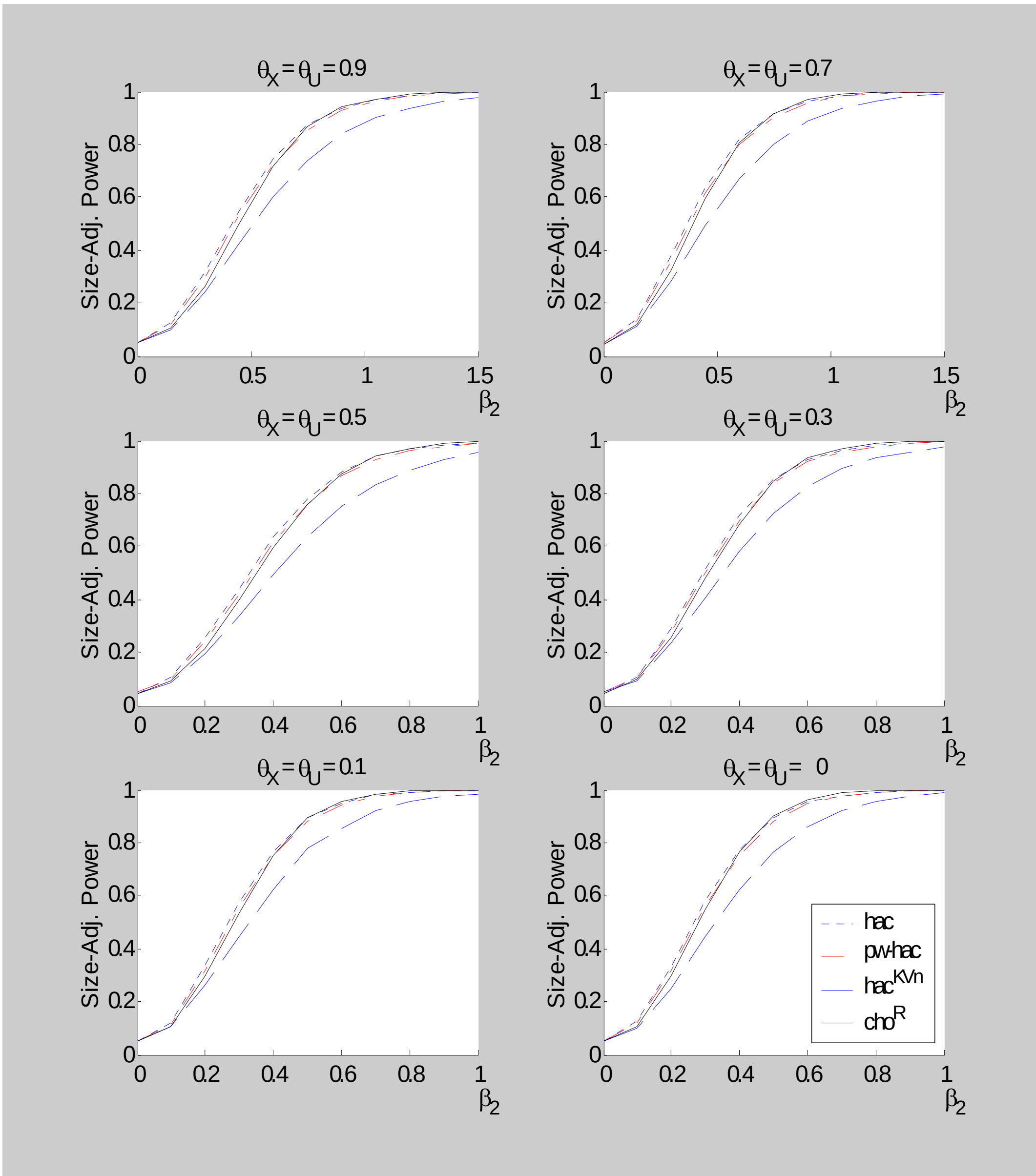


Figure 6: **Heteroskedastic MA(1) Regressors and Errors** $\zeta = (0,0,1,0,0)'$ – $H_0 : \beta_2 = 0$



multi-dimension tests for the AR(1) and MA(1) specifications, respectively. Again, we report only the results from setting $\rho_U = \rho_X$ ($\theta_U = \theta_X$).

In testing the hypothesis $\beta_2 - \beta_3 = 0$, the cho^R test provides the best empirical size under all heteroskedasticity specifications and under all degrees of dependence, excepting the case where $\zeta = (0, 1, 0, 0, 0)'$. The size advantage of the cho^R tests in the first and third heteroskedasticity specifications is quite large for both the AR(1) and MA(1) models. The empirical size of the cho^R test is cut by more than half the level of the traditional estimators, and it even exhibits better size than the hac^{KVn} test. In both cases, the empirical size approaches the nominal size rather quickly as the serial correlation decreases. When the heteroskedasticity enters solely from the one of the regressors under test, the cho^R test is not as disadvantaged as in previous specifications, but still cannot be given a strong recommendation.

Last, note that in the model under study, the testing problem is made easier by the introduction of the second parameter to the hypothesis. Comparing the relevant rows of Tables 7 and 8 to the corresponding rows in previous tables where $\rho_U = \rho_X$, we see that the empirical sizes of these tests are lower than the sizes of the single dimension tests across the full range of estimators. The same is also true for the MA(1) models as well as for values of $\rho_U \neq \rho_X$. The size-adjusted power functions for these multi-parameter tests follow the same patterns as the power functions of previous tests. The power figures for the AR(1) models can be found in Appendix C.

Table 7: **Empirical Size of Multi-Dimension Tests - Het. AR(1) Errors**

$H_0 : \beta_2 - \beta_3 = 0$

$\rho_X = \rho_U$	(1) *hac*	(2) *pw*	(3) hac^{KV}	(4) hac^{KVn}	(5) *cho*	(6) cho^R
			Panel A: $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$			
0.9	0.3220	0.2589	0.2560	0.1858	0.2118	0.1389
0.7	0.1930	0.1628	0.1561	0.1111	0.1157	0.0723
0.5	0.1324	0.1252	0.1109	0.0837	0.0835	0.0576
0.3	0.1016	0.1066	0.0859	0.0675	0.0668	0.0474
0.1	0.0902	0.0993	0.0771	0.0632	0.0621	0.0461
0.0	0.0840	0.0950	0.0712	0.0583	0.0593	0.0425
			Panel B: $\zeta = (0, 1, 0, 0, 0)'$			
0.9	0.4185	0.3653	0.3544	0.2600	0.3951	0.3091
0.7	0.2513	0.2184	0.2142	0.1514	0.2768	0.2142
0.5	0.1723	0.1620	0.1505	0.1073	0.2185	0.1739
0.3	0.1346	0.1372	0.1168	0.0882	0.1899	0.1549
0.1	0.1213	0.1307	0.1064	0.0803	0.1752	0.1465
0.0	0.1181	0.1293	0.1043	0.0793	0.1772	0.1458
			Panel C: $\zeta = (0, 0, 0, 1, 0)'$			
0.9	0.3209	0.2545	0.2540	0.1862	0.2133	0.1388
0.7	0.1883	0.1589	0.1519	0.1098	0.1163	0.0753
0.5	0.1356	0.1258	0.1134	0.0812	0.0837	0.0572
0.3	0.1026	0.1073	0.0883	0.0721	0.0678	0.0482
0.1	0.0885	0.0976	0.0751	0.0643	0.0624	0.0446
0.0	0.0887	0.0986	0.0765	0.0636	0.0635	0.0463

Table 8: **Empirical Size of Multi-Dimension Tests - Het. MA(1) Errors**

$H_0 : \beta_2 - \beta_3 = 0$

$\theta_X = \theta_U$	(1) *hac*	(2) *pw*	(3) hac^{KV}	(4) hac^{KVn}	(5) *cho*	(6) cho^R
	Panel A: $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$					
0.9	0.1250	0.1134	0.1021	0.0765	0.0762	0.0524
0.7	0.1157	0.1080	0.0952	0.0705	0.0724	0.0499
0.5	0.1102	0.1084	0.0928	0.0701	0.0700	0.0486
0.3	0.0996	0.1055	0.0844	0.0668	0.0662	0.0468
0.1	0.0870	0.0988	0.0750	0.0634	0.0612	0.0448
0.0	0.0873	0.0967	0.0744	0.0597	0.0588	0.0430
	Panel B: $\zeta = (0, 1, 0, 0, 0)'$					
0.9	0.1624	0.1481	0.1396	0.1021	0.2048	0.1653
0.7	0.1605	0.1495	0.1384	0.1000	0.2048	0.1650
0.5	0.1463	0.1419	0.1283	0.0939	0.1970	0.1610
0.3	0.1316	0.1353	0.1172	0.0863	0.1856	0.1516
0.1	0.1188	0.1294	0.1048	0.0803	0.1954	0.1448
0.0	0.1168	0.1269	0.1037	0.0795	0.1739	0.1444
	Panel C: $\zeta = (0, 0, 0, 1, 0)'$					
0.9	0.1244	0.1137	0.1023	0.0759	0.0762	0.0525
0.7	0.1219	0.1128	0.1011	0.0756	0.0754	0.0519
0.5	0.1097	0.1089	0.0920	0.0704	0.0693	0.0487
0.3	0.0972	0.1023	0.0829	0.0647	0.0643	0.0451
0.1	0.0884	0.0990	0.0766	0.0628	0.0619	0.0444
0.0	0.0846	0.0956	0.0730	0.0596	0.0580	0.0427

4 Empirical Application

To understand the impact of our proposed standard error estimator on applied research, we revisit Lustig and Verdelhan (2007) who use the Newey-West standard error estimator in their study of foreign currency risk premia. Lustig and Verdelhan propose that the excess returns of foreign currency markets can be partially explained as a premium on aggregate consumption growth risk in U.S. markets. The premium arises because returns from high interest rate countries tend to be low when (U.S. aggregate) consumption growth is low. Conversely, the assets of the low interest rate countries negatively covary with consumption growth and so can serve as a hedge against U.S. aggregate consumption growth risk.

To verify their proposition, Lustig and Verdelhan study a panel of countries from which they first assign each country to one of eight currency portfolios. The portfolios, which are formed by sorting on the short-term risk-free interest rate of the country, are rebalanced every period so that the countries with the lowest interest rates are in the first portfolio and countries with the highest interest rates are in the eighth portfolio. With annual data from 1953-2002 they estimate the covariances, which underpin their analysis, between excess currency returns and the growth rate of consumption. For each of the portfolios estimates are obtained from regressions of the form

$$R_{t+1} = \alpha + \beta f_t + U_t,$$

where R_{t+1} is the annual excess return on the portfolio and f_t is the growth of consumption in the previous year. Separate regressions are run for durable and nondurable consumption.

The estimates β along with various estimates of their standard errors are reported in Tables 9 and 10. Table 9 presents the results for the entire 50 year sample while Table 10 reports the results from the 32 year subsample after the demise of Brenton-Woods in 1971. The first row of Panel A (Panel B) reports estimates of the slope coefficients when the nondurable (durable) consumption growth factor is used as the regressor. These estimates replicate the results that are found in Table 6 of Lustig and Verdelhan. As can be seen, the consumption risk is generally increasing in the interest rate, with a difference in consumption betas of the first and seventh portfolios of about 100 basis points for the full sample, and 150 basis points for the post Brenton-Woods era. The remaining rows report the estimated standard errors and critical values which are used to test hypotheses on the regression coefficients. Numbers in braces are estimated standard errors. Numbers in italics represent alternative critical values which can be used in the hypothesis tests. The * represents significance at the 5% level.

When the Newey-West estimator is used to estimate the standard errors, only two portfolios exhibit risk parameters that are significantly different than zero when the entire 50 year sample is used. The excess currency returns are positively correlated with nondurable consumption growth in the second portfolio and positively correlated with durable consumption growth in the seventh portfolio. However, hypothesis rejection varies depending on which standard error estimator is employed. For example, for the second portfolio in Panel A of Table 9, the *hac*, hac^{KV}, and *pw* tests all reject the hypothesis that β_c^2 is equal to zero. The hac^{KV} test rejects the hypothesis even though the test critical value has increased from 1.96 to 2.0794. However, the hac^{KVn}, *cho*, and cho^R tests fail to reject the hypothesis. The hac^{KVn} test uses a standard error estimate that is almost half the size of the traditional Newey-West and prewhitened Newey-West estimates, but the critical value is too large to produce a rejection at a value of 4.813. On the other hand, the *cho* and cho^R tests use more moderate critical values, but the estimated standard errors are almost twice as large as the Newey-West standard errors. For the seventh portfolio in Panel B, the *cho* and cho^R tests also produce rejections in addition to the *hac*, hac^{KV}, and *pw* tests. The critical values of the hac^{KVn} test are once again too large to reject the hypothesis.

The amount of heteroskedasticity in these two models may help to explain the porformance of the standard error estimate in the *cho* and cho^R tests. We test the null hypothesis of homoskedasticiy using the White heteroskedasticity test, running regressions of the squared OLS residuals from each model on a constant, the factor, and the factor squared. For portfolio 2 in Panel A, the χ^2 test statistic with two degrees of freedom gives a test statistic of 1.229 and a p-value of 0.541. The lack of heteroskedasticity in this model may indicate that the standard error estimate in the *cho* and cho^R tests provide the best approximation to the true standard error and the *hac*, hac^{KV}, and *pw* tests may falsely reject the null. On the other hand, the White test of the seventh portfolio in Panel B produces a χ^2 statistic of 4.697 with a p-value of 0.096 indicating that substantial heteroskedasticity may be present. However, the standard error estimate of *cho* and cho^R tests, though slightly larger, are in line with the standard errors from the heteroskedasticity consistent tests.

When the sample size is shortened to the post Brenton-Woods era, none of the nondurable consumption growth betas are significantly different from zero. However, the durable consumption growth betas of portfolios 3, 4, and 7 are significantly different from zero in at least one of the tests. The *hac* and hac^{KV} tests reject the null hypothesis for all three portfolios. The *pw* and *cho* tests reject for the fourth and seventh portfolios, and the hac^{KVn} test rejects for the third and seventh portfolios. When using the Rothenberg second-order critical value adjustment, the cho^R test fails to reject the null hypothesis at the 5% level for any of the portfolios.

Table 9 - Estimation of Factor Betas for Eight Portfolios Sorted on Interest Rates (1953-2002)

		Portfolios							
Estimate	Crit. Val.	1	2	3	4	5	6	7	8
		Panel A: Nondurables							
β_c^j		0.1047	0.7615	0.2635	0.1820	0.6342	0.2605	1.1005	0.0855
hac	1.96	[0.5393]	[0.3559]*	[0.6878]	[1.1620]	[0.5460]	[0.8281]	[0.7742]	[1.1328]
	cv^{KV}	*2.0794*	*2.0794**	*2.0797*	*2.0196*	*2.0794*	*2.0794*	*2.0196*	*2.0794*
pw	1.96	[0.6127]	[0.3479]*	[0.8065]	[1.2876]	[0.5981]	[0.9320]	[0.8653]	[1.2124]
hac^{KVn}	4.813	[0.2390]	[0.1879]	[0.1919]	[0.5543]	[0.1943]	[0.2908]	[0.3531]	[0.4178]
cho	1.96	[0.6434]	[0.6571]	[0.6763]	[0.7677]	[0.6961]	[0.7752]	[0.7869]	[1.4216]
	cv^{R}	*2.0120*	*2.0666*	*2.0241*	*2.0352*	*2.0015*	*2.0476*	*2.0468*	*1.9809*
		Panel B: Durables							
β_d^j		0.2396	0.4889	0.6365	0.8916	0.5501	0.6948	1.2983	0.6753
hac	1.96	[0.4817]	[0.3301]	[0.4340]	[0.6264]	[0.5544]	[0.6126]	[0.5464]*	[0.6057]
	cv^{KV}	*2.0794*	*2.0794*	*2.0794*	*2.0794*	*2.0794*	*2.0196*	*2.0794**	*2.1395*
pw	1.96	[0.5678]	[0.3374]	[0.5302]	[0.7037]	[0.6237]	[0.6745]	[0.5270]*	[0.6275]
hac^{KVn}	4.813	[0.2682]	[0.1816]	[0.3245]	[0.5612]	[0.3168]	[0.3647]	[0.4318]	[0.2432]
cho	1.96	[0.5532]	[0.4862]	[0.5124]	[0.7154]	[0.6250]	[0.6374]	[0.6387]*	[0.9867]
	cv^{R}	*1.9924*	*2.1345*	*2.0155*	*2.0091*	*1.9789*	*2.0647*	*2.0140**	*2.1577*

Table 10 - Estimation of Factor Betas for Eight Portfolios Sorted on Interest Rates (1971-2002)

		Portfolios							
Estimate	Crit. Val.	1	2	3	4	5	6	7	8
		Panel A: Nondurables							
β_c^j		0.0050	0.8962	0.3586	0.6646	0.6979	0.3191	1.5461	-0.4612
hac	1.96	[0.6575]	[0.5050]	[0.8875]	[1.3994]	[0.6270]	[1.0852]	[0.9876]	[1.2458]
	cv^{KV}	*2.1471*	*2.1471*	*2.1471*	*2.0532*	*2.1471*	*2.0532*	*2.0532*	*2.2416*
pw	1.96	[0.7378]	[0.5122]	[1.0504]	[1.5284]	[0.6926]	[1.1484]	[1.0389]	[1.5223]
hac^{KVn}	4.813	[0.3251]	[0.3590]	[0.3049]	[0.5569]	[0.2401]	[0.4036]	[0.4236]	[0.4495]
cho	1.96	[0.8913]	[0.7158]	[0.9462]	[1.1022]	[0.9247]	[1.0683]	[1.1270]	[1.7593]
	cv^R	*2.1118*	*2.0951*	*2.0759*	*2.0373*	*2.1608*	*2.0957*	*2.0854*	*2.0004*
		Panel B: Durables							
β_d^j		0.5367	0.7857	1.2881	2.0321	1.2249	1.3590	2.1827	0.8447
hac	1.96	[0.7177]	[0.5437]	[0.5503]*	[0.7371]*	[0.7453]	[0.9185]	[0.7997]*	[0.8608]
	cv^{KV}	*2.1471*	*2.1471*	*2.1471**	*2.0532**	*2.1471*	*2.0532*	*2.0532**	*2.2416*
pw	1.96	[0.8108]	[0.5796]	[0.6712]	[0.7762]*	[0.8125]	[0.9431]	[0.7532]*	[0.8516]
hac^{KVn}	4.813	[0.5759]	[0.4333]	[0.2016]*	[0.4434]	[0.4511]	[0.4013]	[0.3185]*	[0.4286]
cho	1.96	[0.8597]	[0.7086]	[0.8178]	[0.9615]*	[0.8409]	[0.9694]	[1.0425]*	[1.5109]
	cv^R	*2.2717*	*2.2636*	*2.2417*	*2.2678*	*2.3921*	*2.2654*	*2.2422*	*2.2476*

The White heteroskedasticity tests for the third, fourth, and seventh portfolios in Panel B of Table 10 produce χ^2 test statistics of 1.613, 3.242, and 8.310, with p-values of 0.446, 0.198, and 0.016, respectively. There appears to be little evidence of heteroskedasticity for portfolio 3, slight evidence of heteroskedasticity in portfolio 4, and substancial evidence of heteroskedasticity in portfolio 7.

5 Conclusion

The simulation results of this paper suggest that when sample sizes are small, modeling the heterogeneity of a process is secondary to accounting for dependence. We find that a conditionally homoskedastic covariance matrix estimator (when used in conjunction with Rothenberg's second-order critical value adjustment) improves test size with only a minimal loss in test power, even when the data manifest significant amounts of heteroskedasticity. In some specifications, the size inflation was cut by nearly 40% over the traditional HAC test. Much of this size gain can be attributed to the manner in which the second-order theory adjusts the critical value according to the degree of correlation found in the data. Strong dependence usually leads to estimated standard errors which are too small, resulting in tests which reject the null hypothesis too often. Rothenberg's second-order critical value refinements serve to dampen this effect by producing larger critical values when standard errors are small, and more moderate critical values when standard errors are large.

In addition to improved small-sample test size, the conditionally homoskedastic estimator offers improvements over traditional HAC estimators in both computational ease and testing continuity. While the second-order critical value calculation is slightly complicated, the process is completely data dependent, and requires no user choices in implementation. Conversely, traditional HAC estimators require the user to specify a weighting kernel, prewhitening filter, and a bandwidth selection procedure, which previous authors have shown can have a significant impact on the performance of the estimator.

To be sure, the conditionally homoskedastic covariance estimator does not perform the best under all heteroskedastic conditions. As noted by Rothenberg, statistical inference is especially problematic when the error variance is highly correlated with the regressor of interest. Care must be taken when heteroskedasticity is caused primarily by the regressor whose coefficient is under test. However, the adjusted critical values and estimator that we propose deliver a testing procedure with substantial gains in controlling size, at little cost in power.

6 Appendix A

Accurate estimation of $Var\left(n^{\frac{1}{2}}(B-\beta)|X\right)$ hinges on accurate estimation of $n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}E\left(U_sX_sU_tX_t'|X\right)$. To describe the estimation problem let $V_t=(Y_t-X_t'\beta)X_t$ and $\hat{V}_t=(Y_t-X_t'B)X_t$. The average conditional autocovariance at lag j is

$$\Gamma_n(j)=\frac{1}{n}\sum_{t=j+1}^{n}E\left(V_tV_{t-j}'|X\right),$$

and $n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}E\left(U_sX_sU_tX_t'|X\right)$ is equivalently expressed as

$$J_n=\Gamma_n(0)+\sum_{j=1}^{n-1}\left(\Gamma_n(j)+\Gamma_n(j)'\right).$$

The issue of lag truncation arises immediately. If one simply replaces the latent errors, which form $\Gamma_n(j)$, with estimated residuals, which form $\hat{\Gamma}_n(j)=n^{-1}\sum_{t=j+1}^{n}\hat{V}_t\hat{V}_{t-j}'$, then the resulting sum is identically zero. To obtain a non-zero estimator, White and Domowitz truncate the autocovariance summation at $m<n-1$. They establish (for a class of models that includes linear regression) that if m is allowed to grow with n, but more slowly than $n^{\frac{1}{4}}$ ($m=o\left(n^{\frac{1}{4}}\right)$), then replacing latent errors with estimated residuals yields a consistent estimator of $Var\left(n^{\frac{1}{2}}(B-\beta)|X\right)$.[21]

Unfortunately, the truncation of the sum proposed by White and Domowitz does not always yield a positive semi-definite estimator of the covariance matrix. The introduction of a weight function, or kernel, yields an estimator that is positive semi-definite. Newey and West show that, with the (modified) Bartlett kernel, the estimator

$$\hat{J}_n=\hat{\Gamma}_n(0)+\sum_{j=1}^{m}\left(1-\frac{j}{m+1}\right)\left(\hat{\Gamma}_n(j)+\hat{\Gamma}_n(j)'\right)$$

is positive semi-definite. That simply downweighting autocovariances at far lags ensures positive semi-definiteness may not be intuitive. The key insight

[21] White and Domowitz report that $m=o(n^{\frac{1}{3}})$. The corrected rate is reported in Newey and West.

is that

$$s'\hat{J}_n s = \gamma(0) + 2\sum_{j=1}^{m} w(j,m)\ \gamma_n(j),$$

where the scalar quantity $\gamma_n(j) = \frac{1}{n}\sum_{t=j+1}^{n}\left(s'\hat{V}_t\right)\left(s'\hat{V}_{t-j}\right)$ corresponds to the elements of the symmetric $(m+1)$-dimensional matrix $P = [p_{ij}]$ (with $\gamma_n(|i-j|) = p_{ij}$). As P is positive semi-definite by construction (McLeod and Jimenez 1984), if there exists a vector v such that $s'\hat{J}_n s = v'Pv$, then $\hat{J}_n$ is positive semi-definite as well. The required condition linking the kernel and v is

$$w(j,m) = \left[\sum_{i=0}^{m-j} v(i,m)\, v(i+j,m)\right] / \left[\sum_{i=0}^{m} v(i,m)^2\right]. \tag{2}$$

For the Bartlett kernel $w(j,m) = 1 - \left(\frac{j}{m+1}\right)$ and, with $v = \iota/\sqrt{m+1}$ for ι the $(m+1)$ vector of ones, (2) is satisfied for all m. For the truncated kernel, there is no vector v for which (2) is satisfied for all m.

How rapidly m grows with n affects the rate of convergence of $\hat{J}_n$. For the Bartlett kernel, the convergence is most rapid if m is proportional to $n^{\frac{1}{3}}$, in which case the rate of convergence of $\hat{J}_n$ is also $n^{\frac{1}{3}}$. Although this relatively rapid rate of growth for m violates the conditions for consistency derived by Newey and West, Andrews establishes that $\hat{J}_n$ is consistent (that is, $\hat{J}_n - J_n \xrightarrow{p} 0$) if $m = o\left(n^{\frac{1}{2}}\right)$.[22] To provide guidance regarding the choice of m in finite samples, Andrews derives the asymptotic (truncated) MSE of $\hat{J}_n$. As the asymptotic MSE depends upon the unknown serial correlation structure of V_t, selection of m to minimize the asymptotic MSE requires an estimator of this correlation structure. To implement the method, Andrews suggests approximating the correlation structure by estimating an AR(1) for each element of V_t.

Andrews shows that the asymptotic MSE can be reduced further by selecting a kernel that assigns non-zero weight to all lags (the quadratic-spectral kernel), hence m is termed a smoothing parameter (and need not be integer valued). Even though the quadratic-spectral kernel does not satisfy (2), the resulting covariance estimator is positive semi-definite because the kernel yields a non-negative spectral density estimator (for a univariate time series).

[22] Indeed, under strict conditions Andrews establishes that $\hat{J}_n$ is consistent if $m = o(n)$.

Selection of a kernel that minimizes the asymptotic MSE of the estimator of J_n may not result in test statistics with optimal performance. While the quadratic-spectral kernel yields a smaller MSE than the Bartlett kernel in simulations, due in large measure to smaller bias, both of these covariance estimators produce coverage probabilities for estimated confidence intervals that are lower than the nominal confidence levels. Indeed, no clear consensus has emerged regarding the optimal kernel for constructing test statistics on the elements of β.

One clear insight that has emerged is the importance of selecting m. Of particular importance is to allow the potential contributions of far lags, which mirrors the suggested use of the quadratic-spectral kernel that admits all lags. With larger values of m, the number of allowable lags in the Bartlett kernel becomes such a large fraction of the sample size that the asymptotic theory based on m growing slowly relative to n may not deliver an accurate finite sample approximation. To address the issue, Kiefer and Vogelsang derive asymptotic theory for $\hat{J}_n$ under the assumption that $\frac{m}{n} = b$ with $b \in (0, 1]$.[23] As could be inferred from Andrews, $\hat{J}_n$ is inconsistent if m grows at the same rate as n ($m = O(n)$).[24] However, test statistics for β (such as t and F statistics) remain asymptotically pivotal and their limit distributions can be calculated. Further, as the distributions depend both upon the kernel and b, asymptotic local power can be used to determine an optimal choice of bandwidth and kernel.

Kiefer and Vogelsang show that the limit distribution of a t statistic for a single element of β is $Q_{KV}(b)^{-\frac{1}{2}} \cdot N(0,1)$, where $Q_{KV}(b)$ is a scalar random variable with moments that depend on both the kernel and b. As the expected value of $Q_{KV}(b)$ (for the Bartlett kernel) is less than 1 for all considered values of b, the presence of $Q_{KV}(b)$ in the limit distribution tends to increase the dispersion of the t statistic. Although the mean and variance of $Q_{KV}(b)$ can be represented analytically, there is no general analytic expression for the limit distribution. Instead the critical values are obtained by simulation, which for the Bartlett kernel and a 5 percent nominal size,

[23] Even if $b = 1$, the resultant estimator is not identically zero because of the weights $w(j, m)$.

[24] If $m = O(n)$, then $\hat{J}_n$ converges to a random quantity. For $b = 1$, the expected value of this quantity is only $\frac{1}{3}$ of the asymptotic covariance matrix, which motivates why standard errors constructed from $\hat{J}_n$ tend to be too small.

Kiefer and Vogelsang calculate as

$$1.96 + 2.97b + .42b^2 - .53b^3.$$

The critical values increase rapidly with b, for $b = .5$ the critical value is 3.48 while for $b = 1$ the critical value is 4.82. (The increase in critical values is even more pronounced for the quadratic-spectral kernel.) These critical values provide confidence interval coverage probabilities that are closer to their nominal levels over a wide range of bandwidth parameters, though slight over rejections still persist in finite samples. Unfortunately, the improved coverage probability comes at the expense of lengthening the interval, and therefore reduces the power of the test considerably.

Allowing m to grow at the same rate as n is only one way to account for the contributions of far lags. Phillips, Sun and Jin (2006) alter the kernel, rather than the rate of growth of m, and study

$$\hat{J}_n(\rho) = \hat{\Gamma}_n(0) + \sum_{j=1}^{m} \left(1 - \frac{j}{m+1}\right)^{\rho} \left(\hat{\Gamma}_n(j) + \hat{\Gamma}_n(j)'\right)$$

where $\rho \geq 1$. Clearly, $\rho = 1$ returns the Bartlett kernel and $\hat{J}_n(1) = \hat{J}_n$. For larger values of ρ, the weights decline more steeply and the kernel is termed the steep-origin kernel. The presence of ρ implies that one can establish that $\hat{J}_n(\rho)$ is consistent even if $m = O(n)$ by allowing ρ to grow at an appropriate rate with n. In consequence, m is set equal to n in studying the properties of $\hat{J}_n(\rho)$. Phillips, Sun and Jin establish that $\hat{J}_n(\rho)$ is consistent if ρ grows more slowly than $n/\ln(n)$ ($\rho = o\left(\frac{n}{\ln(n)}\right)$). The convergence is most rapid if ρ is proportional to $n^{\frac{2}{3}}$, in which case the rate of convergence for $\hat{J}_n(\rho)$ is $n^{\frac{1}{3}}$ in parallel with the earlier results of Andrews.

If the value of ρ is fixed independently of n, then Phillips, Sun and Jin show that $\hat{J}_n(\rho)$ converges to a random quantity. The limit distribution of a t statistic for a single element of β is $Q_P(\rho)^{-\frac{1}{2}} \cdot N(0,1)$ where the moments of Q_P depend on ρ. For the case $\rho = 1$ the asymptotic critical value is again 4.82 and these values decline as ρ increases (for $\rho = 16$ the critical value is 2.32). In terms of (size-adjusted) power, simulations conducted by Phillips, Sun and Jin show surprisingly few gains for $\hat{J}_n(\rho)$ over a traditional estimator ($\hat{J}_n(1)$ with m determined by the data rule proposed by Andrews). For the case of $\rho = 16$, there are power gains for $n = 50$.

7 Appendix B

CALCULATION OF SECOND-ORDER CRITICAL VALUES

We are interested in testing the single restriction hypothesis $H_0 : c'\beta = c'\beta_0$. Under the assumption of conditional homoskedasticity, we form the test statistic

$$t_{cho} = \frac{c'\sqrt{n}\left(B - \beta_0\right)}{\left[c'\left(n^{-1}X'X\right)^{-1}\hat{J}_{cho}\left(n^{-1}X'X\right)^{-1}c\right]^{1/2}}.$$

Using Edgeworth expansions, Rothenberg shows that the second-order adjusted critical value of this test statistic, defined as $\Pr\left(t_{cho} > cv_\alpha^R\right) = \alpha$, can be expressed as

$$cv_\alpha^R = Z_\alpha\left(1 + \frac{\frac{1}{4}\left(1 + Z_\alpha^2\right)\hat{V}_W - \hat{a}\left(Z_\alpha^2 - 1\right) - \hat{b}}{2n}\right)$$

where Z_α is the corresponding α critical value from the standard normal distribution. The formula for $\hat{V}_W$ comes from rewriting the test statistic t_{cho} as

$$t_{cho} = \frac{T_{cho}}{\left(1 + \frac{W}{\sqrt{n}}\right)^{1/2}}.$$

Here, T_{cho} is the test statistic formed using the true value of J_{cho}, and by definition,

$$T_{cho} = \frac{c'\left(B - \beta_0\right)}{\left[c'\left(X'X\right)^{-1}J_{cho}\left(X'X\right)^{-1}c\right]^{1/2}}$$

and

$$W = \sqrt{n}\frac{c'\left(X'X\right)^{-1}\left(n\hat{J}_{cho} - J_{cho}\right)\left(X'X\right)^{-1}c}{c'\left(X'X\right)^{-1}J_{cho}\left(X'X\right)^{-1}c}.$$

If the regression errors are known to be conditionally homoskedastic and stationary with unknown autocovariances, Rothenberg (pg. 1006) derives the specific form of W as well as its variance. The estimator of the variance of W is

$$\hat{V}_W = \frac{2\sum_k\left(\sum_j r_j\hat{\delta}_{j+k}\right)^2}{\left(\sum_k r_k\hat{\delta}_k\right)^2}, \qquad \begin{array}{l} k = -\left(n-1\right),\ldots,0,\ldots\left(n-1\right) \\ j = -\left(n-1\right),\ldots,0,\ldots\left(n-1\right) \end{array}$$

where $\hat{\delta}_j$ is the j^{th} sample autocovariance as defined in section 2 and

$$r_j = n^{-1} \sum_{t=1}^{n-|j|} x_t x_{t+|j|} \qquad j = -(n-1), \ldots, 0, \ldots, n-1$$

with $x_t = nX (X'X)^{-1} c$.

The parameters $\hat{a}$ and $\hat{b}$ are defined as

$$\hat{a} = \frac{\sum_k r_k \bar{r}_k}{\sum_k r_k \hat{\delta}_k} \quad \text{and} \quad \hat{b} = \frac{\sum_k r_k \bar{q}_{kk}}{\sum_k r_k \hat{\delta}_k}, \qquad k = -(n-1), ..., (n-1)$$

where $\bar{r}_k = (n-k)^{-1} \sum_t \hat{z}_t \hat{z}_{t+k}$, and

$$\begin{aligned} \bar{q}_{kk} &= trace\left[(X'X)^{-1}(X'X_{-k})(X'X)^{-1}\left(n\hat{J}_{cho}\right)\right] \\ &\quad -2 \times trace\left[(X'X)^{-1} X'\hat{\Delta}X_{-k}\right]. \end{aligned}$$

In the equation for $\bar{r}_k$, $\hat{z}_t$ is defined as the t^{th} element of the $n \times 1$ vector

$$\hat{z} = \frac{M\hat{\Delta}x}{\sqrt{n^{-1}x'\hat{\Delta}x}}$$

where M is the $n \times n$ matrix $I_n - X (X'X)^{-1} X'$, and $\hat{\Delta}$ is the $n \times n$ estimated covariance matrix for U with (s,t) element equal to $\hat{\delta}_{|s-t|}$. The lagged cross product matrices $X'X_{-k}$ and $X'\hat{\Delta}X_{-k}$ are formed by summing over the $n-|k|$ common observations.

8 Appendix C

Table C1: **Empirical Size of *iid* and *par* - Heteroskedastic Errors** - $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$

		AR(1)		MA(1)	
ρ_X/θ_X	ρ_U/θ_U	*iid*	*par*	*iid*	*par*
0.9	0.9	0.4422	0.3002	0.1760	0.1412
	0.7	0.3171	0.2056	0.1735	0.1418
	0.5	0.2220	0.1547	0.1632	0.1385
	0.3	0.1531	0.1270	0.1457	0.1321
	0.1	0.1010	0.1092	0.1233	0.1292
	0.0	0.0822	0.1024	0.1102	0.1265
0.7	0.9	0.3464	0.2441	0.1762	0.1446
	0.7	0.2765	0.1959	0.1699	0.1410
	0.5	0.2146	0.1677	0.1619	0.1383
	0.3	0.1602	0.1412	0.1459	0.1340
	0.1	0.1192	0.1251	0.1211	0.1261
	0.0	0.0991	0.1155	0.1087	0.1226
0.5	0.9	0.2596	0.1955	0.1636	0.1380
	0.7	0.2188	0.1676	0.1595	0.1365
	0.5	0.1820	0.1509	0.1507	0.1322
	0.3	0.1487	0.1348	0.1365	0.1275
	0.1	0.1215	0.1250	0.1209	0.1244
	0.0	0.1079	0.1197	0.1121	0.1204
0.3	0.9	0.1876	0.1565	0.1454	0.1292
	0.7	0.1710	0.1439	0.1432	0.1285
	0.5	0.1544	0.1363	0.1401	0.1285
	0.3	0.1326	0.1238	0.1324	0.1259
	0.1	0.1203	0.1209	0.1183	0.1192
	0.0	0.1136	0.1188	0.1109	0.1160
0.1	0.9	0.1342	0.1283	0.1261	0.1222
	0.7	0.1305	0.1249	0.1255	0.1215
	0.5	0.1285	0.1245	0.1244	0.1212
	0.3	0.1226	0.1207	0.1179	0.1159
	0.1	0.1134	0.1134	0.1174	0.1178
	0.0	0.1128	0.1142	0.1136	0.1151
0.0	0.9	0.1157	0.1199	0.1142	0.1169
	0.7	0.1155	0.1202	0.1157	0.1189
	0.5	0.1143	0.1169	0.1135	0.1150
	0.3	0.1126	0.1143	0.1115	0.1124
	0.1	0.1140	0.1153	0.1128	0.1135
	0.0	0.1127	0.1135	0.1150	0.1151

Table C2: **Empirical Size of *iid* and *par* - Heteroskedastic Errors**

$\rho_X = \rho_U$ / $\theta_X = \theta_U$	AR(1) *iid*	AR(1) *par*	MA(1) *iid*	MA(1) *par*
Panel A: $\zeta = (0,1,0,0,0)'$ – $H_0 : \beta_2 = 0$				
0.9	0.5919	0.4642	0.3358	0.2929
0.7	0.4412	0.3567	0.3248	0.2876
0.5	0.3436	0.3047	0.3083	0.2841
0.3	0.2801	0.2697	0.2790	0.2696
0.1	0.2608	0.2612	0.2559	0.2559
0.0	0.2550	0.2556	0.2541	0.2539
Panel B: $\zeta = (0,0,1,0,0)'$ – $H_0 : \beta_2 = 0$				
0.9	0.3337	0.1945	0.0984	0.0746
0.7	0.1720	0.1084	0.0937	0.0733
0.5	0.0995	0.0779	0.0821	0.0691
0.3	0.0681	0.0628	0.0675	0.0636
0.1	0.0564	0.0569	0.0577	0.0578
0.0	0.0555	0.0562	0.0566	0.0566
Panel C: $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$ – $H_0 : \beta_2 - \beta_3 = 0$				
0.9	0.3365	0.1915	0.0987	0.0744
0.7	0.1717	0.1083	0.0902	0.0696
0.5	0.1017	0.0788	0.0805	0.0680
0.3	0.0698	0.0639	0.0689	0.0637
0.1	0.0583	0.0584	0.0572	0.0572
0.0	0.0550	0.0555	0.0556	0.0556
Panel D: $\zeta = (0,1,0,0,0)'$ – $H_0 : \beta_2 - \beta_3 = 0$				
0.9	0.5167	0.3754	0.2393	0.2006
0.7	0.3488	0.2623	0.2341	0.1998
0.5	0.2456	0.2089	0.2132	0.1920
0.3	0.1933	0.1837	0.1876	0.1791
0.1	0.1699	0.1700	0.1672	0.1678
0.0	0.1686	0.1698	0.1661	0.1669
Panel E: $\zeta = (0,0,0,1,0)'$ – $H_0 : \beta_2 - \beta_3 = 0$				
0.9	0.3360	0.1932	0.0988	0.0737
0.7	0.1734	0.1095	0.0941	0.0730
0.5	0.1059	0.0811	0.0794	0.0665
0.3	0.0698	0.0632	0.0668	0.0625
0.1	0.0586	0.0590	0.0580	0.0583
0.0	0.0598	0.0597	0.0538	0.0544

Table C3: **Empirical Size - Homoskedastic AR(1) Errors**

ρ_X	ρ_U	(1) *iid*	(2) *par*	(3) *hac*	(4) *pw*	(5) hac^{KV}	(6) hac^{KVn}	(7) *cho*	(8) cho^R
0.9	0.9	0.4229	0.2428	0.3746	0.3006	0.3019	0.2304	0.2537	0.1695
	0.7	0.2957	0.1598	0.2772	0.2182	0.2241	0.1708	0.1860	0.1175
	0.5	0.1973	0.1210	0.2104	0.1716	0.1743	0.1350	0.1498	0.0951
	0.3	0.1273	0.0954	0.1596	0.1403	0.1354	0.1132	0.1280	0.0769
	0.1	0.0763	0.0812	0.1262	0.1221	0.1088	0.0955	0.1128	0.0647
	0.0	0.0570	0.0740	0.1120	0.1151	0.0947	0.0884	0.1041	0.0558
0.7	0.9	0.3106	0.1623	0.2779	0.2203	0.2194	0.1589	0.1557	0.0985
	0.7	0.2272	0.1224	0.2218	0.1771	0.1782	0.1309	0.1296	0.0828
	0.5	0.1630	0.1015	0.1811	0.1520	0.1509	0.1124	0.1146	0.0738
	0.3	0.1102	0.0847	0.1439	0.1305	0.1232	0.0959	0.0989	0.0635
	0.1	0.0713	0.0732	0.1155	0.1158	0.0984	0.0840	0.0860	0.0538
	0.0	0.0540	0.0670	0.1028	0.1079	0.0871	0.0771	0.0795	0.0506
0.5	0.9	0.2087	0.1135	0.2012	0.1638	0.1586	0.1133	0.1006	0.0711
	0.7	0.1659	0.0982	0.1765	0.1469	0.1420	0.1034	0.0971	0.0661
	0.5	0.1261	0.0850	0.1502	0.1329	0.1253	0.0932	0.0880	0.0595
	0.3	0.0945	0.0747	0.1267	0.1179	0.1077	0.0814	0.0806	0.0546
	0.1	0.0668	0.0661	0.1081	0.1101	0.0922	0.0762	0.0728	0.0491
	0.0	0.0544	0.0618	0.0963	0.1032	0.0832	0.0691	0.0669	0.0458
0.3	0.9	0.1351	0.0874	0.1521	0.1285	0.1220	0.0880	0.0722	0.0591
	0.7	0.1150	0.0811	0.1400	0.1242	0.1159	0.0867	0.0777	0.0566
	0.5	0.0963	0.0719	0.1268	0.1193	0.1069	0.0807	0.0726	0.0521
	0.3	0.0781	0.0663	0.1138	0.1136	0.0976	0.0757	0.0693	0.0494
	0.1	0.0630	0.0612	0.1025	0.1077	0.0890	0.0715	0.0647	0.0461
	0.0	0.0569	0.0597	0.0973	0.1041	0.0844	0.0682	0.0643	0.0459
0.1	0.9	0.0761	0.0697	0.1083	0.1018	0.0899	0.0671	0.0531	0.0507
	0.7	0.0719	0.0663	0.1068	0.1042	0.0891	0.0719	0.0613	0.0474
	0.5	0.0682	0.0638	0.1040	0.1069	0.0885	0.0711	0.0635	0.0477
	0.3	0.0639	0.0613	0.0997	0.1061	0.0859	0.0675	0.0623	0.0455
	0.1	0.0583	0.0577	0.0955	0.1039	0.0833	0.0656	0.0607	0.0445
	0.0	0.0564	0.0571	0.0942	0.1026	0.0818	0.0656	0.0601	0.0431
0.0	0.9	0.0575	0.0649	0.0925	0.0928	0.0769	0.0618	0.0463	0.0465
	0.7	0.0560	0.0608	0.0937	0.0974	0.0781	0.0639	0.0561	0.0445
	0.5	0.0563	0.0596	0.0931	0.0986	0.0791	0.0647	0.0586	0.0434
	0.3	0.0570	0.0584	0.0929	0.0999	0.0801	0.0666	0.0591	0.0440
	0.1	0.0569	0.0580	0.0945	0.1035	0.0821	0.0663	0.0601	0.0441
	0.0	0.0559	0.0560	0.0930	0.1010	0.0802	0.0651	0.0593	0.0434

Table C4: **Empirical Size - Homoskedastic MA(1) Errors**

ρ_X	ρ_U	(1) *iid*	(2) *par*	(3) *hac*	(4) *pw*	(5) hac^{KV}	(6) hac^{KVn}	(7) *cho*	(8) cho^R
0.9	0.9	0.1193	0.0789	0.1386	0.1168	0.1135	0.0838	0.0781	0.0533
	0.7	0.1168	0.0773	0.1350	0.1158	0.1112	0.0803	0.0773	0.0525
	0.5	0.1064	0.0748	0.1309	0.1165	0.1103	0.0827	0.0757	0.0521
	0.3	0.0903	0.0724	0.1236	0.1168	0.1061	0.0800	0.0748	0.0519
	0.1	0.0693	0.0690	0.1095	0.1110	0.0935	0.0739	0.0737	0.0484
	0.0	0.0571	0.0669	0.1032	0.1084	0.0880	0.0727	0.0711	0.0471
0.7	0.9	0.1138	0.0764	0.1347	0.1164	0.1106	0.0819	0.0750	0.0523
	0.7	0.1096	0.0750	0.1330	0.1150	0.1106	0.0810	0.0750	0.0531
	0.5	0.1030	0.0734	0.1294	0.1168	0.1091	0.0811	0.0746	0.0509
	0.3	0.0884	0.0722	0.1206	0.1147	0.1035	0.0788	0.0744	0.0511
	0.1	0.0665	0.0658	0.1072	0.1096	0.0918	0.0729	0.0692	0.0470
	0.0	0.0554	0.0638	0.0996	0.1055	0.0850	0.0705	0.0668	0.0450
0.5	0.9	0.1055	0.0743	0.1294	0.1164	0.1077	0.0808	0.0739	0.0526
	0.7	0.1008	0.0732	0.1271	0.1133	0.1058	0.0783	0.0733	0.0518
	0.5	0.0974	0.0727	0.1261	0.1167	0.1067	0.0795	0.0741	0.0510
	0.3	0.0849	0.0695	0.1170	0.1136	0.0990	0.0753	0.0712	0.0504
	0.1	0.0649	0.0644	0.1054	0.1090	0.0900	0.0724	0.0679	0.0466
	0.0	0.0562	0.0622	0.0992	0.1053	0.0850	0.0682	0.0649	0.0457
0.3	0.9	0.0907	0.0722	0.1203	0.1130	0.1014	0.0766	0.0718	0.0510
	0.7	0.0873	0.0687	0.1186	0.1127	0.0998	0.0757	0.0689	0.0500
	0.5	0.0820	0.0677	0.1146	0.1118	0.0980	0.0734	0.0687	0.0492
	0.3	0.0742	0.0656	0.1080	0.1090	0.0931	0.0741	0.0682	0.0491
	0.1	0.0630	0.0619	0.1000	0.1055	0.0864	0.0698	0.0644	0.0458
	0.0	0.0559	0.0587	0.0953	0.1026	0.0823	0.0671	0.0622	0.0444
0.1	0.9	0.0659	0.0625	0.1021	0.1044	0.0857	0.0693	0.0610	0.0445
	0.7	0.0658	0.0619	0.1032	0.1050	0.0880	0.0689	0.0620	0.0450
	0.5	0.0673	0.0631	0.1040	0.1081	0.0882	0.0711	0.0641	0.0474
	0.3	0.0646	0.0616	0.1008	0.1077	0.0873	0.0687	0.0639	0.0473
	0.1	0.0590	0.0590	0.0966	0.1056	0.0841	0.0667	0.0625	0.0458
	0.0	0.0561	0.0569	0.0940	0.1028	0.0807	0.0670	0.0604	0.0445
0.0	0.9	0.0547	0.0586	0.0920	0.0977	0.0774	0.0652	0.0565	0.0423
	0.7	0.0566	0.0594	0.0946	0.1009	0.0803	0.0654	0.0581	0.0433
	0.5	0.0562	0.0589	0.0942	0.1021	0.0806	0.0659	0.0596	0.0436
	0.3	0.0570	0.0584	0.0944	0.1032	0.0830	0.0658	0.0599	0.0445
	0.1	0.0554	0.0559	0.0929	0.1062	0.0802	0.0636	0.0599	0.0426
	0.0	0.0554	0.0558	0.0937	0.1019	0.0814	0.0663	0.0587	0.0435

Figure 7: **Heteroskedastic AR(1) Regressors and Errors** $\zeta = \left(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)'$ – $H_0 : \beta_2 - \beta_3 = 0$

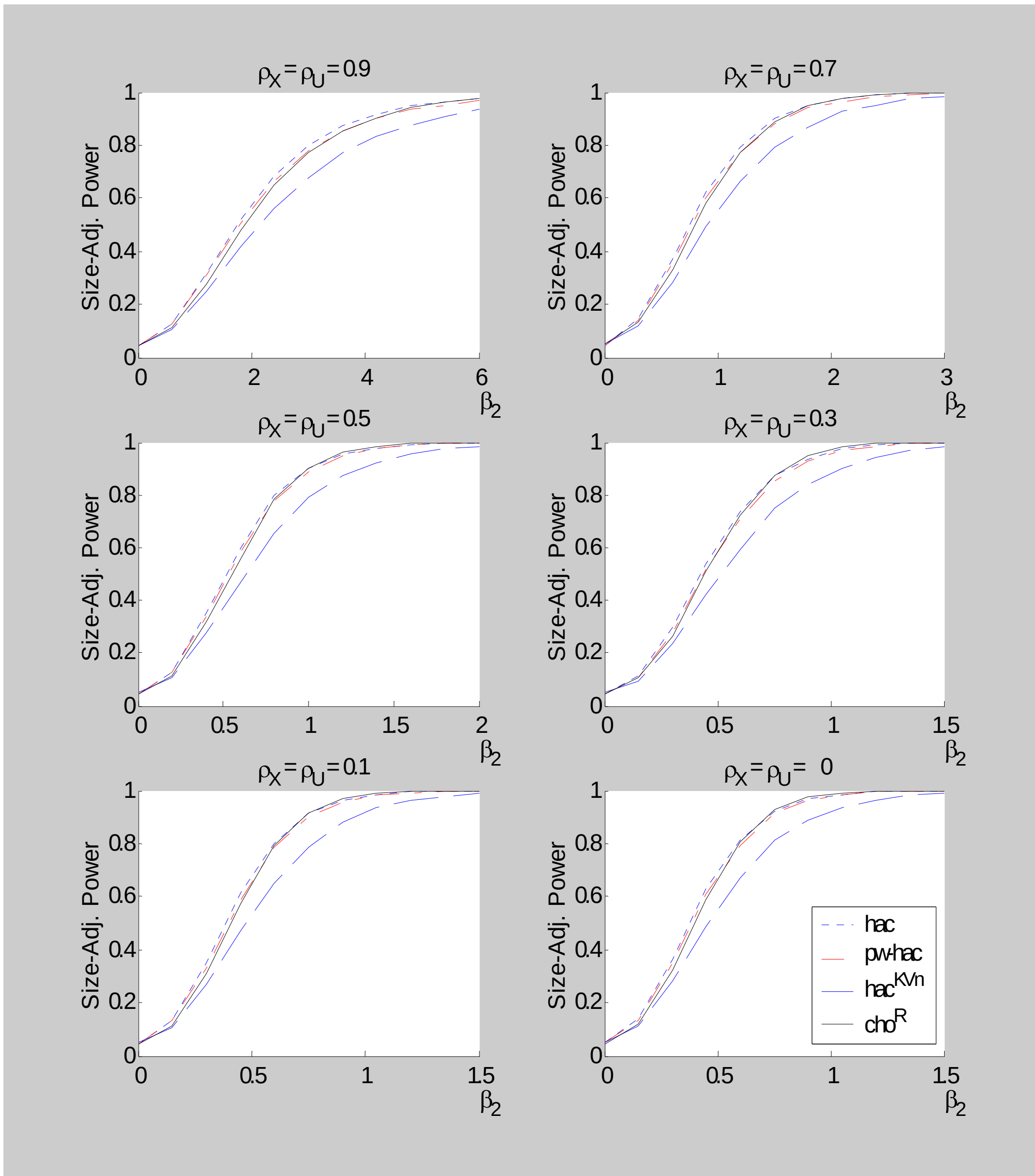


Figure 8: **Heteroskedastic AR(1) Regressors and Errors** $\zeta = (0,1,0,0,0)' \; - \; H_0 : \beta_2 - \beta_3 = 0$

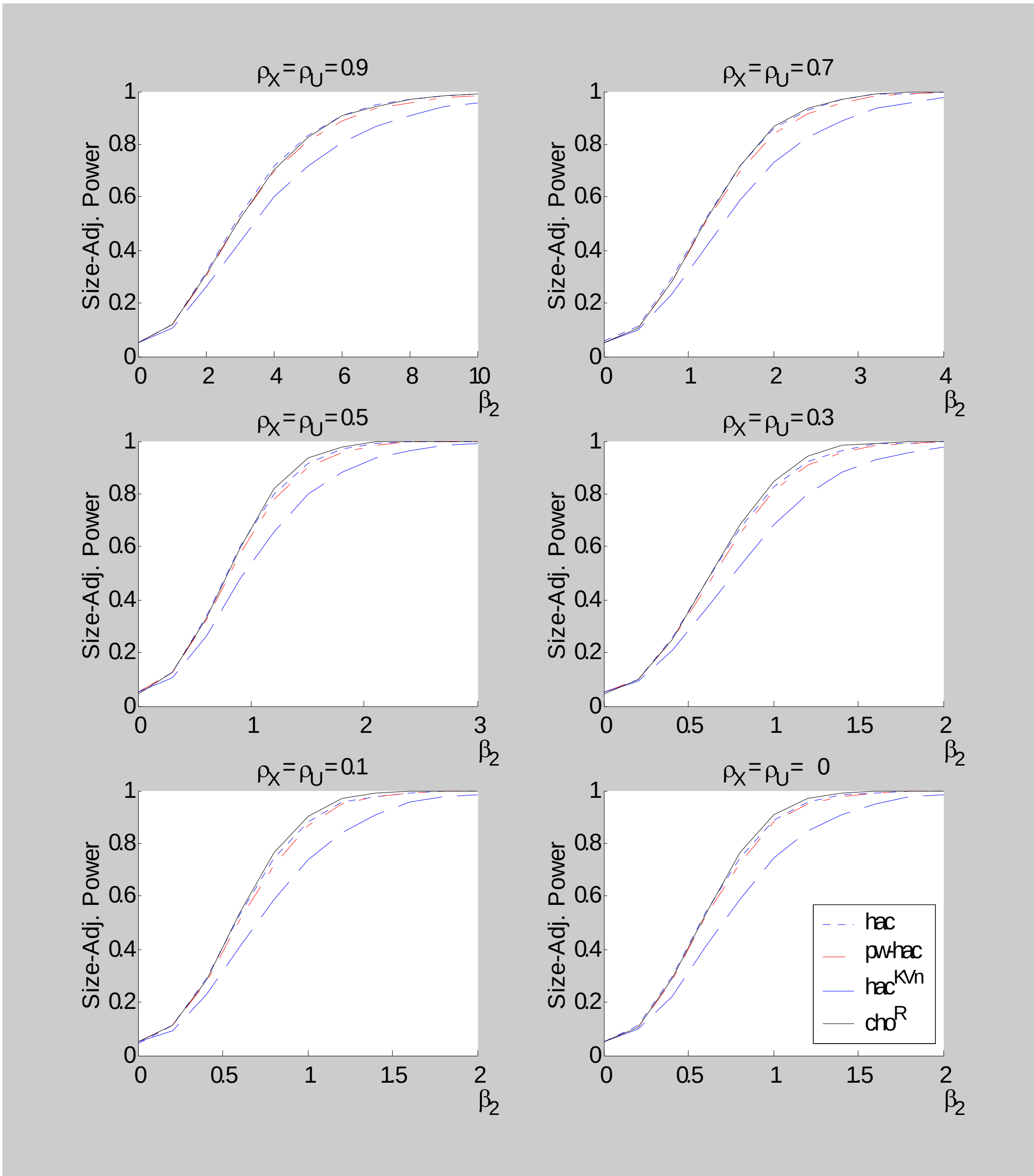
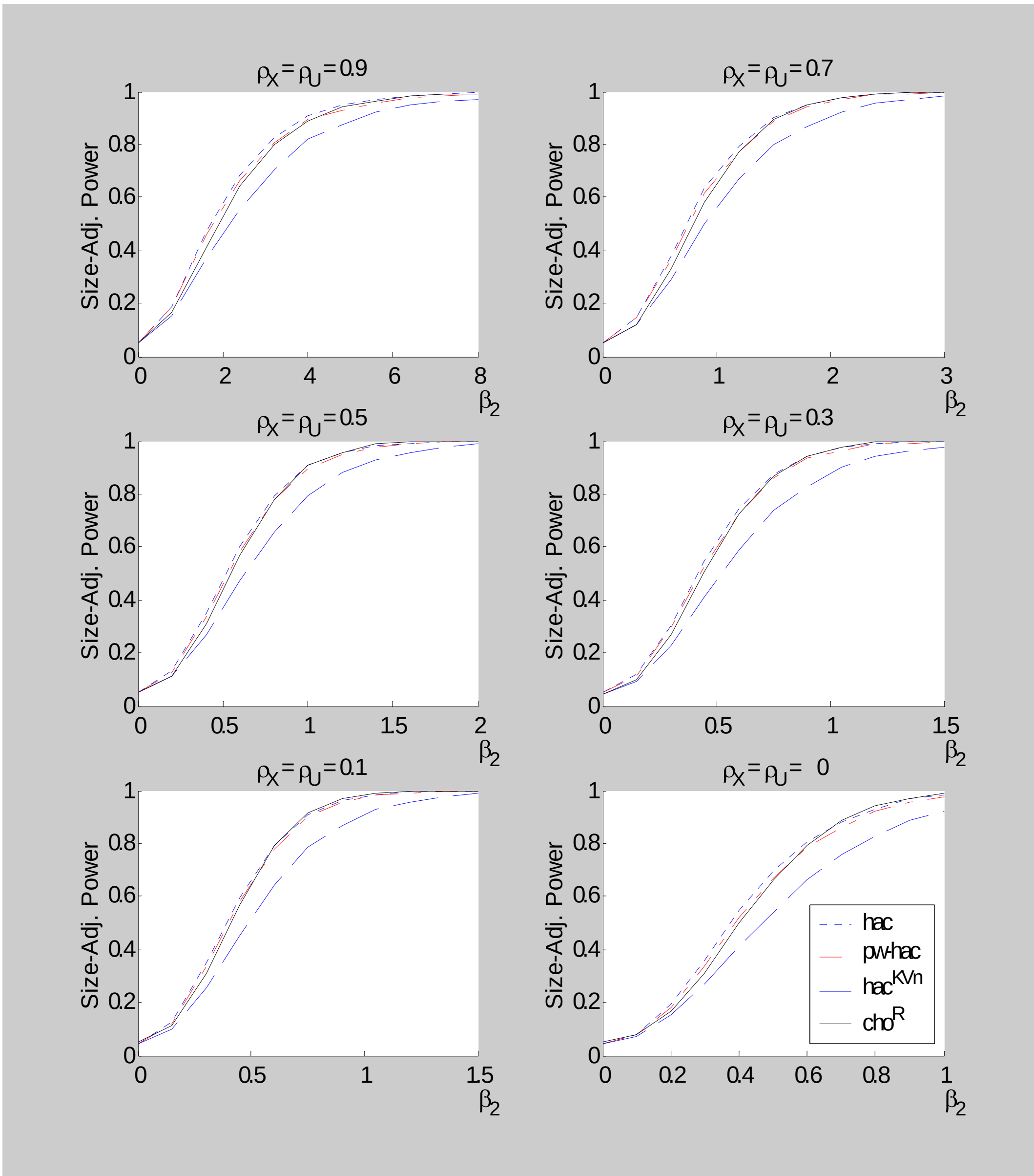


Figure 9:

Figure 10: **Heteroskedastic AR(1) Regressors and Errors** $\zeta = (0,0,0,1,0)' \; - \; H_0 : \beta_2 - \beta_3 = 0$



Bibliography

Andrews, D., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica* 59, 817-858.

Andrews, D. and J. Monahan, 1992, "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica* 60, 953-966.

Kiefer, N. and T. Vogelsang, 2005, "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests", *Econometric Theory* 21, 1130-1164.

Lustig, H. and Verdelhan, A., 2007, "The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk", *American Economic Review* 97, 89-117.

McLeod, A. and C. Jimenez, 1984, "Nonnegative Definiteness of the Sample Autocovariance Function", *American Statistician* 38, 297-298.

Newey, W. and K. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708.

Phillips, P., Y. Sun and S. Jin, 2006, "Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels Without Truncation", *International Economic Review* 47, 837-894.

Rothenberg, T., 1988, "Approximate Power Functions for Some Robust Tests of Regression Coefficients", *Econometrica* 56, 997-1019.

White, H. and I. Domowitz, 1984, "Nonlinear Regression with Dependent Observations", *Econometrica* 52, 143-162.