



# UNSW Business School

## Working Paper

Never Stand Still

Business School

Economics

UNSW Business School Research Paper No. 2016 ECON 06

Alternative HAC Covariance Matrix Estimators with Improved Finite Sample Properties

Luke Hartigan

This paper can be downloaded without charge from  
The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=2783612>

# Alternative HAC Covariance Matrix Estimators with Improved Finite Sample Properties

Luke Hartigan\*

University of New South Wales

May 23, 2016

## Abstract

HAC estimators are known to produce test statistics that reject too frequently in finite samples. One neglected reason comes from using the OLS residuals when constructing the HAC estimator. If the regression matrix contains high leverage points, such as from outliers, then the OLS residuals will be negatively biased. This reduces the variance of the OLS residuals and the HAC estimator takes this to signal a more accurate coefficient estimate. Transformations to reflate the OLS residuals and offset the bias have been used in the related HC literature for many years, but these have been overlooked in the HAC literature. Using a suite of simulations I provide strong evidence in favour of replacing the OLS residual-based HAC estimator with estimators related to extensions of either of the two main HC alternatives. In an empirical application I show how different inference from using the alternative HAC estimators can be important, not only from a statistical perspective, but also from an economic one as well.

**Keywords:** Covariance matrix estimation; Finite sample analysis; Leverage points; Auto-correlation; Hypothesis testing; Monte Carlo simulation; Inference.

**JEL Classification:** C12; C13; C15; C22.

---

\* *Correspondence:* Luke Hartigan, PhD Candidate, School of Economics, University of New South Wales, Sydney, 2052, NSW, Australia. Email: luke.hartigan@unsw.edu.au. I am grateful to Seojeong Lee, Timothy Neal, and James Morley for helpful comments and suggestions. Any remaining errors are my own.

# 1 Introduction

Accurate inference about the estimated coefficients from an ordinary least squares (OLS) regression model relies crucially on a consistent estimator of the coefficient covariance matrix. It is widely understood that if the OLS residuals have heteroskedasticity and/or serial correlation, then the usual coefficient covariance matrix will not be consistent and if it is used to perform inference, then it will lead erroneous conclusions. To overcome this, alternative covariance matrix estimators have been proposed in the literature which are consistent to unknown forms of heteroskedasticity (HC), see Eicker (1963) and White (1980) and/or autocorrelation (HAC), see Newey and West (1987). However, the performance of these so-called robust estimators can differ significantly in finite samples encountered in empirical applications.

Focusing on the HAC covariance matrix estimator, a number of authors including for example: Andrews (1991), Andrews and Monahan (1992), Burnside and Eichenbaum (1996), Christiano and den Haan (1996), Den Haan and Levin (1997) and West (1997) have shown in various simulation studies that estimating HAC covariance matrices with small to moderate sample sizes (as encountered in macroeconomics for example), tend to produce coefficient standard error estimates that lead to test statistics that are oversized (i.e., they reject the null hypothesis in a two-sided test too frequently) and confidence intervals with shorter length than otherwise.

As a means of achieving a more efficient estimator the HAC literature has devoted much attention to the problem of selecting the optimal kernel (Andrews 1991), the right bandwidth (Andrews 1991 and Newey and West 1994) or whether or not to pre-whiten the OLS residuals beforehand (Andrews and Monahan 1992 and den Haan and Levin 2000). However, only minor attention has been given to potential problems caused from using the OLS residuals as an estimate for the unobserved stochastic error term in the regression model. Note that the standard HAC estimator is just a generalisation of the consistent covariance matrix estimator popularised by White (1980). While this covariance matrix estimator (sometimes called the Eicker-White estimator or ‘HC0’ in the HC literature) is asymptotically consistent, it was

soon understood that the HC0 estimator could be seriously biased in finite samples. As a result, alternative estimators were suggested instead. These use transformations of the OLS residuals to offset the bias and do have improved finite sample properties (see MacKinnon and White 1985). These alternative estimators have been used in the cross-section literature and available in many statistical software packages for some time. However, no attention has been given in the HAC literature to extend these alternative HC estimators to the HAC setting, although Davidson and Mackinnon (1993) did conjecture that doing so might provide some improvement.

The potential problem with using the OLS residuals is that they are negatively biased estimator of the true unobserved stochastic error process. This means that the OLS residuals will usually be smaller than otherwise and hence the variance of the OLS residuals will also be smaller. The covariance matrix estimator takes these smaller variance estimates to signify more accurate coefficient estimates, leading to smaller standard errors and hence oversized test statistics and under covered confidence intervals. The bias in the OLS residuals is related to the predictor matrix itself, and is more apparent if the predictor matrix is ‘unbalanced’ in which case some observations exert more influence on the regression fit than others, such as when the data contain outliers (see Chesher and Jewitt 1987).

In the HAC literature, only two previous studies have investigated ways of overcoming the downward bias problem in the OLS residuals (see Keener et al. 1991 and Kuan and Hsieh 2008). Keener et al. (1991) attempt to estimate the bias term explicitly and then remove it as part of the estimation procedure, but only for a finite number of lags. Whereas Kuan and Hsieh (2008) suggested using the forecast error residuals in place of the OLS residuals. However, neither consider using the alternative estimators already available in the HC literature. Furthermore, both methods have some issues that seem to have prevented them from being used in practice. In relation to using the forecast errors, Diebold (2015) argues this technique can be unsatisfactory as the convergence of the model coefficients to their population values as the sample size increases can introduce non-stationarities into the estimated forecast errors. An additional issue is the practical aspect that a portion of the sample is lost in the estimation process. The method proposed by Keener et al. (1991) is

also unsatisfactory because it is not guaranteed to generate a positive definite covariance matrix. This is because it uses a sharp truncation when estimating the long-run coefficient covariance matrix. While in their simulation results the occurrence of this was small, the fact that the occurrence has a non-zero probability makes using it less appealing in practice.

Instead of these two methods, I propose replacing the OLS residuals with either one of the alternative residuals related to the HC estimators ‘HC3’ and ‘HC4m’ when estimating the HAC coefficient covariance matrix. The reason for doing this is that these two alternatives have been shown to be the better performing estimators (in the sense of generating empirical sizes closer to the true nominal size) from the HC literature. To examine if these alternative estimators provide some improvement compared to the OLS residuals, I undertake a suite of simulations based on the work of Andrews (1991), but expanded to include Student-t and Chi-squared distributed random variables in addition to Gaussian distributed random variables. Both of these distributions are more likely to lead to unbalanced predictor matrices and therefore more negatively biased OLS residuals.

My simulation results show that across all scenarios considered here the alternative estimators from the HC setting outperform the OLS residual-based HAC estimator. Although the improved performance comes at a cost of a more variable variance estimator. However, Simonoff (1993) suggests that since variances are estimated primarily to be used for inference, then a more useful method of comparison for such estimators could be related to their intended use, such as evaluations based on the true nominal size of test statistics or the true coverage probability of confidence intervals.

One situation when HAC estimators play an important role is in tests comparing predictive accuracy (Diebold and Mariano 1995). These tests can involve small samples or data with large outliers depending on the type of loss function used to compute the loss differential. As such, these tests could generate DM test statistics which are more likely to be oversized due to negatively biased OLS residual-based estimates of the standard error used to compute the DM test statistic. The results of this paper suggest the proposed alternative estimators will be more appropriate to use in these types of situations. To demonstrate I replicate the

empirical application of Diebold and Mariano (1995), but using the UK Pound/US Dollar exchange rate, and compare the performance of the OLS residual-based HAC estimator to the two alternative estimators and find that the null hypothesis is rejected less frequently when using the latter than it is using the former over a range of sample sizes and forecast horizons.

The remainder of the paper is organised as follows: Section 2 introduces the sample kernel-based HAC covariance matrix estimator as well as an explanation of why the OLS residuals are negatively biased and discusses how to offset it. Section 3 presents the Monte Carlo simulation design and the simulation results comparing the OLS residuals-based HAC estimator to the two proposed alternatives. Section 4 documents an application involving tests comparing predictive accuracy using the OLS residual and the two alternative residual estimators, while Section 6 concludes. Auxiliary computations as well as additional simulation results are presented in the Appendix.

## 2 Alternative HAC Covariance Matrix Estimators

### 2.1 HAC Covariance Matrix Estimation and OLS Residual Bias

The standard linear regression model can be expressed as:

$$y_t = x_t' \beta_0 + \epsilon_t \quad t = 1, \dots, T \quad (1)$$

Where there are  $T$  observations on the scalar response variable,  $y_t$ , and the vector of predictor variables,  $x_t = (x_{t1}, x_{t2}, \dots, x_{tN})'$ . The scalar stochastic error term,  $\epsilon_t$ , is unobserved, while  $\beta_0$  is an unknown  $N \times 1$  vector of population coefficients that indicate how the predictors relate to the responses. In matrix notation this relationship is written as:

$$Y = X\beta_0 + \epsilon \quad (2)$$

In this case  $Y$  is a  $T \times 1$  vector of responses,  $X$  is a  $T \times N$  matrix of predictors with rows  $x'_t$ , while  $\epsilon$  is a  $T \times 1$  vector of errors. The asymptotic covariance matrix for the estimated coefficients,  $\hat{\beta}_T$ , can be specified as:

$$\text{Var} \left( \sqrt{T} \left( \hat{\beta}_T - \beta_0 \right) \right) = Q_{xx}^{-1} \Omega_0 Q_{xx}^{-1} \quad (3)$$

The ‘ $T$ ’ subscript signifies that the estimation is based on a sample size of  $T$ .  $Q_{xx}^{-1}$  is the probability limit of  $\left( \frac{1}{T} X'X \right)^{-1}$  while  $\Omega_0$  is the long run covariance matrix of  $X'\epsilon$ , which has the following form:

$$\Omega_0 = \text{Var} \left( \frac{1}{\sqrt{T}} X'\epsilon \right) \quad (4)$$

$$= \mathbb{E} \left( \frac{1}{T} X'\epsilon\epsilon'X \right) \quad (5)$$

$$= \mathbb{E} \left( \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T x_t \epsilon_t \epsilon_s x'_s \right) \quad (6)$$

Note, as  $T \rightarrow \infty$  this expression becomes equivalent to  $2\pi$  times the spectral density of  $X'\epsilon$  at frequency zero (see den Haan and Levin 1997). A consistent estimator of  $\Omega_0$  is:

$$\hat{\Omega}_T = \frac{T}{T-N} \sum_{j=1}^{T-1} \kappa \left( \frac{j}{S_T} \right) \left( \hat{\Gamma}_j + \hat{\Gamma}'_j \right) \quad (7)$$

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^T x_t \hat{\epsilon}_t \hat{\epsilon}_{t-j}' x'_{t-j} \quad (8)$$

Here  $\kappa(\cdot)$  is a real-valued kernel function,  $S_T$  is a bandwidth parameter, and the term  $T/(T-N)$  is a small sample adjustment recommended by Andrews (1991) to offset the effect of estimating the coefficient vector  $\hat{\beta}_T$ . Note, non-kernel-based methods for consistently estimating  $\Omega_0$  exist as well, see for example Kiefer et al. (2000) or West (1997), but these methods do not seem to be as popular as the kernel-based method and are not readily available in many of the main statistical software packages.

The kernel function apportions declining weights on covariance terms further apart in time and is based on the assumption that these more distant covariance terms should have relatively less impact on the overall variance of the process as  $t \rightarrow T$ . The two most commonly

used kernel functions in empirical applications tend to be the Bartlett (BT) and Quadratic Spectral Density (QS) kernels. The BT kernel was used in the first kernel-based HAC estimator introduced into the econometrics literature by Newey and West (1987) and can be defined as:

$$\kappa_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Notice that the BT kernel assigns linearly declining weights as  $x$  increases and reach zero at the value  $S_T$ . In this setting the bandwidth parameter is also referred to as a lag truncation parameter since lags of order  $x > S_T$  receive zero weights. The QS kernel was introduced into the HAC estimator literature by Andrews (1991) and was also shown to be the optimal kernel (in an asymptotic truncated mean squared error sense). The QS kernel can be defined as:

$$\kappa_{QS}(x) = \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right) \quad (10)$$

The QS kernel assigns weights that decline non-linearly, reaching zero at a lag order of about 1.2 times the bandwidth parameter  $S_T$  and then oscillates around zero for higher lag orders up to the sample length  $T$  (for a more fuller explanation see den Haan and Levin 1997). Other kernels which may be used, but which are less popular than the preceding two are the Truncated, Parzan and Tukey-Hanning kernels. See Andrews (1991) for a comprehensive analysis of these alternative kernels.

There are two alternative methods of computing an appropriate bandwidth parameter; either as a function of the sample size (fixed interval) or a function of the data (data-dependent or automatic). Newey and West (1987) originally used a fixed interval denoted as  $m$ . We can relate their term to the bandwidth parameter using the expression  $S_T = m + 1$ . Asymptotic theory related to using the fixed interval bandwidth assumes it grows as  $T \rightarrow \infty$  but at a rate more slowly than  $T$  does so that  $m/T \rightarrow 0$ .

Various ‘rules of thumb’ have been suggested for determining  $m$ , one such method is  $m = 4(T/100)^{2/9}$  as recommended by Wooldridge (2013). In contrast, the automatic or data-dependent method uses the idea that the bandwidth should not simply be based on the



sample size, but should also consider the time series features of the data in addition to the kernel used (see Andrews 1991 and Newey and West 1994).

Finally, the OLS residual is used in place of the unobservable stochastic error process. In this specification the HAC estimator is closely related to the HC estimator of White (1980). Both of these estimators use the squared OLS residuals as an estimate of the variance of the unobserved stochastic error term. Asymptotically, this has been established to be a consistent estimator for both; however, using this specification can lead to seriously biased estimates in finite samples. The reason for this is that the OLS residuals are a downward biased estimator for the true stochastic error process (for more detail see Meloun and Militký 2001). To see this note that the OLS residual is just a by-product of Least Squares estimation:

$$\hat{\epsilon} = Y - X\hat{\beta}_T = Y - \hat{Y} = (I_T - P)Y = MY = M(X\beta_0 + \epsilon) = M\epsilon \quad (11)$$

Where  $I_T$  is an identity matrix,  $P$  is the projection matrix defined as  $X(X'X)^{-1}X'$  while  $M$  is the orthogonal projection matrix (also called the ‘annihilator matrix’). Equation (11) shows that the OLS residuals are in fact a linear combination of the true stochastic errors and the predictors (or regressors). This is more evident when considering the  $t^{\text{th}}$  OLS residual given as:

$$\hat{\epsilon}_t = \epsilon_t(1 - h_t) - \sum_{i \neq t}^T P_{ti}\epsilon_i \quad (12)$$

When the error variances are heteroskedastic then the covariance matrix for  $\epsilon$  is given as  $\text{Var}(\epsilon | X) = \Sigma$  which is a diagonal matrix with the individual variance terms along the main diagonal. However,  $\text{Var}(\hat{\epsilon} | X) \neq \Sigma$ , in fact the variance is defined as:

$$\text{Var}(\hat{\epsilon} | X) = \text{Var}(M\epsilon | X) = M\text{Var}(\epsilon | X)M = M(\Sigma)M = \Sigma - P\Sigma - \Sigma P + P\Sigma P \quad (13)$$

Hence, the variance of the  $t^{\text{th}}$  error term is equal to  $\sigma_t^2$ , whereas, the variance of the  $t^{\text{th}}$  OLS residual is instead given by:

$$\hat{\sigma}_t^2 = \sigma_t^2(1 - h_t)^2 - 2\left(\epsilon_t(1 - h_t)\sum_{i \neq t}^T P_{ti}\epsilon_i\right) + \left(\sum_{i \neq t}^T P_{ti}\epsilon_i\right)^2 \quad (14)$$

The  $h_t$ ,  $t = 1, \dots, T$  appearing in Equations (12) and (14) are sometimes referred to as the ‘leverage points’. They indicate the amount of influence exerted on the fitted response value  $\hat{y}_t$  by the actual response value  $y_t$ . The leverage points are computed from the diagonal elements of the projection matrix,  $P$ . It is always the case that  $0 \leq h_t \leq 1$  and since  $\text{trace}(P) = N$ , the number of columns in the predictor matrix, the average size of the  $h_t$  can be computed as  $N/T$ . When values of  $h_t$  are close to the mean value, then the predictor matrix  $X$  is described as being ‘well balanced’ since no one observation dominates the estimated fit more so than any other. As a rough guide, Hoaglin and Welsch (1978) suggested that a ‘reasonable rule of thumb’ for large leverage points as being values for  $h_t$  greater than  $2N/T$  in magnitude.

Given the bounds on the values  $h_t$  may take, it becomes clear that the possibility of severe negative bias arises when there are large leverage points. This is because the OLS residuals associated to these large leverage points have smaller magnitude on average. The covariance matrix estimator takes the smaller OLS residuals as evidence of small error variances and hence a sign of more accurate coefficient estimates (Chesher and Jewitt 1987). The negative bias will tend to be more pronounced if the sample size is small,  $\max(h_t)$  is close to 1, or if the data has positive autocorrelation (Künsch 1989). It is this that leads to smaller standard errors and hence oversized hypothesis tests and under covered confidence intervals. The direction of the bias is important as well. If it were positive then tests would be relatively more conservative and reject less often. It is the fact that the bias is negative that leads to the unappealing size properties in practice.

## 2.2 Alternative Residual Estimators

In the HC covariance matrix estimator literature there have been no less than six different estimators besides the OLS residual-based estimator proposed. These alternative estimators have become known by the labels: HC1, HC2, HC3, HC4, HC4m and HC5, while the original Eicker-White estimator is known as HC0. The first three (HC1, HC2 and HC3) were originally introduced by MacKinnon and White (1985). HC1 is the most basic estimator

and is just a simple degrees of freedom adjustment of HC0. The HC3 estimator as it is known today is based on the specification suggested by Davidson and Mackinnon (1993) as an approximation of the jackknife version originally proposed in MacKinnon and White (1985). However, Hansen (2016) shows that this approximation is really just an alternative expression for the original estimator. The latter three (HC4, HC4m and HC5) were suggested by Cribari-Neto (2004), Cribari-Neto and da Silva (2011), and Cribari-Neto et al. (2007) respectively. All three are closely related in design with HC4m and HC5 based on alterations of the HC4 estimator.

A range of different simulation studies related to cross-section data have demonstrated that these alternative residuals generally perform much better than the HC0 estimator, (see for example; MacKinnon and White 1985, Cribari-Neto et al. 2000, Kauermann and Carroll 2001, or Godfrey 2006, as well as the detailed simulation results of Long and Ervin 2000), in producing test statistics with empirical size being closer to the true nominal size. The way they achieve this is to offset the negative bias present in the OLS residuals via various transformations that ‘inflate’ them. Given these alternatives have been successful in the HC literature, it seems reasonable to assume there would be a commiserate improvement in the empirical size properties of test statistics in the HAC estimator setting. An idea that seems to have been first mentioned by Davidson and Mackinnon (1993). However, it does not appear as if this has been investigated at all in the HAC literature as yet, were most attention has been devoted to determining the optimal kernel function to use, the best method of computing the bandwidth parameter and whether or not one should use pre-whitened residuals.

I suggest replacing the OLS residual-based HAC estimator with these alternative estimators from the HC setting and investigate if doing so also leads to an improvement in finite sample performance. The previously cited studies suggest that of the six alternatives, the HC3 and the more recent HC4m estimators seem to dominate the others (in the sense of generating test statistics closest to the true nominal size). As well be shown, the HC3 estimator has some appealing theoretical properties to go along with the empirical justification. However, in some situations, particularly when the data contain high leverage points, the HC4m

estimator tends to outperform the HC3 estimator. Hence, I will focus on using these two alternative estimators in place of the standard HAC estimator based on the OLS residuals.

Note, while no other study has investigated using these alternative residuals from the HC literature, two previous studies in the HAC literature have acknowledged the problem of estimating the coefficient covariance matrix using the OLS residuals in finite samples. The first was by Keener et al. (1991), they presented an expression for the bias in the covariance matrix estimator and suggested a new estimator with the bias removed. However, they used a method that appeared to be similar to a truncated kernel HAC covariance matrix estimator. The problem with this, which the authors do note, is that it is not guaranteed to generate an estimator which is positive definite.

The second was by Kuan and Hsieh (2008), they discuss using an alternative residual as well; however, they recommended using the forecast error residual. Their method is not without some problems either. Firstly, it is slightly more computational intensive than the method I am suggesting and requires dropping a portion of the sample in estimation (in their paper Kuan and Hsieh 2008 drop the first 10% of their data set). A bigger concern though is in the way the forecast errors are computed by recursively re-estimating the OLS model on an increasing sample size. This procedure can introduce non-stationarities into the estimated forecast errors as the model coefficients converge to their population values (Diebold 2015). Nonetheless, both studies report simulation results which indicate that addressing the negative bias in the residuals can lead to more correctly sized test statistics compared to the HAC estimator using the OLS residuals.

I now describe the two transformations applied to the OLS residuals to construct the HC3 and HC4m estimators respectively. To do so I introduce the following two alternative residuals. The first alternative residual is the one used in the HC3 specification. Hansen (2016) refers to it as the prediction error residual and gives it the label of  $\tilde{\epsilon}_t$ . It can be defined as:

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{(1 - h_t)} \quad (15)$$

The appealing feature of this residual is it partially offsets the impact of the leverage points on

the OLS residuals. The second alternative residual is the one used in the HC4m specification. It differs from the HC3 method by specifying a function for the discount factor applied the term  $(1 - h_t)$  instead of a constant as with the HC3 estimator. Since Cribari-Neto and da Silva (2011) did not propose an explicit name for this residual I will refer to it in this paper as the ‘modified discounted error’ residual and label it as  $\dot{\epsilon}_m$ . The ‘ $m$ ’ is to signify it is a modification of an existing estimator. It can be defined as:

$$\dot{\epsilon}_{m,t} = \frac{\hat{\epsilon}_t}{(1 - h_t)^{\delta_t/2}} \quad (16)$$

Where  $\delta_t = \min \{\gamma_1, Th_t/N\} + \min \{\gamma_2, Th_t/N\}$  for  $t = 1, \dots, T$ . In this expression  $\gamma_1$  and  $\gamma_2$  are two real-valued positive constants which need to be set by the researcher. Based on a variety of factors the authors suggest setting  $\gamma_1 = 1.0$  and  $\gamma_2 = 1.5$ . This leads to a discount factor with a maximum value of 2.5, which is very close to the figure used to compute  $\tilde{\epsilon}_t$ , and could be one reason why its performance in the HC literature is usually close to that of the prediction error residual. This specification attempts to correct for overly large leverage points in the predictor matrix. Cribari-Neto and Zarkos (2001) show that the presence of high leverage points can be more detrimental to the finite sample performance of the coefficient covariance matrix estimator than the intensity of heteroskedasticity.

Computing the leverage points is not a difficult task. If the  $QR$  decomposition is used as the method of computing the regression function Equation (2), then Davidson and Mackinnon (1993) provide a technique for calculating the leverage points,  $h_t$ , as a by-product.<sup>1</sup> For a detailed explanation please refer to the Appendix.

As already mentioned, both the alternative residuals have shown improved performance relative to the OLS residuals in previous empirical results in the cross-section literature. However, it must be shown that they both still lead to a consistent estimator for the population HAC estimator as described in Equation (3). To prove consistency using the alternative residuals, all that one must show is that either alternative residual is a consistent estimator for the true stochastic error term in Equation (17), since the alternative residuals are both

---

<sup>1</sup>Additionally Davidson and Mackinnon (1993) also note that another benefit in using the  $QR$  decomposition is that it produces the most accurate results.

just a transformation of the OLS residuals. This is stated in Proposition 1.

**Proposition 1.** *If standard regularity conditions for Least Squares estimation with mixing time series processes hold (for example Theorem 3.78 from White 2000), then the leverage points are asymptotically negligible, and as a consequence, the alternative residuals are a consistent estimator of the population error process.*

The implication of Proposition 1 is that existing proofs in the literature for the consistency of HAC covariance matrix estimators, see for example: Newey and West (1987), Andrews (1991), Hansen (1992), de Jong and Davidson (2000), de Jong (2000), or Jansson (2002), remain applicable. The proof is presented in Appendix.

## 3 Monte Carlo Simulation Analysis

### 3.1 Simulation Design

In my simulation experiments I follow the same basic set-up as first used in Andrews (1991) and which has also been employed by other studies related to investigating the performance of different HAC estimators (see for example, Andrews and Monahan 1992, West 1997, Kiefer et al. 2000, and Kuan and Hsieh 2008). All the simulations contain the following features: each predictor and the error term are modelled as an AR(1) or MA(1) process. The error term is then specified to be either: homoskedastic (HOM) or heteroskedastic by making the error term correlated with either the first non-constant predictor (HET1) or correlated with all the predictors (HET2). The method of doing this is explained shortly.

One potential limitation with the simulations used by the previous studies is that they all only considered Gaussian distributed random variables. In the literature related to simulations analysing the properties of HC covariance matrix estimators, Chesher and Jewitt (1987) recommended that the actual predictors (or design) matrix used in simulations should also be varied across different simulations. As has already been discussed, points of high leverage are one important source of downward bias in the OLS residuals and as a consequence the

associated HAC covariance matrix estimator. Hence, to get a more comprehensive analysis of the performance of different HAC covariance matrix estimators based on the alternative residuals it seems prudent to use distributions in which the generation of large leverage points are relatively more probable.

To achieve this, I follow Godfrey (2006) and also consider random variables generated from a Student-t distribution with 5 degrees of freedom as well as a Chi-squared distribution with 2 degrees of freedom, in addition to a Gaussian distribution (each of which are abbreviated to ‘ST’, ‘CS’, or ‘GA’ respectively in the tables and figures to follow). The Student-t distribution is used as an example of a symmetric non-Gaussian distribution with more mass located towards the tails of the distribution (i.e., ‘fat tailed’) and hence a greater probability of producing large deviations from the mean relative to a Gaussian distribution. In comparison, the Chi-squared distribution is an example of a non-normal and asymmetric distribution and so will provide the possibility of generating extreme values from only the right half of the distribution.

Before using any random variables I normalise them to have zero mean and unit variance. Let  $\nu$  represent the degrees of freedom parameter, then for the Student-t random variables, which already have a zero mean, this is achieved by dividing the generated random variables by  $\nu/(\nu - 2)$ . However, with the Chi-squared random variables it was necessary to also demean the random variables by subtracting  $\nu$  before dividing by  $2\nu$  to get a process with mean zero and unit variance.

In all simulations the linear regression model used is of the form:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \epsilon_t \quad t = 1, \dots, 128 \quad (17)$$

I set  $\beta_i$   $i = 0, \dots, 4$  to 1 in all experiments. Each predictor  $x_{ti}$   $i = 1, \dots, 4$  and the error process  $\epsilon_t$  are generated as follows: First a set of random variables,  $\xi_{ti}$ , where  $i = 1, \dots, 5$  and  $t = 1, \dots, 128$ , are generated from one of the three distributions (standardised with mean zero and unit variance) previously mentioned. The  $\xi_{ti}$  are then used to construct AR(1) or

MA(1) processes as follows:

$$\text{AR}(1) : z_t = \phi z_{t-1} + \eta(\xi_t) \quad \text{MA}(1) : z_t = \eta(\xi_t) + \theta \eta(\xi_{t-1}) \quad (18)$$

The variable,  $\eta$ , is a rescaling factor and is necessary to ensure the overall variance of the generated processes remains unity. To achieve this I give  $\eta$  the following structure depending on which process is being generated:

$$\eta = \begin{cases} (1 - \phi^2)^{1/2} & \text{if AR}(1) \text{ process} \\ (1 + \theta^2)^{-1/2} & \text{if MA}(1) \text{ process} \end{cases} \quad (19)$$

When generating either AR(1) or MA(1) model process, I use a range of values for both  $\phi$  and  $\theta \in \{-0.9, -0.8, \dots, 0.8, 0.9\}$ , giving 19 different model processes in total including the case of 0 (i.e., no dependence). Note the initial condition for all generated processes was drawn from the stationary distribution for either a AR(1) or MA(1) model. To aide comparison each simulation is grouped by the residual being analysed. Within each group I seed the random number generator with the same value so that the same stream of (pseudo) random numbers is generated. Doing this removes a possible source of difference between the alternative residuals and the OLS residuals and ensures greater comparability. In each new experiment a new set of predictors and error term is generated. Finally, as in Andrews (1991), I also transform the predictor matrix  $X$  before estimating the model to ensure that  $\frac{1}{T}X'X = I_T$ .

The errors generated by this method as just described are independent of the predictors and are therefore homoskedastic (HOM). In addition two different forms of heteroskedastic (HET) errors are then constructed by introducing multiplicative heteroskedasticity to the error term of the HOM model in Equation (17). This is achieved by multiplying  $\epsilon_t$  with  $k^{-1/2} \left| \sum_{i=1}^k x_{ti} \right|$  (i.e., only the non-constant predictors). This technique is standard in the HAC simulation literature and was first used by Andrews (1991), but has also been employed by Andrews and Monahan (1992), West (1997), Kiefer et al. (2000), and Kuan and Hsieh (2008). For the first form of heteroskedasticity (HET1)  $k = 1$ , which means the error



term is correlated with the first non-constant predictor ( $x_{t1}$ ) only. Whereas for the second form of heteroskedasticity (HET2)  $k = 4$ , and so the error term is correlated with all the non-constant predictors. Using this specification for the two forms of heteroskedasticity ensures that the variance of the overall error process generated remains unity for each form of heteroskedasticity investigated in the Monte Carlo simulations.

There are a number of choices that must be made when estimating the HAC covariance matrix for the estimated coefficients in Equation (17), such as which kernel to use, how to specify the bandwidth parameter and whether to apply a pre-whitening adjustment or not. Given that I am solely focused on comparing the performance of three different residuals, I seek to avoid using methods which might accentuate any differences between them. Kiefer et al. (2000) made the important observation that if an automatic bandwidth parameter and/or pre-whitening adjustment is used then different test statistics can result depending on any transformation applied to the data before estimating a regression model. This means that different test statistics could result which would impact the ability to fairly compare the three different residuals.

To overcome this potential invariance problem, they recommend using a fixed interval bandwidth parameter and avoid pre-whitening, which I have done in the following simulations. It is worth mentioning that the standard pre-whitening method as documented in Andrews and Monahan (1992) can impart additional bias into the estimation during the re-colouring stage. The source of this extra bias comes from the bias in estimating the VAR(1) model parameters (see the discussion in Sul et al. 2005). Additionally, Kuan and Hsieh (2008) suggest that pre-whitening can be avoided in practice based on the results of their simulations.

I focus primarily on using the BT kernel in the various simulations; however, the choice of kernel is less important so long as the same kernel is used for each residual. To determine the fixed interval bandwidth parameter I use the integer part of  $4(T/100)^{2/9}$ . As a check, I also consider simulations using the QS kernel and the Andrews (1991) data-dependent bandwidth using an AR(1) process estimated by regression. Generally, the estimated automatic bandwidth parameter showed only a relatively small difference between the three differ-

ent residuals across the experiments considered. Furthermore, the choice of kernel and/or method of determining the bandwidth parameter was found not to materially impact any of the simulation results. Nonetheless, I only present results based on the BT kernel and the fixed interval bandwidth (however, these extra results are available upon request).

To ascertain the finite sample properties of the two alternative residuals discussed in Section 2 relative to the OLS residual I undertake three different simulation experiments. The first experiment will be a fixed- $T$  simulation in which a range of comparison statistics for a single coefficient hypothesis test using a HAC estimator based on one of the three different residuals will be computed for a sample size of 128 observations. The second experiment entails an increasing- $T$  simulation. In this case the empirical size for single and a joint coefficient hypothesis test will be calculated as in the fixed- $T$  case but for an increasing sequence of sample sizes starting at 64 observations, ending at 512 observations, and increasing by one observation with each new simulation. The final experiment will investigate the empirical power of single and joint coefficient hypotheses tests based on a fixed sample size of 128 observations.

I choose to focus primarily on a sample size of 128 observation for the first and third experiments because it is a common length used in the simulation studies previously cited and is about the size of data sets frequently encountered in macroeconomics; being just over ten years for monthly data or 32 years for quarterly data. Additionally, this is a relatively small sample size and will highlight the impact of downward bias when estimating the coefficient covariance matrix using the OLS residuals.

In the fixed- $T$  simulation the main variable of interest is the estimated asymptotic variance of the OLS coefficient of the first non-constant predictor,  $x_1$ , that is, the ‘estimand’ is the (2, 2) element of the  $5 \times 5$  coefficient covariance matrix  $\text{Var} \left( \sqrt{T} \left( \hat{\beta}_T - \beta_0 \right) \right)$  which I will denote as  $\hat{v}_{22}$ . It has a true value of 1 throughout all the simulations undertaken, no matter which process or form of heteroskedasticity is used.

For each variance estimator and each scenario, the following statistics are also computed: the bias of the variance estimator, the variance estimator’s standard deviation, the root mean

squared error (RMSE) of the variance estimator, the empirical size, the average confidence interval width, the fixed interval bandwidth parameter used and the average maximal leverage point ( $h_t$ ) of the predictor matrix. To evaluate the empirical size, I conduct the following single coefficient (two-sided) hypothesis test:

$$\mathbb{H}_0: \beta_1 = 1 \tag{20}$$

$$\mathbb{H}_1: \beta_1 \neq 1 \tag{21}$$

The test statistic based on the asymptotic Gaussian approximation is given as:

$$z = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{1}{T}\hat{v}_{22}}} \tag{22}$$

And then compute the proportion of times out of the 10,000 repetitions the null hypothesis is rejected at the nominal size of  $\alpha = 0.05$ . Finally, the confidence interval width is defined as  $2 \left( z_{\alpha/2} \sqrt{\frac{1}{T}\hat{v}_{22}} \right)$ , with  $\alpha$  also set equal to 0.05.

The increasing- $T$  simulations have the same features as the fixed- $T$  simulations but only focuses on estimating the empirical size using a sample range of 64 to 512, incrementing the sample size by one observation with each new simulation (449 different simulations in total). This range of sample sizes will provide more granularity about the impact of sample size on empirical size estimation than previous simulations have been able to provide. To keep the number of simulations manageable I focus only on the heteroskedastic cases (HET1 and HET2) and use  $\phi = 0.5$  for the AR(1) process and  $\theta = 0.5$  for the MA(1) process. I use this value for both processes as it provides a moderate amount of dependence and at least with the autoregressive process, mimics the growth rates of some macroeconomic variables such as US GDP as noted by West (1997).

In addition to the single coefficient hypotheses test I also investigate a joint test that all the

non-constant coefficients are jointly equal to one:

$$\mathbb{H}_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1, \quad \text{against} \quad (23)$$

$$\mathbb{H}_1: \text{at least one } \beta_i \neq 1, \quad i = 1, \dots, 4 \quad (24)$$

To do this, I use the standard Wald test (see for example White 2000) defined as:

$$W = T \left( R\hat{\beta}_T - r \right)' \left( R\hat{V}_{\hat{\beta}_T}R' \right)^{-1} \left( R\hat{\beta}_T - r \right) \quad (25)$$

Where  $T$  is the sample size for that particular simulation (64, ..., 512),  $R$  is a  $4 \times 5$  dimension matrix defined as  $[0 : I_4]$ , and  $r$  is a  $4 \times 1$  column vector of 1s.  $\hat{\beta}_T$  are the OLS coefficient estimates and  $\hat{V}_{\hat{\beta}_T}$  is the associated HAC estimated coefficient covariance matrix. The empirical size is then computed as was the case in the fixed- $T$  simulations using a nominal size  $\alpha = 0.05$ .

The final experiment focuses on comparing empirical power for tests using the three different residuals. I use the same model as in Equation (17) as well as the same data generation processes used in the fixed- $T$  and increasing- $T$  simulations. Again, I focus exclusively on a model parameter of 0.5 for the AR(1) and MA(1) generated processes. Each simulation used one of the three different distributions considered and only the two forms of heteroskedasticity (HET1 and HET2) were examined. I undertake the same two hypotheses tests as in the increasing- $T$  experiment: one involving the first non-constant predictor and the second involving all the non-constant predictors. In each case, the model is generated with a set of alternative coefficient values  $\beta_i$   $i = 0, \dots, 4$  and  $\beta_i$  is given a value from an equally spaced grid of 100 values from 0.01 to 2 (with 1 the true value under the null hypothesis). I then test the following two hypotheses: 1)  $\mathbb{H}_0: \beta_1 = 1$  using a standard  $t$ -test as described by Equation (22) and 2)  $\mathbb{H}_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$  using the Wald test described by Equation (25) and calculate the frequency of rejections of the null hypothesis for each test using a nominal size of  $\alpha = 0.05$  and 10,000 repetitions.

Given the large number of simulations to be undertaken, all computations were programmed

in C++ making extensive use of the ‘Armadillo’ template Linear Algebra library (Sanderson 2010), which provides a convenient interface to optimised (and multi-threaded) linear algebra libraries such as OpenBLAS.<sup>2</sup>

## 3.2 Simulation Results

As a starting point it will be useful to compare the performance of the two alternative residuals relative to the OLS residual when the underlying model analysed uses independently and identically distributed (*iid*) data. To do this I generate two sets of 128 random variables from the three different distributions and then construct the responses,  $Y$ , using the predictors,  $X$ , the error process,  $\epsilon$ , and with all the  $\beta_i$  set equal to 1 as specified by Equation (17). The model is then estimated by regression and I compute the suite of statistics discussed previously in relation to the fixed- $T$  simulations using the three different types of residuals. The results are presented in Table 1.

From Table 1 we can make the following comments: all the residuals produce estimands which are biased. As was discussed, the OLS residuals produce estimates of  $\hat{v}_{22}$  which are negatively biased and the amount of the bias *increases* as the average maximum leverage point increases such as with the two non-Gaussian distributions. In comparison, both of the alternative residuals estimands are positively biased. The impact of the different directions of bias becomes apparent when looking at the empirical sizes produced using each residual. The OLS residual results in oversized test statistics in all cases. In contrast, the two alternative residuals produce test statistics with empirical sizes much closer to the true nominal size across the three different distributions considered.

---

<sup>2</sup>OpenBLAS is an open source C library which provides access to BLAS and LAPACK. These are the standard numerical linear algebra libraries used extensively by many modern numerical computing software such as R or MATLAB. It can be accessed from <http://www.openblas.net/>.

Table 1: Alternative Residual Estimator Comparison

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	max $h_t$
Gaussian distributed random variables								
$\hat{\epsilon}$	0.953	-0.047	0.275	0.279	0.063	0.335	4	0.123
$\tilde{\epsilon}$	1.065	0.065	0.312	0.319	0.052	0.354	4	0.123
$\dot{\epsilon}_m$	1.082	0.082	0.320	0.330	0.050	0.357	4	0.123
Student-t distributed random variables with degrees of freedom = 5								
$\hat{\epsilon}$	0.933	-0.067	0.464	0.469	0.065	0.327	4	0.231
$\tilde{\epsilon}$	1.100	0.100	0.656	0.663	0.051	0.353	4	0.231
$\dot{\epsilon}_m$	1.143	0.143	0.782	0.795	0.048	0.359	4	0.231
Chi-squared distributed random variables with degrees of freedom = 2								
$\hat{\epsilon}$	0.916	-0.084	0.538	0.544	0.080	0.321	4	0.265
$\tilde{\epsilon}$	1.112	0.112	0.754	0.762	0.061	0.351	4	0.265
$\dot{\epsilon}_m$	1.164	0.164	0.840	0.856	0.058	0.358	4	0.265

*Note:* Numbers represent simulation results from 10,000 replications with *iid* data. In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{v}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘max  $h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Fixed- $T$  simulation plots related to the empirical size simulations for AR(1) and MA(1) processes using the three different residuals are presented in Figures 1 and 2 respectively. The first column corresponds to Gaussian distributed simulations, the second column relates Student-t (with 5 degrees of freedom) distributed simulations, and the last column corresponds to Chi-squared (with 2 degrees of freedom) distributed simulation results. The first row relates to errors which are homoskedastic, while rows two and three correspond to errors with heteroskedasticity specified by HET1 and HET2 respectively. For comparison purposes the nominal size used ( $\alpha = 0.05$ ) as well as the average maximum leverage point for each simulation is also plotted. More detailed fixed- $T$  simulation results related to the single coefficient test case similar to Table 1 are provided in Tables A1–A6 in the Appendix.

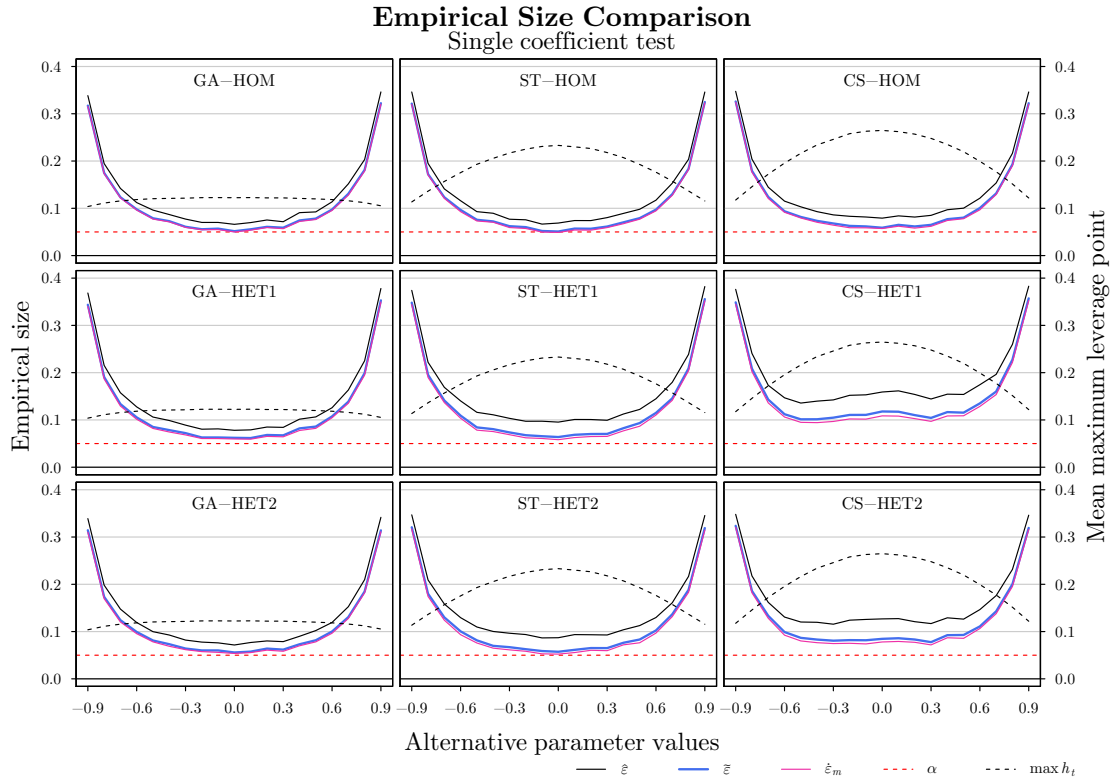


Figure 1: AR(1) Process with  $T = 128$  and  $N = 4$

It is clear that in all of the AR(1) fixed- $T$  simulations the empirical size computed using the two alternative residuals outperform the OLS residuals, particularly with lower levels of dependence. The difference is largest for the two non-Gaussian distributions which are associated with larger average maximum leverage points as well. It is interesting to observe how all the empirical sizes appear symmetric for all the process parameter values analysed, with the centre located at a parameter value of 0. Using the Hoaglin and Welsch (1978) ‘rule of thumb’ to distinguish points of high leverage which for the present case is equal to 0.0625, indicates many of the fixed- $T$  empirical size simulations are potentially being affected. The degree of heteroskedasticity present in the errors also seems to be linked to the difference in the empirical sizes obtained from the alternative residuals and the OLS residuals with a larger gap associated with HET1 and HET2 compared to the homoskedastic case. However, the gap between the OLS residuals and the two other residuals empirical sizes narrows for more highly persistent AR processes with  $\phi > \pm 0.8$ . The reason why this happens might have something to do with changes that occur to the leverage points for different values of  $\phi$ . Notice also how the average maximum leverage point has a concave parabolic shape over

the set of parameter values, reaching a maximum value at  $\phi = 0$  (which is opposite of the empirical size results). This point is the same point where the gap between the empirical sizes is largest as well.

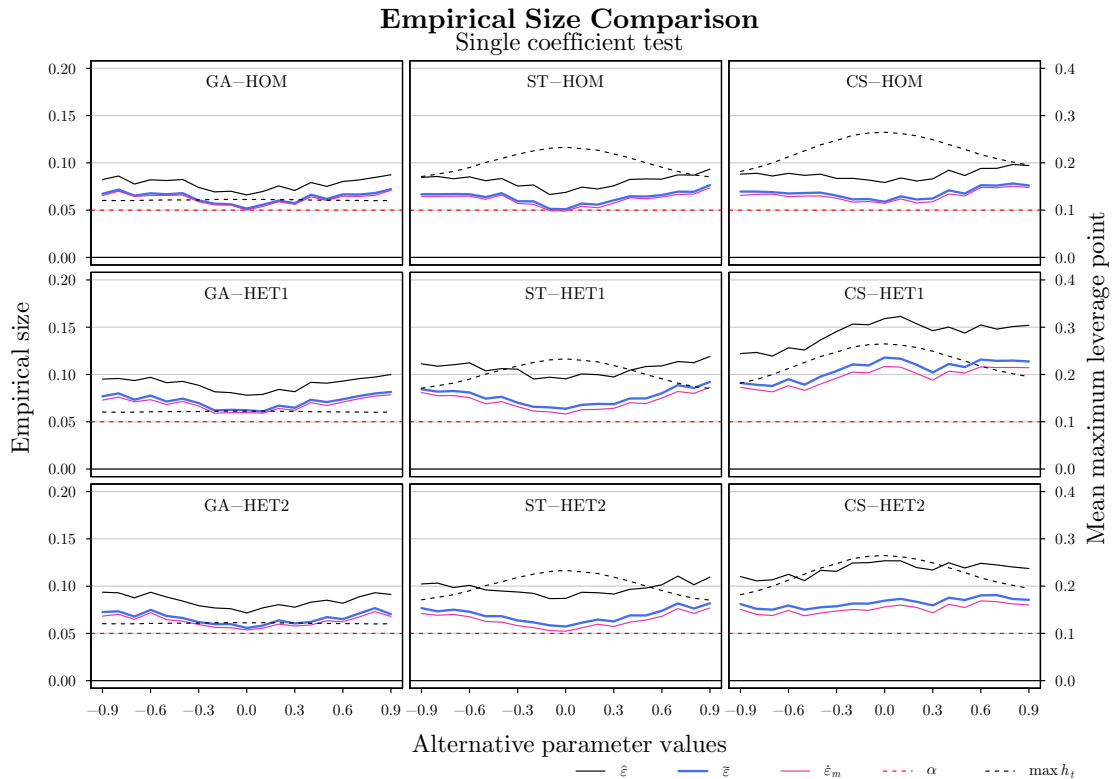


Figure 2: MA(1) Process with  $T = 128$  and  $N = 4$

The fixed- $T$  simulation results for the MA(1) processes are generally consistent with those of the AR(1) results, except the performance of the two alternative residuals relative to the OLS residual does not deteriorate as  $\theta$  increases towards the endpoints of the parameter space considered. Furthermore, the gap between the empirical sizes for each residual appears to remain relatively steady for all the parameter values analysed.

For both processes, the modified discounted residuals perform marginally better than the prediction error residuals, especially for the non-Gaussian distributions and for parameter values for  $\phi$  and  $\theta$  close to 0. This reflects the findings of Cribari-Neto and da Silva (2011) who also found only a marginal improvement in empirical size performance between the two in relation to cross-section data. It could be argued that this is due to the fact that the maximum discount factor that can be applied by the modified discounted error residuals method is 2.5 while that of the prediction error residuals is only slightly lower at 2.



Finally, I briefly summarise the results from Tables A1–A4. The OLS residual always provides the most oversized and most downward biased estimand while the modified discounted residual always produced the least oversized and the most upward biased estimate for the variance term  $\hat{v}_{22}$ . Recall, the more upward or positive biased a variance estimator is, the less likely it is to produce oversized test statistics. The bias is largest for Chi-squared distributed errors with HET1 and smallest for Gaussian homoskedastic errors. The standard deviation and RMSE of the estimand is always largest for the modified discounted error residual and smallest for the OLS residual and these increase when the error term has HET1-type multiplicative heteroskedasticity.

The fixed- $T$  simulations help illustrate just how points of high leverage in the predictor matrix can adversely impact coefficient variance estimation and the test statistics computed from them. However, when the sample size increases, we would expect to see some reduction in this and therefore some convergence in the empirical sizes resulting from the three different residuals. This is the main reason for undertaking the increasing- $T$  simulations. The empirical size results are displayed in Figure 3 for the single coefficient test and Figure 4 for the joint coefficient test. In both cases a Gaussian distribution was assumed and the nominal size ( $\alpha$ ) was set to 0.05. Results for the two alternative distributions are illustrated in Figures A1–A4 in the Appendix.

While there is a noticeable amount of variability in the three plots, all three residuals produce over sized test statistics for the single coefficient test. Nonetheless, the OLS residuals are dominated by the two alternative residuals, especially at smaller sample sizes (less than 200 observations). Each of the three residuals show a downward movement towards the nominal size of 0.05 as the sample size increases; however, the OLS residuals appear to asymptote relatively faster than the other two residuals.

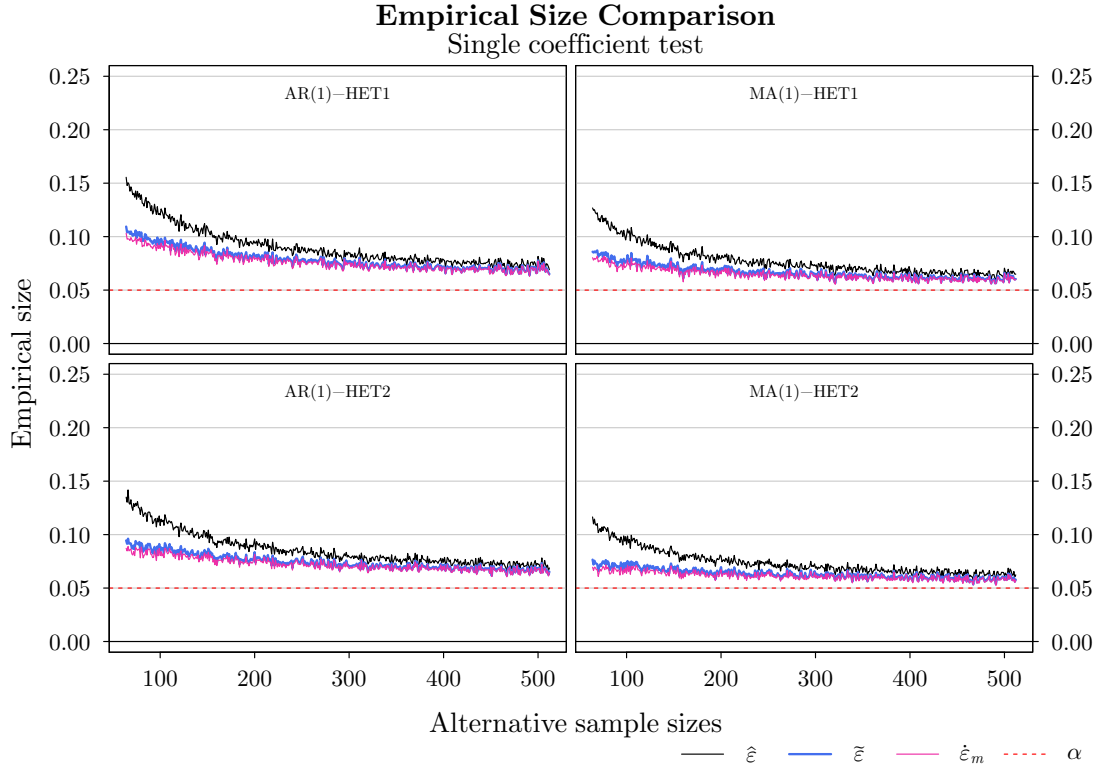


Figure 3: Gaussian AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

This finding seems to be because of the increasing sample size reducing the impact of larger leverage points on the OLS Residuals. Recall, the transformation process applied to  $\hat{\epsilon}$  in order to generate the two alternative residuals has already offset this effect to some degree. The AR(1) processes appear to generate the largest oversized test statistics, while the MA(1) processes appear to cause a smaller deterioration in the empirical sizes of the test statistics across all three residuals and was also evident when comparing Figures 1 and 2. For sample sizes less than approximately 200 observations the modified discounted residuals out performs the prediction error residuals by a small margin, but this difference disappears as the sample size increases. These findings also carry over to the Student-t and Chi-squared distributions as well, where the oversized problem in small samples for all three residuals is more evident.

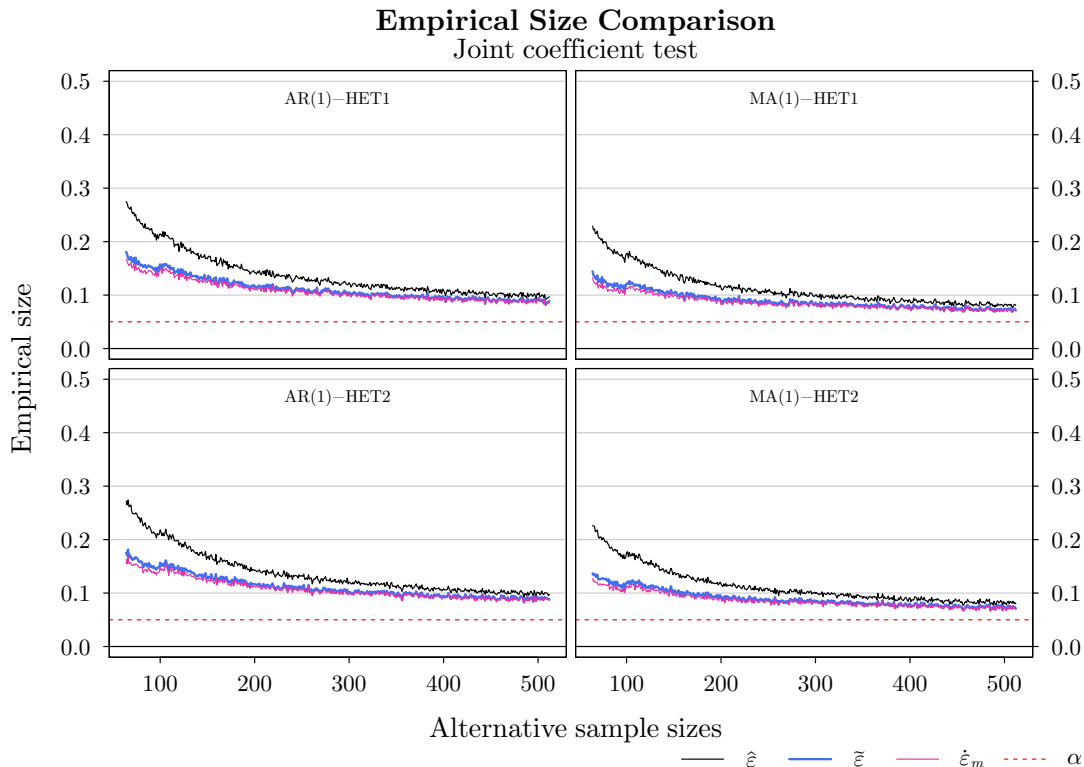


Figure 4: Gaussian AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

Comparable to Figure 3, all residuals produce Wald test statistics that are significantly oversized.<sup>3</sup> This finding adds to the list of previous studies which have discussed the poor empirical size properties of Wald test statistics (see Chesher and Austin 1991 for the cross-section case and Burnside and Eichenbaum 1996, Christiano and den Haan 1996, and Kiefer et al. 2000 for the time series case).

What is notable; however, is the scale of the deterioration in the empirical size of the Wald test statistic based on the OLS residuals when the data are both dependent and heteroskedastic such as with the AR(1)-HET1 simulation. For this experiment the empirical size is almost *five* times that of the true nominal size. These findings indicate that joint tests, even with only a small number of coefficients, are likely to be very unreliable when the sample size is relatively small to modest and the data have some dependence and/or heteroskedasticity. For larger sample sizes (such as 500 observations and above), there is some

<sup>3</sup>Note, there are two up-ticks in the empirical size plots located around sample sizes of 100 and 273 observations. At these two points the fixed interval bandwidth, which is a function of  $T$ , increases from 3 to 4 and from 4 to 5 respectively. Burnside and Eichenbaum (1996) suggest that an increase in the bandwidth parameter used in estimation, all else equal, will cause a deterioration in the performance of the Wald test statistic.

improvement evident in the empirical size estimated using all three residuals, although the OLS residuals continue to remain the worst performer. Similar to the single coefficient case the small out performance noted for the modified discounted residuals over the prediction error residuals appears to disappear at larger samples sizes. The joint test results also indicate that size distortions appear to be more of an issue for the AR(1) processes than the MA(1) processes as was also the case with the single coefficient test case. The two non-Gaussian distributions show similar results in relation to the relative empirical size performance of the three different residuals as in the Gaussian case.

It is interesting that the gap between the empirical size estimated using the OLS residual and the other two residuals for all the different cases considered persists for a majority of the sample sizes investigated. Indeed, even up to a sample size as large as 500 observations there still remains a small difference between the OLS residuals and the two alternatives. This indicates that when the data are moderately dependent and possibly heteroskedastic then there is a real benefit from using the alternative residuals in estimating the HAC covariance matrix instead of relying on the OLS residuals. It is interesting to contrast this finding about sample size with those of Long and Ervin (2000) in the HC literature. Their results lead them to recommend using the prediction error residuals for (*iid* cross-section data) sample sizes of around 250 observations. In the case of time series data, it would appear that when there is moderate dependence in the data, then a similar type recommendation would be to use the alternative residual-based estimators for sample sizes of up to 500 observation in the time dimension, approximately twice as large as their number in the cross-section.

The proof of consistency of the alternative residuals in the Appendix relies on the asymptotic negligibility of the leverage points. While the preceding results appears to support this, I now investigate this using simulation. To do this I undertake three ‘large sample’ experiments using sample sizes of 1,000, 5,000 and 10,000 observations. Each regression model continues to have four predictors plus a constant as with the previous simulations. I again consider three distributions and use the BT kernel with a fixed interval bandwidth parameter. The model used for each simulation is an AR(1) with  $\phi = 0.5$  and the form of heteroskedasticity is HET1. These specifications were chosen because autoregressive models and heteroskedas-

ticity of the form of HET1 tended to result in the worst empirical size performance in the simulations when using a sample size of 128. The simulation results comparing the OLS residuals with the two alternative residuals are presented in Table A7 in the Appendix.

Comparing these results to those in Table A1 which uses a sample size of 128 (also in the Appendix), it becomes clear that increasing the sample size does materially reduce the magnitude of the average maximum leverage point observed. For example, in the Gaussian distribution case, the average maximum leverage point from a sample size of 128 is 0.12, in addition the estimated empirical size for the comparable simulation set-up is 0.102 for  $\hat{\epsilon}$ , 0.086 for  $\tilde{\epsilon}$ , and 0.082 for  $\dot{\epsilon}_m$ . In contrast, the average maximum leverage point attained declines to only 0.003 when the sample size is increased to 10,000 observations while the estimated empirical size for each residual is approximately the same value and all are close to the true nominal size of 0.05.

Having established that the alternative residuals offer superior empirical size performance relative to the traditional OLS residuals, I now move to see if this result comes at any loss of test power. Size-corrected power curves for the single coefficient test and Gaussian distributed random variables are presented in Figure 5, while those for the joint coefficient test are presented in Figure 6. Size-corrected power curves of Student-t and Chi-squared distributed random variables are presented in the Appendix in Figures A1–A4.

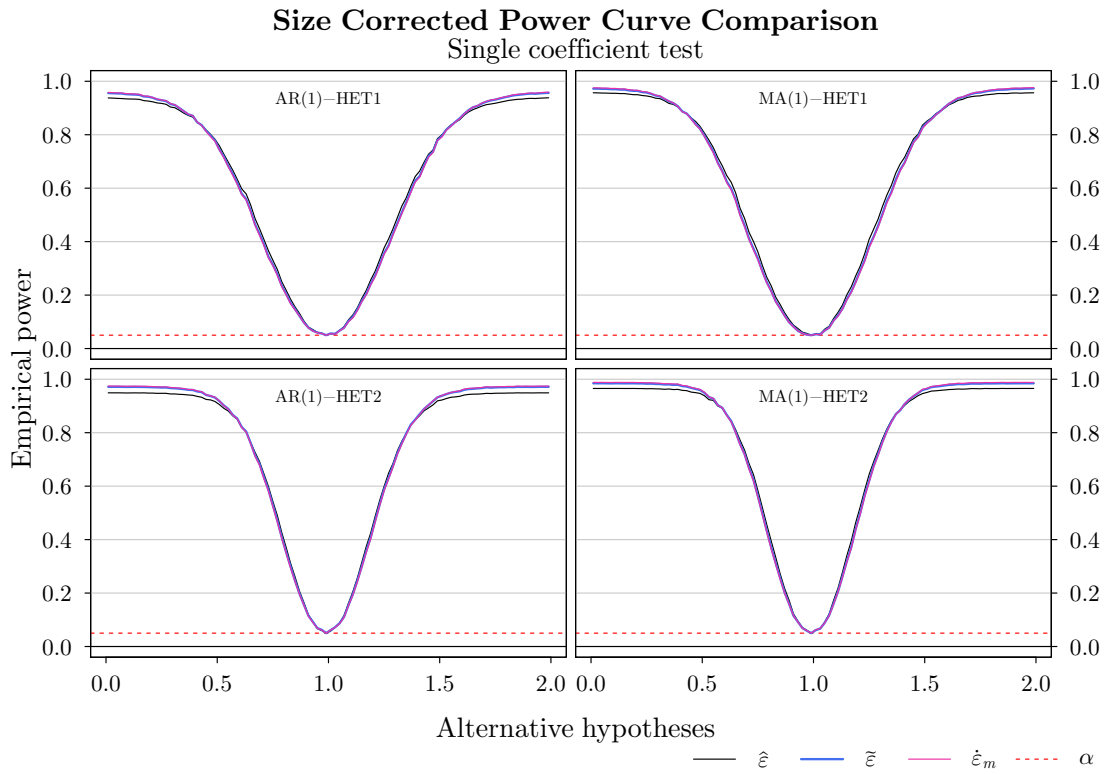


Figure 5: Gaussian AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

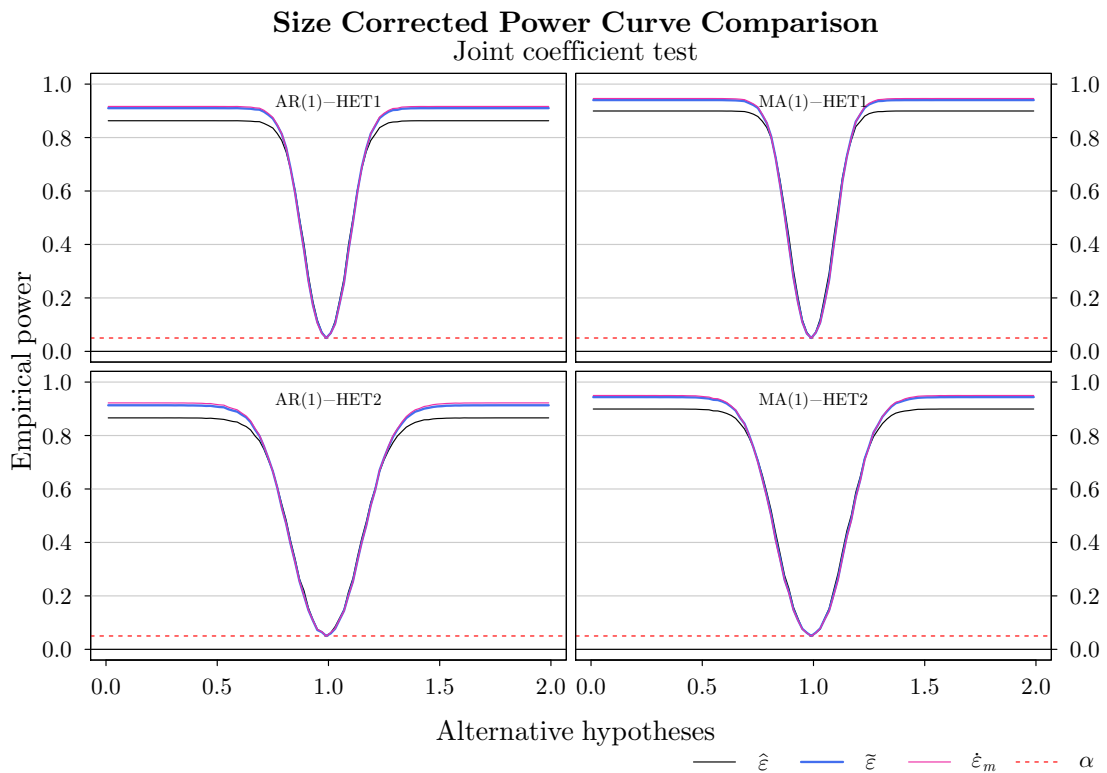


Figure 6: Gaussian AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

All the plots illustrate that even with different processes as well as different types of het-

eroskedasticity in the model's error term, there is no apparent change in test power when using either the prediction error residuals or the modified discounted error residuals in place of the OLS residuals.

The results from all three different simulations presented so far document the better performance of the two alternative residual estimators relative to the OLS residuals in the sense that their empirical size is closest to the true nominal size obtained from the approximating standard normal distribution relative to the OLS residual without any loss in power. However, it is important to highlight some of the (potentially) less attractive attributes of the two alternative residual estimators as well. Generally the alternative residual estimators achieve their improved empirical size results by having more variation in the estimate of the variance term compared to the OLS residual-based estimate. As a consequence, this leads to wider estimated coefficient confidence intervals as well.

The impact of this to confidence interval widths for different model parameter values, distributions and forms of heteroskedasticity for the prediction error and modified discounted error residuals is illustrated in Figures 7 and 8. The plots display average coefficient confidence interval widths estimated using the prediction error and the modified discounted error residuals as a ratio of the OLS residual-based confidence interval width. A value greater than one indicates that the estimated width is longer than the width estimated using the OLS residual on average.

In all cases investigated both alternative residuals have longer confidence intervals relative to OLS residuals, ranging from just under 10% longer for Gaussian distributed variables to just over 20% longer for Chi-squared distributed variables. Furthermore, the estimated average widths for the two variables is largest for values of  $\phi$  or  $\theta$  closer to 0, which corresponds to the region where the average maximum leverage point reaches its largest value as well. However, such a result is not really a negative though if we think about why we estimate variances in the first place. We want these quantities to compute test statistics. Hence, what is important is that the test statistic leads to the correct inference, that is to say, the test statistic is corrected sized. This is the argument used by Simonoff (1993) when he suggested

that it was advisable to use a variance estimator that might be more variable if it gives the right inference compared to using a less variable estimator that performs relatively poorly in inference.

Comparing the prediction error residuals to the discounted residuals, we see that the former always has a smaller relative confidence interval width in all simulations undertaken. One reason that the modified discounted error residual has more variation could be because the discount factor used in its construction can take one of four possible values. Whereas, the prediction error's 'discount factor' is a constant value.

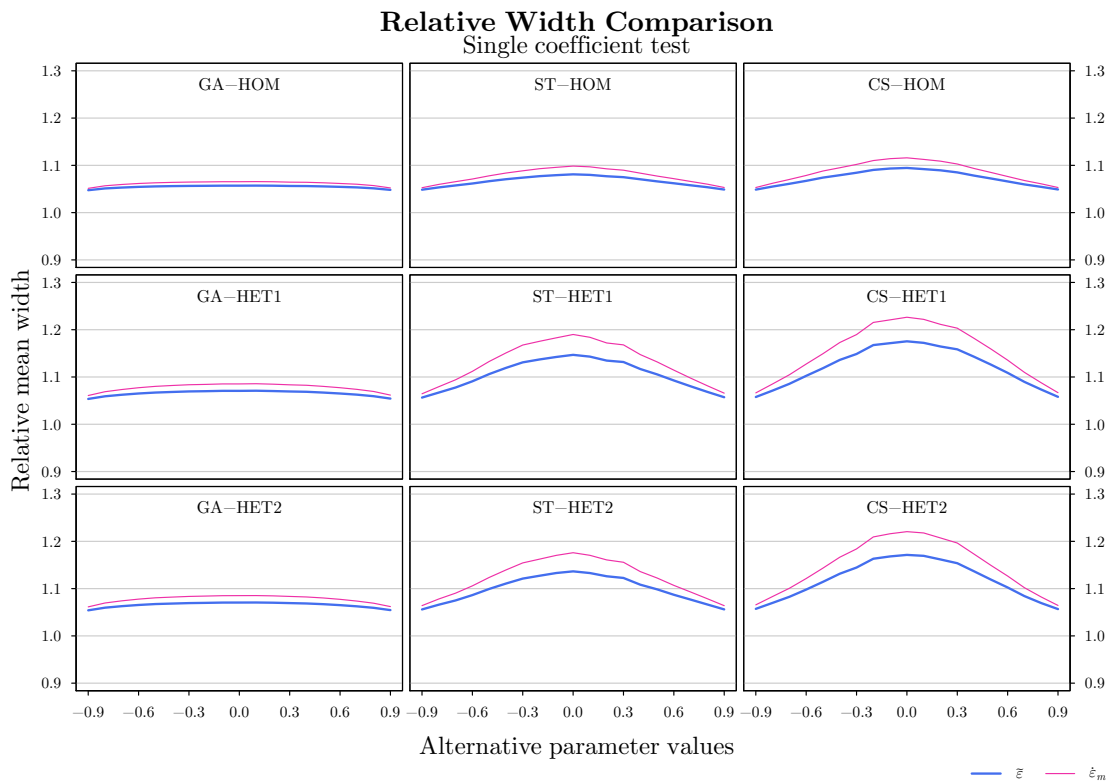


Figure 7: AR(1) Process with  $T = 128$  and  $N = 4$



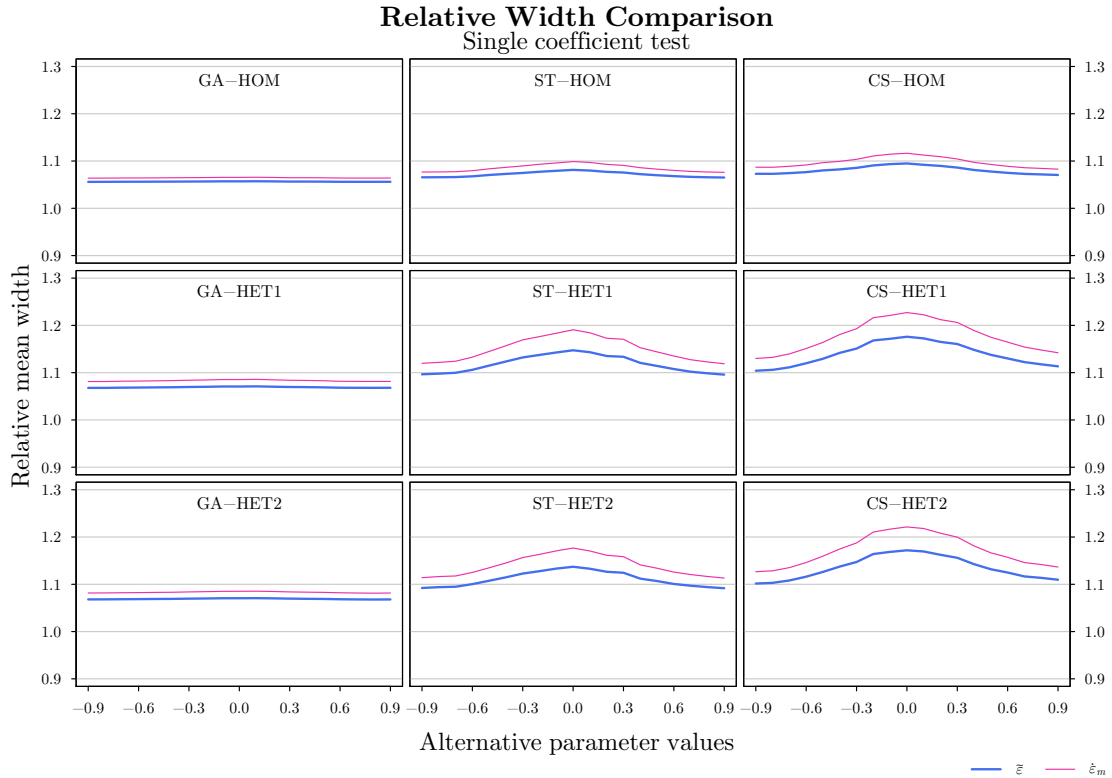


Figure 8: MA(1) Process with  $T = 128$  and  $N = 4$

## 4 Empirical Application

### 4.1 Exchange Rate Forecasting Evaluation

In this Section I use an empirical application to demonstrate how inference can differ depending on which residual is used in estimating the HAC covariance matrix. Differences in inference are important not only from a statistical point of view but also because they could have a real economic impact as well. To do this I replicate the forecasting accuracy comparison exercise from Diebold and Mariano (1995) where a consistent estimate of the long run covariance matrix is necessary to compute the DM test statistic. However, I focus on the UK Pound Sterling/US Dollar exchange rate (UKP/USD) instead of using the Dutch guilder/US Dollar exchange rate as Diebold and Mariano (1995) originally did since it ceased trading when the Euro was introduced in 2002.

There are two reasons for focusing on the UKP/USD exchange rate. Firstly, the data are

easily obtainable from the Bank of England's (BOE) web site.<sup>4</sup> Secondly, the period of early 1990s saw significant instability in the UK Pound exchange rate as a result of the UK first joining the European Exchange Rate Mechanism in October 1990 and then its subsequent exit two years later in September 1992. Such dramatic events are likely to result in large outliers which could potentially be points of high leverage.

The series to be forecast is the three-month change (measured monthly) in the nominal UKP/USD end of month spot exchange rate (in US cents). I will compare the accuracy of two forecasts of the UKP/USD exchange rate. The first forecast assumes no change in the exchange rate (i.e., a random walk model), whereas the second is the forecast implicit in the three-month UKP/USD forward rate and is computed as the difference between the three-month forward rate and the spot rate. Diebold and Mariano (1995) considered the period 1977:01–1991:12, giving them 180 observations. The exchange rate data from the BOE on the forward rates for the UKP/USD exchange rate begin in 1979:01, so I use this date as my starting point and I set the last observation to be 2015:12, giving a total of 443 observations. The three time series are plotted in Figure 9. Note the large amount of variability in the spot UKP/USD exchange rate, especially during the early 1990s for the reasons previously highlighted.

When testing for equal forecast accuracy Diebold and Mariano (1995) stressed that any loss type could be used; however, in the empirical application the authors only considered a symmetric loss function (absolute loss). However, potential users of forecasts for the exchange rate, such as by exporters or importers for instance, are more likely to be concerned about unanticipated appreciations or depreciations. For example, an exporter located in the UK will be more concerned with positive forecast errors as opposed to negative forecast errors as the former represents an appreciation of the UK Pound relative to the US Dollar while the latter reflects a depreciation of the UK Pound relative to the US Dollar. By the same token, an importer in the UK will be more concerned with negative forecast errors than positive forecast errors since these represent a depreciation of the UK Pound relative to the US Dollar.

---

<sup>4</sup><http://www.bankofengland.co.uk/>

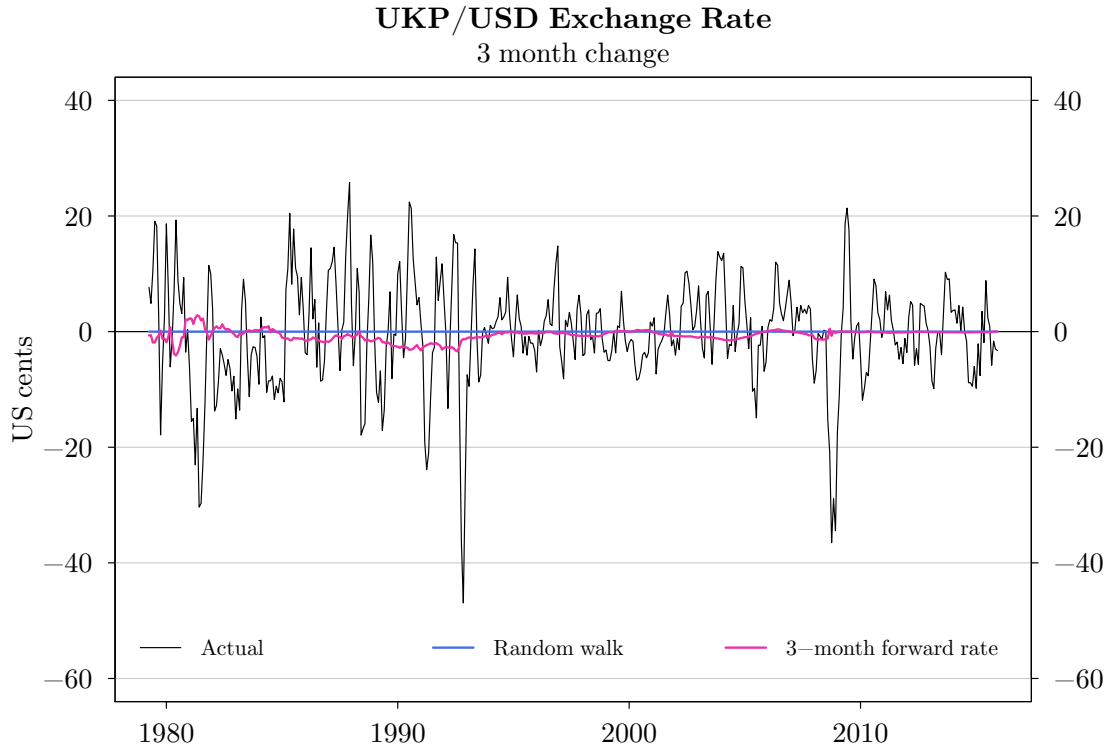


Figure 9: Exchange Rate Forecasts

In this context, it is more realistic to compare forecasting accuracy using an asymmetric loss function. One such loss type is the ‘Linex’ loss function which has the following form:

$$L(e_t) = \exp(\theta e_t) - \theta e_t - 1 \quad (26)$$

The parameter  $\theta$  controls the amount of asymmetry in the loss function. If  $\theta > 0$ , there are large losses from positive forecast errors and the losses increase in magnitude the bigger the value of  $\theta$ . Whereas, if  $\theta < 0$ , large losses are incurred from negative forecast errors and the losses increase the smaller the value of  $\theta$  is (for a detailed discussion see Elliot and Timmermann 2004). Plots for the loss differential using the Linex loss function and values of  $\theta$  equal to  $\pm 0.2$  are presented in Figure 10a for positive losses and Figure 10b for negative losses. These values for  $\theta$  were selected after considered a range of values from  $\pm 0.1$ – $0.5$  because they generated outliers without being too large in absolute terms. The resultant loss differentials should cause a greater downward bias impact to the OLS residual-based HAC estimated standard error than to the estimator using the alternative residuals.

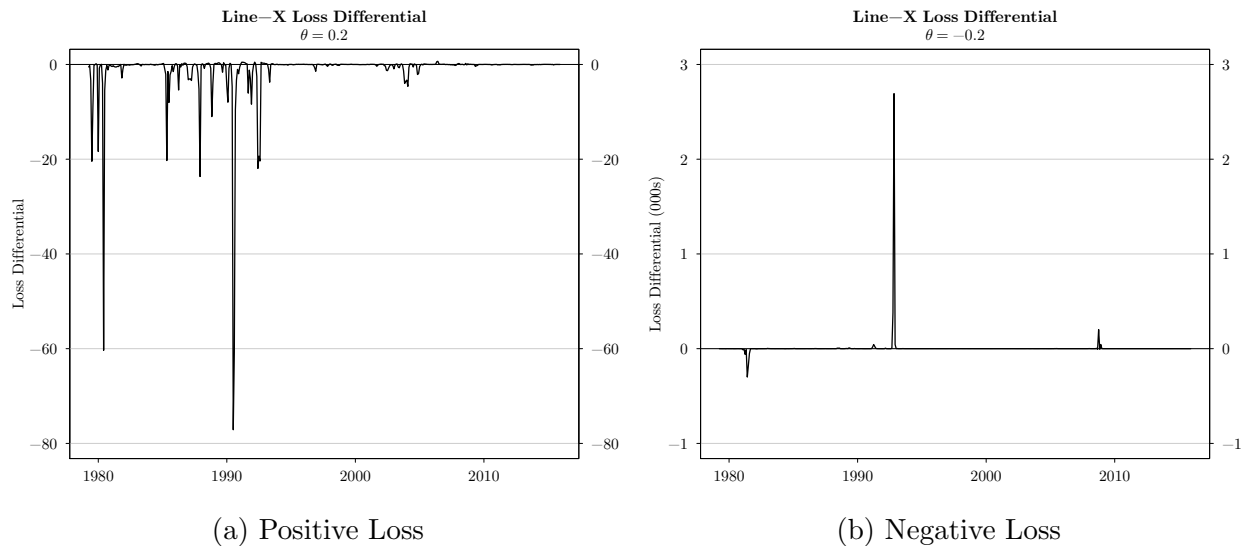


Figure 10: Asymmetric Loss Function Comparison

The DM test involves regressing each of the two forecast loss differentials in Figure 10 on a constant and test if the estimated constant is statistically different from zero using a  $t$ -test computed with a HAC standard error. To be agnostic about the time period considered or sample size used, I compute a sequence of DM test statistics from a series of rolling regressions using different window lengths, starting at 40 observations and increasing by 20 observations until reaching 180 observations, over the full sample and then calculate the proportion of rejections using a nominal size of  $\alpha = 0.05$ . This same procedure is done for each of the three different residuals and when computing the HAC standard error of the constant term in the regression I use a BT kernel with a fixed interval bandwidth. This is also the same set-up used in Diebold and Mariano (1995). The results are displayed in Table 2.

Table 2: Proportion of DM Test Rejections

Residual	Rolling Regression Window Length							
	40	60	80	100	120	140	160	180
$\theta = 0.2$								
$\hat{\epsilon}$	0.26	0.34	0.49	0.49	0.55	0.59	0.74	0.81
$\tilde{\epsilon}$	0.24	0.30	0.47	0.45	0.51	0.54	0.67	0.74
$\dot{\epsilon}_m$	0.24	0.30	0.47	0.45	0.51	0.54	0.67	0.74
$\theta = -0.2$								
$\hat{\epsilon}$	0.02	0.04	0.01	0.00	0.00	0.00	0.00	0.00
$\tilde{\epsilon}$	0.01	0.03	0.01	0.00	0.00	0.00	0.00	0.00
$\dot{\epsilon}_m$	0.01	0.03	0.01	0.00	0.00	0.00	0.00	0.00

*Note:* Numbers represent the proportion of rejections of the null hypothesis of equal predictive accuracy at  $\alpha = 0.05$  for a given rolling regression window length and residual estimator. The HAC standard errors were computed using a BT kernel and a fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$ .

DM tests using a HAC standard error computed with either the prediction error or the modified discounted error residuals result in fewer rejects of the null hypothesis of equal predictive accuracy for both the positive and negative loss differentials. Also, the results for the two alternative residuals are identical for both the positive and negative loss differential cases and this could be because of the reasons previously highlighted. The evidence is stronger for the positive loss differential test results than it is for the negative loss differential test results and might be a consequence of there being only a few non zero values for the negative loss differential.

Focusing on the positive loss differential DM test results, the proportion of rejections starts out fairly similar when using a rolling window size equal to 40 observations; nonetheless, the proportion of rejections obtained by the OLS residuals relative to the other two residuals starts to diverge as the sample size increase. One explanation for this is finding is that using a bigger window length means that the more extreme values will remain in the sample used to compute the DM test statistic for larger proportion of the tests undertaken. In comparison, with a smaller sample size, these same large values will more quickly drop out of the test computation as the window moves through the full sample. Finally, the results highlight how inference could change depending on which residual is used to estimate the HAC standard error and in some cases this difference could also have economic consequences.

## 5 Implications and Recommendations

The simulation and empirical results provides evidence towards using either of the alternative residuals in place of the OLS residuals in HAC covariance matrix estimation, especially for relatively small samples and data with outliers and/or asymmetries. Test statistics computed using HAC covariance matrix estimators using the alternative residuals suffer less size distortions relative to those based on the OLS residuals without any loss of power. This finding indicates that alleviating the downward bias from the OLS residuals is more important than mean squared error minimisation in the estimation of the HAC covariance matrix to achieve a more accurate test size and supports the argument of Simonoff (1993) in relation to variance estimators in general.

The simulation results related to joint coefficient testing provides additional evidence that researchers should be cautious about using Wald-type joint tests even with a small number of restrictions when the sample size they are working with is relatively small and the data being investigated are weakly dependent and/or heteroskedastic.

Making the alternative residuals available in exiting statistical software would not be overly difficult either. Most statistical programs use the  $QR$  decomposition as the method of computing the solution to the Least Squares estimation problem. Furthermore, it is often the case that software which does provide robust covariance matrix estimation tends to offer both HC and HAC estimators, and so routines already exist to compute the alternative residuals at least for the HC version of the covariance matrix estimator (see for instance the ‘Sandwich’ R package of Zeileis 2004).

Of the two different alternative residuals, the modified discounted errors tended to perform slightly better than the prediction errors in small samples and when the data have points of high leverage. However, this gap quickly reduces as the sample size increases and comes at a price of noticeable larger variance estimates and hence wider confidence intervals. This increased variability seems to come from the discounting term used to compute the modified discounted error residual taking one of four possible values.

Given the improvements in empirical size from using the modified discounted error compared to the prediction error are not too great (indeed, they were identical in the empirical application of Section 4) and the since the former requires slightly more involved computations, it would seem that the prediction errors would generally be suitable for most empirical applications in practice.

The difference in inference that resulted from replacing the OLS residuals with the alternative residuals in the Wald-type tests in the simulations and empirical application raises the possibility that similar differences might be observed in tests which use the OLS residuals explicitly such as in Likelihood Ratio or Lagrange Multiplier tests. This might be an interesting avenue for future research.

Finally, it is worth emphasising that if there is a very high degree of dependence in the data being analysed, then it is not appropriate to rely on HAC estimators to provide correctly sized test statistics, even in large samples. In these situations, it is recommended that researchers try specifying the dynamics of their model more fully (such as including relevant lagged variables in the model) or investigate alternative parametric methods such as Feasible Generalised Least Squares (see for example Hamilton 1994).

## 6 Conclusion

In the presence of heteroskedasticity and/or autocorrelation the traditional coefficient covariance matrix is inconsistent and hence standard inference is incorrect. As an alternative, the HAC coefficient covariance matrix can be used and is consistent to both heteroskedasticity and autocorrelation of unknown forms. The problem is that in finite samples the performance of the HAC estimator can differ greatly and has been shown to produce oversized test statistics.

One of the primary reasons for the oversized test statistics is the use of the OLS residuals in the construction of the coefficient covariance matrix which tend to downward biased in finite samples. The degree of bias is related to how balanced or unbalanced the predictor

matrix is as indicated by the presence or absence of high leverage points. High points of leverage causes the OLS residuals to be smaller than otherwise and this leads to smaller error variance estimates.

The downward bias in the OLS residuals can be partially offset using well known methods from the related HC literature. The better performing alternatives include the prediction error and modified discounted error residuals. Both use a transformation of the OLS residuals which inflate those residuals linked to high points of leverage and thereby result in more correctly sized test statistics. To date, these HC related residuals have not been analysed in relation to HAC estimation, although two related studies have shown that offsetting the OLS residual bias does help to improve the empirical size of test statistics.

A suite of simulations combined with an empirical application provides strong support for replacing the OLS residual with either the prediction error or modified discounted error residual in the construction of the HAC estimator. Both alternative residuals help to reduce the tendency for HAC covariance matrix estimators to produce oversized test statistics in finite samples. This recommendation is especially important in sample sizes less than 500 observations as highlighted in the simulation experiments with error processes having heteroskedasticity and moderate amounts of dependence.

The difference in inference from using the OLS residuals or one of the two alternative residuals was shown to also have potentially real economic significance as well. Using a forecasting comparison application with data containing large outliers it was found that there was a greater tendency to reject the null hypothesis of equal predictive accuracy when using the OLS residuals than was the case when using the two alternative residuals.



## References

- Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59(3), 817–858.
- Andrews, D. W. K. and J. C. Monahan (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60(4), 953–966.
- Burnside, C. and M. Eichenbaum (1996). Small-Sample Properties of GMM-Based Wald Tests. *Journal of Business & Economic Statistics* 14(3), 294–308.
- Chesher, A. and G. Austin (1991). The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics. *Journal of Econometrics* 47(1), 153–173.
- Chesher, A. and I. Jewitt (1987). The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator. *Econometrica* 55(5), 1217–1222.
- Christiano, L. J. and W. J. den Haan (1996). Small-Sample Properties of GMM for Business-Cycle Analysis. *Journal of Business & Economic Statistics* 14(3), 309–327.
- Cribari-Neto, F. (2004). Asymptotic Inference Under Heteroskedasticity of Unknown Form. *Computational Statistics & Data Analysis* 45(2), 215–233.
- Cribari-Neto, F. and W. B. da Silva (2011). A New Heteroskedasticity-Consistent Covariance Matrix Estimator for the Linear Regression model. *AStA Advances in Statistical Analysis* 95(2), 129–146.
- Cribari-Neto, F., S. L. P. Ferrari, and G. M. Cordeiro (2000). Improved Heteroscedasticity-Consistent Covariance Matrix Estimators. *Biometrika* 87(4), 907–918.
- Cribari-Neto, F., T. C. Souza, and K. L. P. Vasconcellos (2007). Inference Under Heteroskedasticity and Leveraged Data. *Communications in Statistics - Theory and Methods* 36(10), 1877–1888.
- Cribari-Neto, F. and S. G. Zarkos (2001). Heteroskedasticity-Consistent Covariance Matrix

- Estimation: White's Estimator and the Bootstrap. *Journal of Statistical Computation and Simulation* 68(4), 391–411.
- Davidson, R. and J. G. Mackinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- de Jong, R. M. (2000). A Strong Consistency Proof for Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimators. *Econometric Theory* 16(2), 262–268.
- de Jong, R. M. and J. Davidson (2000). Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices. *Econometrica* 68(2), 407–423.
- den Haan, W. J. and A. T. Levin (1997). A Practitioner's Guide to Robust Covariance Matrix Estimation. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 15 of *Robust Inference*, Chapter 12, pp. 299–342. Amsterdam: Elsevier.
- den Haan, W. J. and A. T. Levin (2000). Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order. Technical Working Paper Series 255, NBER.
- Diebold, F. X. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. *Journal of Business & Economic Statistics* 33(1), 1–9.
- Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Eicker, F. (1963). Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions. *The Annals of Mathematical Statistics* 34(2), 447–456.
- Elliot, G. and A. Timmermann (2004). Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions. *Journal of Econometrics* 122(1), 47–79.
- Godfrey, L. G. (2006). Tests for Regression Models with Heteroskedasticity of Unknown Form. *Computational Statistics & Data Analysis* 50(10), 2715–2733.

- Hamilton, J. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.
- Hansen, B. E. (1992). Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes. *Econometrica* 60(4), 967–972.
- Hansen, B. E. (2016). *Econometrics*. Unpublished manuscript.
- Hoaglin, D. C. and R. E. Welsch (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician* 32(1), 17–22.
- Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* 18(6), 1449–1459.
- Kauermann, G. and R. Carroll (2001). A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association* 96(456), 1387–1396.
- Keener, R. W., J. Kmenta, and N. C. Weber (1991). Estimation of the Covariance Matrix of the Least-Squares Regression Coefficients when the Disturbance Covariance Matrix is of Unknown form. *Econometric Theory* 7(1), 22–45.
- Kiefer, K. M., T. J. Vogelsang, and H. Bunzel (2000). Simple Robust Test of Regression Hypotheses. *Econometrica* 68(3), 695–714.
- Kuan, C.-M. and Y.-W. Hsieh (2008). Improved HAC Covariance Matrix Estimation based on Forecast Errors. *Economic Letters* 99(1), 89–92.
- Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* 17(3), 1217–1241.
- Long, J. S. and L. H. Ervin (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician* 54(3), 217–224.
- MacKinnon, J. G. and H. White (1985). Some Heteroskedasticity-Consistent Covariance Ma-

- trix Estimators with Improved Finite Sample Properties. *Journal of Econometrics* 29(3), 305–325.
- Meloun, M. and J. Militký (2001). Detection of Single Influential Points in OLS Regression Model Building. *Analytica Chimica Acta* 439(2), 169–191.
- Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55(3), 703–708.
- Newey, W. K. and K. D. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies* 61(4), 631–653.
- Sanderson, C. (2010). Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments. Technical report, NICTA.
- Simonoff, J. S. (1993). The Relative Importance of Bias and Variability in the Estimation of the Variance of a Statistic. *Journal of the Royal Statistical Society. Series D (The Statistician)* 42(1), 3–7.
- Sul, D., P. C. B. Phillips, and C.-Y. Choi (2005). Prewhitening Bias in HAC Estimation. *Oxford Bulletin of Economics and Statistics* 67(4), 517–546.
- West, K. D. (1997). Another Heteroskedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator. *Journal of Econometrics* 76(1–2), 171–191.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48(4), 817–838.
- White, H. (2000). *Asymptotic Theory for Econometricians* (Second Revised ed.). Academic Press.
- Wooldridge, J. A. (2013). *Introductory Econometrics: A Modern Approach* (Fifth ed.). Mason, OH: South-Western.

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software* 11(10), 1–17.

# Appendix

## Estimating the Leverage Points

For any matrix of predictors with full column rank  $N$ , it is possible to perform the  $QR$  decomposition. This involves finding a  $T \times N$  matrix  $Q$  and a  $N \times N$  upper triangular matrix  $R$  such that:

$$X = QR \quad \text{and} \quad Q'Q = I_T$$

The second condition implies that the columns of  $Q$  are orthonormal (i.e., they each have unit Euclidean norm). We can construct the projection matrix  $P$  given the matrix  $Q$  as:

$$P = QQ'$$

And then  $h_t = \text{diag}(P)$ . However, when  $T$  is large this method is not recommended and a vastly more efficient procedure for computing  $h_t$  is:

$$h_t = \sum_{i=1}^N Q_{ti}^2$$

## Consistency of the Alternative Residuals

Suppose the standard regularity conditions for Least Squares estimation with mixing time series processes holds (for example Theorem 3.78 from White 2000). Next, recall the two alternative residuals,  $\tilde{\epsilon}_t$ , and  $\dot{\epsilon}_{m,t}$ , as well as the two discount factors applied to the term  $(1 - h_t)$  which were defined in Equations (15) and (16). Let  $\check{\epsilon}_t \in \{\tilde{\epsilon}_t, \dot{\epsilon}_{m,t}\}$  and let  $\zeta \in \{1, \delta_t/2\}$ . I use Hansen (2016) Section 6.21 to show that  $\check{\epsilon}_t - \epsilon_t \xrightarrow{p} 0$  as  $T \rightarrow \infty$ . The

proof of this statement proceeds as follows:

$$\begin{aligned}
\check{\epsilon}_t - \epsilon_t &= \check{\epsilon}_t - \hat{\epsilon}_t + \hat{\epsilon}_t - \epsilon_t \\
|\check{\epsilon}_t - \epsilon_t| &\leq |\check{\epsilon}_t - \hat{\epsilon}_t| + |\hat{\epsilon}_t - \epsilon_t| \\
&\leq \left| \hat{\epsilon}_t (1 - h_t)^{-\zeta} - \hat{\epsilon}_t \right| + |\hat{\epsilon}_t - \epsilon_t| \\
&\leq |\hat{\epsilon}_t| \left| (1 - h_t)^{-\zeta} - 1 \right| + |\hat{\epsilon}_t - \epsilon_t| \\
&\leq |\hat{\epsilon}_t - \epsilon_t + \epsilon_t| \left| (1 - h_t)^{-\zeta} - 1 \right| + |\hat{\epsilon}_t - \epsilon_t| \\
&\leq (|\epsilon_t| + |\hat{\epsilon}_t - \epsilon_t|) \left| (1 - h_t)^{-\zeta} - 1 \right| + |\hat{\epsilon}_t - \epsilon_t| \\
&\leq \left( |\epsilon_t| + \left| x_t' (\hat{\beta}_T - \beta_0) \right| \right) \left| (1 - h_t)^{-\zeta} - 1 \right| + \left| x_t' (\hat{\beta}_T - \beta_0) \right| \\
&\leq \left( |\epsilon_t| + \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \right) \left| (1 - h_t)^{-\zeta} - 1 \right| + \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \\
&\leq \left( |\epsilon_t| + \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \right) \left| \frac{1 - (1 - h_t)^\zeta}{(1 - h_t)^\zeta} \right| + \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \\
\max_{1 \leq t \leq T} |\check{\epsilon}_t - \hat{\epsilon}_t| &\leq \max_{1 \leq t \leq T} \left( |\epsilon_t| + \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \right) \left| \frac{1 - (1 - h_t)^\zeta}{(1 - h_t)^\zeta} \right| + \max_{1 \leq t \leq T} \|x_t\| \left\| \hat{\beta}_T - \beta_0 \right\| \\
&= o_p(1) + o_p(T^{-1/2+1/r}) \\
&= o_p(T^{-1/2+1/r})
\end{aligned}$$

Where the third last line makes use of the assumption  $\mathbb{E}|\epsilon_t|^r < \infty$  for  $r \geq 1$  as well as Theorem 6.22.1 of Hansen (2016) which shows that the leverage points are asymptotic negligible, i.e.,  $\max_{1 \leq t \leq T} h_t = o_p(1)$ . In addition,  $\left\| \hat{\beta}_T - \beta_0 \right\| = O_p(T^{-1/2})$  from Theorem 6.3.2 in Hansen (2016). Finally, we can invoke Theorem 5.12.1 (Hansen 2016) to show that  $T^{-1/r} \max_{1 \leq t \leq T} \|x_t\| \xrightarrow{p} 0$ . Hence:

$$\check{\epsilon}_t = \epsilon_t + o_p(T^{-1/2+1/r})$$

As a result, the alternative errors are a uniformly consistent estimator for the true stochastic errors,  $\epsilon_t$ , and as such, exiting asymptotic results proving consistency of the sample kernel-based HAC covariance matrix estimator remain valid.

# Alternative Residual Estimator Comparison Tables

Table A1: Gaussian AR(1) process with  $\phi = 0.5$

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	$\max h_t$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.281	0.281	0.468	0.546	0.092	0.386	4	0.120
$\tilde{\epsilon}$	1.428	0.428	0.528	0.680	0.078	0.408	4	0.120
$\dot{\epsilon}_m$	1.450	0.450	0.540	0.703	0.076	0.411	4	0.120
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	3.341	2.341	1.915	3.024	0.106	0.613	4	0.120
$\tilde{\epsilon}$	3.817	2.817	2.255	3.608	0.086	0.654	4	0.120
$\dot{\epsilon}_m$	3.917	2.917	2.343	3.741	0.082	0.662	4	0.120
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.696	0.696	1.040	1.252	0.102	0.435	4	0.120
$\tilde{\epsilon}$	1.938	0.938	1.227	1.545	0.082	0.464	4	0.120
$\dot{\epsilon}_m$	1.989	0.989	1.275	1.614	0.078	0.470	4	0.120

*Note:* Numbers represent simulation results from 10,000 replications using an AR(1) process for the predictors and the error with  $\phi = 0.5$  and data  $\sim \mathcal{N}(0, 1)$ . In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{v}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘ $\max h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Table A2: Gaussian MA(1) process with  $\theta = 0.5$

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	$\max h_t$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.149	0.149	0.375	0.403	0.075	0.367	4	0.121
$\tilde{\epsilon}$	1.283	0.283	0.425	0.510	0.062	0.387	4	0.121
$\dot{\epsilon}_m$	1.303	0.303	0.435	0.530	0.060	0.390	4	0.121
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	3.075	2.075	1.624	2.634	0.091	0.590	4	0.121
$\tilde{\epsilon}$	3.525	2.525	1.933	3.180	0.071	0.631	4	0.121
$\dot{\epsilon}_m$	3.622	2.622	2.016	3.307	0.067	0.639	4	0.121
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.566	0.566	0.887	1.052	0.085	0.420	4	0.121
$\tilde{\epsilon}$	1.796	0.796	1.057	1.323	0.067	0.449	4	0.121
$\dot{\epsilon}_m$	1.846	0.846	1.103	1.390	0.063	0.455	4	0.121

*Note:* Numbers represent simulation results from 10,000 replications using an MA(1) process for the predictors and the error with  $\theta = 0.5$  and data  $\sim \mathcal{N}(0, 1)$ . In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{v}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘ $\max h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).



Table A3: Student-t AR(1) process with  $\phi = 0.5$ 

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	$\max h_t$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.253	0.253	0.638	0.686	0.098	0.378	4	0.193
$\tilde{\epsilon}$	1.430	0.430	0.760	0.873	0.079	0.403	4	0.193
$\dot{\epsilon}_m$	1.463	0.463	0.796	0.921	0.077	0.408	4	0.193
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	4.416	3.416	5.317	6.320	0.122	0.670	4	0.193
$\tilde{\epsilon}$	5.680	4.680	9.945	10.991	0.093	0.741	4	0.193
$\dot{\epsilon}_m$	6.082	5.082	12.333	13.338	0.087	0.758	4	0.193
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.946	0.946	2.441	2.617	0.113	0.450	4	0.193
$\tilde{\epsilon}$	2.427	1.427	3.816	4.074	0.083	0.494	4	0.193
$\dot{\epsilon}_m$	2.570	1.570	4.457	4.725	0.076	0.505	4	0.193

*Note:* Numbers represent simulation results from 10,000 replications using an AR(1) process for the predictors and the error with  $\phi = 0.5$  and data  $\sim t_{(5)}$  standardised to have unit variance. In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{v}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘ $\max h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Table A4: Student-t MA(1) process with  $\theta = 0.5$ 

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	$\max h_{ii}$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.122	0.122	0.572	0.585	0.083	0.358	4	0.200
$\tilde{\epsilon}$	1.290	0.290	0.695	0.753	0.064	0.383	4	0.200
$\dot{\epsilon}_m$	1.324	0.324	0.736	0.804	0.062	0.388	4	0.200
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	4.308	3.308	5.525	6.439	0.108	0.658	4	0.200
$\tilde{\epsilon}$	5.680	4.680	10.906	11.867	0.075	0.733	4	0.200
$\dot{\epsilon}_m$	6.138	5.138	13.863	14.784	0.069	0.753	4	0.200
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.845	0.845	2.348	2.495	0.098	0.437	4	0.200
$\tilde{\epsilon}$	2.351	1.351	3.878	4.106	0.069	0.484	4	0.200
$\dot{\epsilon}_m$	2.510	1.510	4.650	4.888	0.064	0.496	4	0.200

*Note:* Numbers represent simulation results from 10,000 replications using an MA(1) process for the predictors and the error with  $\theta = 0.5$  and data  $\sim t_{(5)}$  standardised to have unit variance. In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{v}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘ $\max h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Table A5: Chi-squared AR(1) process with  $\phi = 0.5$

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	max $h_t$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.236	0.236	0.670	0.710	0.100	0.374	4	0.220
$\tilde{\epsilon}$	1.430	0.430	0.820	0.926	0.080	0.401	4	0.220
$\dot{\epsilon}_m$	1.469	0.469	0.863	0.982	0.078	0.406	4	0.220
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	4.982	3.982	6.611	7.717	0.154	0.696	4	0.220
$\tilde{\epsilon}$	6.562	5.562	10.202	11.619	0.115	0.784	4	0.220
$\dot{\epsilon}_m$	7.050	6.050	11.576	13.061	0.109	0.807	4	0.220
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.996	0.996	2.393	2.592	0.126	0.449	4	0.220
$\tilde{\epsilon}$	2.584	1.584	3.634	3.964	0.093	0.503	4	0.220
$\dot{\epsilon}_m$	2.759	1.759	4.083	4.445	0.086	0.517	4	0.220

*Note:* Numbers represent simulation results from 10,000 replications using an AR(1) process for the predictors and the error with  $\phi = 0.5$  and data  $\sim \chi_{(2)}^2$  standardised to have mean zero and unit variance. In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{\nu}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘max  $h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Table A6: Chi-squared MA(1) process with  $\theta = 0.5$

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	max $h_t$
HOM: Residuals are uncorrelated with the predictors								
$\hat{\epsilon}$	1.107	0.107	0.609	0.618	0.087	0.354	4	0.229
$\tilde{\epsilon}$	1.295	0.295	0.767	0.822	0.068	0.381	4	0.229
$\dot{\epsilon}_m$	1.335	0.335	0.817	0.882	0.065	0.387	4	0.229
HET1: Residuals are correlated with $x_1$								
$\hat{\epsilon}$	4.997	3.997	7.176	8.214	0.144	0.691	4	0.229
$\tilde{\epsilon}$	6.754	5.754	11.480	12.841	0.108	0.786	4	0.229
$\dot{\epsilon}_m$	7.315	6.315	13.129	14.568	0.102	0.812	4	0.229
HET2: Residuals are correlated with all $x_i$ ( $i = 1, \dots, 4$ )								
$\hat{\epsilon}$	1.929	0.929	2.469	2.638	0.119	0.439	4	0.229
$\tilde{\epsilon}$	2.567	1.567	3.951	4.250	0.085	0.497	4	0.229
$\dot{\epsilon}_m$	2.765	1.765	4.500	4.833	0.077	0.513	4	0.229

*Note:* Numbers represent simulation results from 10,000 replications using an MA(1) process for the predictors and the error with  $\theta = 0.5$  and data  $\sim \chi_{(2)}^2$  standardised to have mean zero and unit variance. In each case the number of observations,  $T$ , was set to 128, while the number of variables,  $N$ , was set to 4. The ‘Estimand’ refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{\nu}_{22}$ , which has a true value of 1. ‘RMSE’ refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. ‘Width’ refers to the average estimated confidence interval width when using a nominal size of 0.05. ‘BW’ refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ‘max  $h_t$ ’ refers to the average of largest estimated leverage point from the generated matrix of predictors, values larger than  $2N/T = 0.0625$  are considered to be points of high leverage (see Hoaglin and Welsch 1978).

Table A7: Large Sample AR(1) process with  $\phi = 0.5$

Residual	Estimand	Bias	Std. Dev.	RMSE	Size	Width	BW	$\max h_t$
Gaussian distributed random variables with sample size = 1,000								
$\hat{\epsilon}$	4.272	3.272	1.016	3.426	0.066	0.255	6	0.021
$\tilde{\epsilon}$	4.347	3.347	1.039	3.505	0.063	0.257	6	0.021
$\dot{\epsilon}_m$	4.362	3.362	1.045	3.520	0.063	0.257	6	0.021
Gaussian distributed random variables with sample size = 5,000								
$\hat{\epsilon}$	4.486	3.486	0.511	3.523	0.062	0.117	9	0.005
$\tilde{\epsilon}$	4.502	3.502	0.514	3.539	0.061	0.117	9	0.005
$\dot{\epsilon}_m$	4.505	3.505	0.514	3.542	0.061	0.117	9	0.005
Gaussian distributed random variables with sample size = 10,000								
$\hat{\epsilon}$	4.544	3.544	0.373	3.563	0.055	0.083	11	0.003
$\tilde{\epsilon}$	4.552	3.552	0.374	3.571	0.055	0.084	11	0.003
$\dot{\epsilon}_m$	4.553	3.553	0.374	3.573	0.055	0.084	11	0.003
Student-t distributed random variables with sample size = 1,000								
$\hat{\epsilon}$	7.140	6.140	8.724	10.667	0.064	0.314	6	0.058
$\tilde{\epsilon}$	7.611	6.611	11.924	13.634	0.060	0.321	6	0.058
$\dot{\epsilon}_m$	7.744	6.744	13.092	14.727	0.059	0.323	6	0.058
Student-t distributed random variables with sample size = 5,000								
$\hat{\epsilon}$	8.151	7.151	10.322	12.557	0.057	0.153	9	0.022
$\tilde{\epsilon}$	8.322	7.322	12.892	14.826	0.055	0.153	9	0.022
$\dot{\epsilon}_m$	8.368	7.368	13.742	15.592	0.055	0.154	9	0.022
Student-t distributed random variables with sample size = 10,000								
$\hat{\epsilon}$	8.301	7.301	6.620	9.855	0.054	0.110	11	0.015
$\tilde{\epsilon}$	8.393	7.393	7.295	10.386	0.054	0.110	11	0.015
$\dot{\epsilon}_m$	8.416	7.416	7.489	10.539	0.053	0.111	11	0.015
Chi-squared distributed random variables with sample size = 1,000								
$\hat{\epsilon}$	8.195	7.195	6.672	9.813	0.076	0.340	6	0.053
$\tilde{\epsilon}$	8.561	7.561	7.246	10.473	0.071	0.347	6	0.053
$\dot{\epsilon}_m$	8.653	7.653	7.401	10.646	0.069	0.348	6	0.053
Chi-squared distributed random variables with sample size = 5,000								
$\hat{\epsilon}$	8.835	7.835	3.398	8.540	0.060	0.162	9	0.015
$\tilde{\epsilon}$	8.916	7.916	3.462	8.639	0.059	0.163	9	0.015
$\dot{\epsilon}_m$	8.935	7.935	3.478	8.664	0.059	0.163	9	0.015
Chi-squared distributed random variables with sample size = 10,000								
$\hat{\epsilon}$	9.015	8.015	2.546	8.409	0.056	0.117	11	0.009
$\tilde{\epsilon}$	9.056	8.056	2.572	8.457	0.055	0.117	11	0.009
$\dot{\epsilon}_m$	9.066	8.066	2.578	8.468	0.055	0.117	11	0.009

Note: Numbers represent simulation results from 10,000 replications with three different distributions: Gaussian,  $\mathcal{N}(0, 1)$ , Student-t,  $t_{(5)}$ , and Chi-squared,  $\chi^2_{(2)}$ . Each is used to generate an AR(1) process with  $\phi = 0.5$  and scaled such that the variance is unity. The error term is constructed to be correlated with the first non-constant predictor (HET1). In each simulation the number of variables,  $N$ , was set to 4. The 'Estimand' refers to the estimated value of the variance for the first non-constant coefficient,  $\hat{\nu}_{22}$ , which has a true value of 1. 'RMSE' refers to the root mean squared error of the estimand. The empirical size was computed by testing the null hypothesis,  $\mathbb{H}_0 : \beta_1 = 1$  against the alternative,  $\mathbb{H}_1 : \beta_1 \neq 1$  and calculating the average number of times the null was rejected at the nominal size of 0.05. 'Width' refers to the average estimated confidence interval width when using a nominal size of 0.05. 'BW' refers to the fixed interval bandwidth parameter computed as the integer part of  $4(T/100)^{2/9}$  and used with a BT kernel. ' $\max h_t$ ' refers to the average of largest estimated leverage point from the generated matrix of predictors.

# Alternative Residual Estimator Comparison Figures

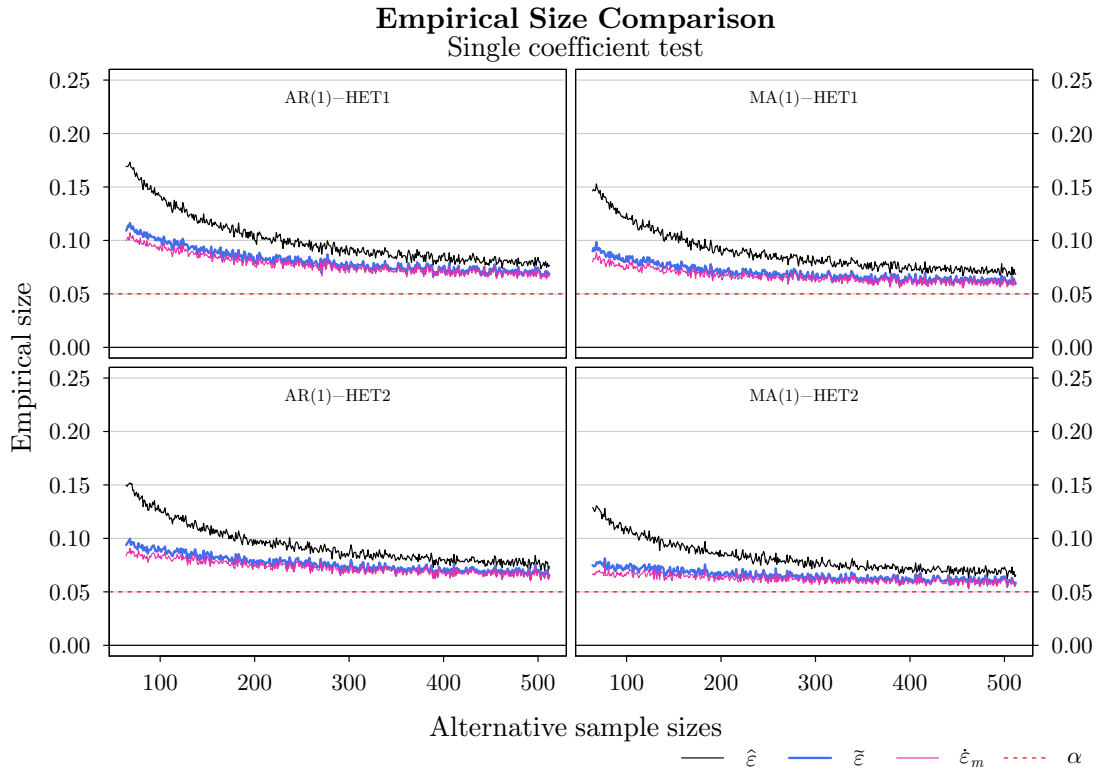


Figure A1: Student-t AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

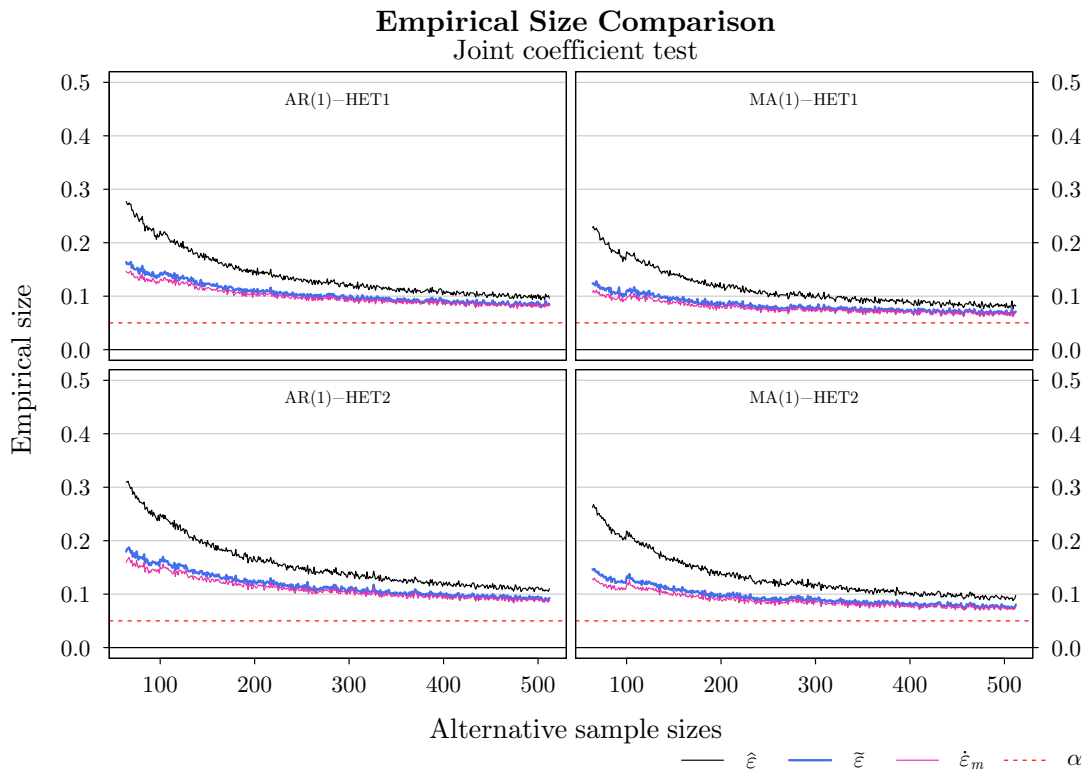


Figure A2: Student-t AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

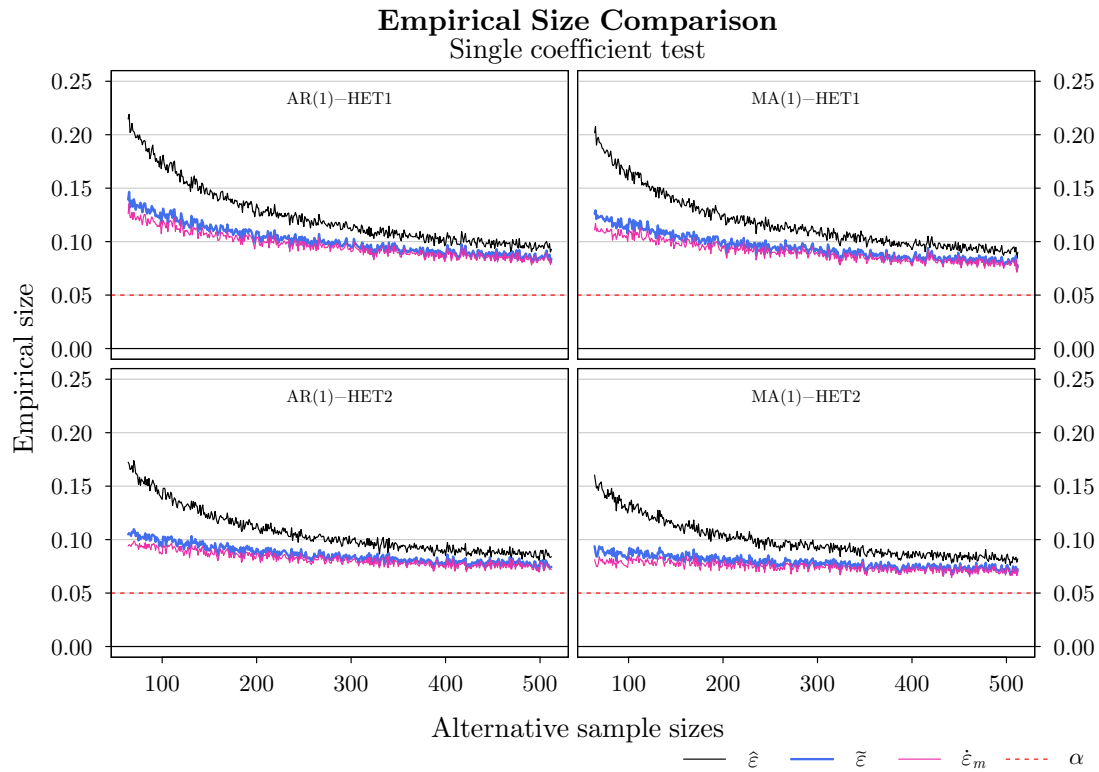


Figure A3: Chi-squared AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

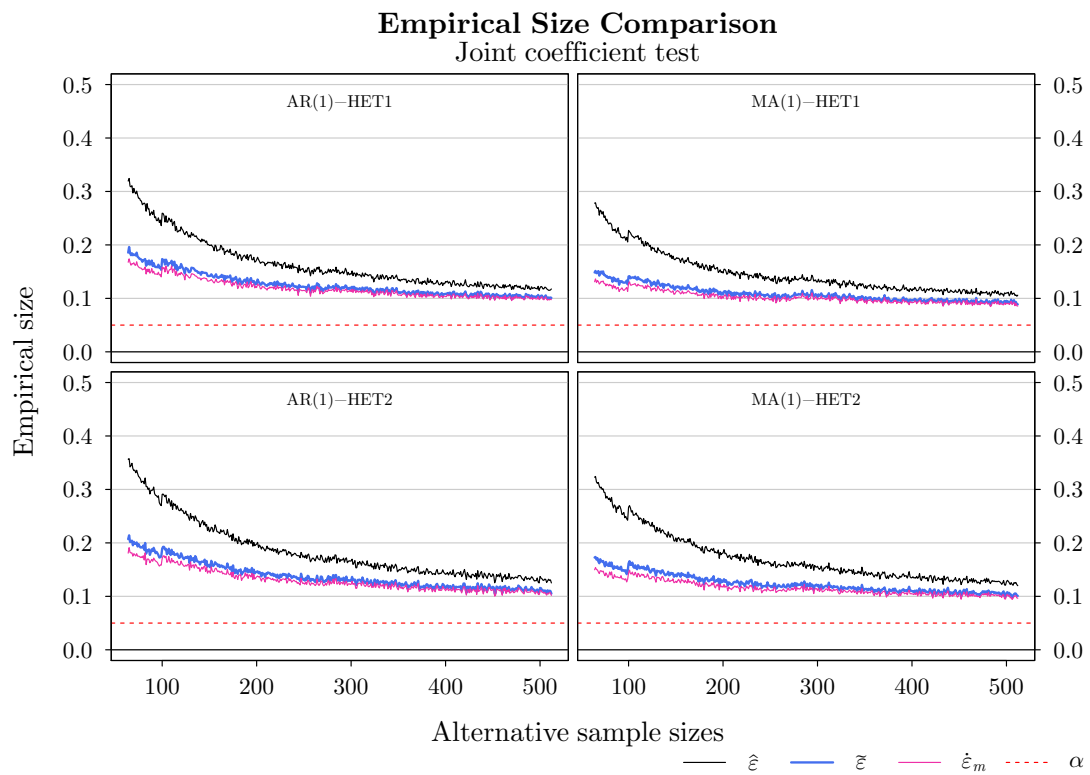


Figure A4: Chi-squared AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

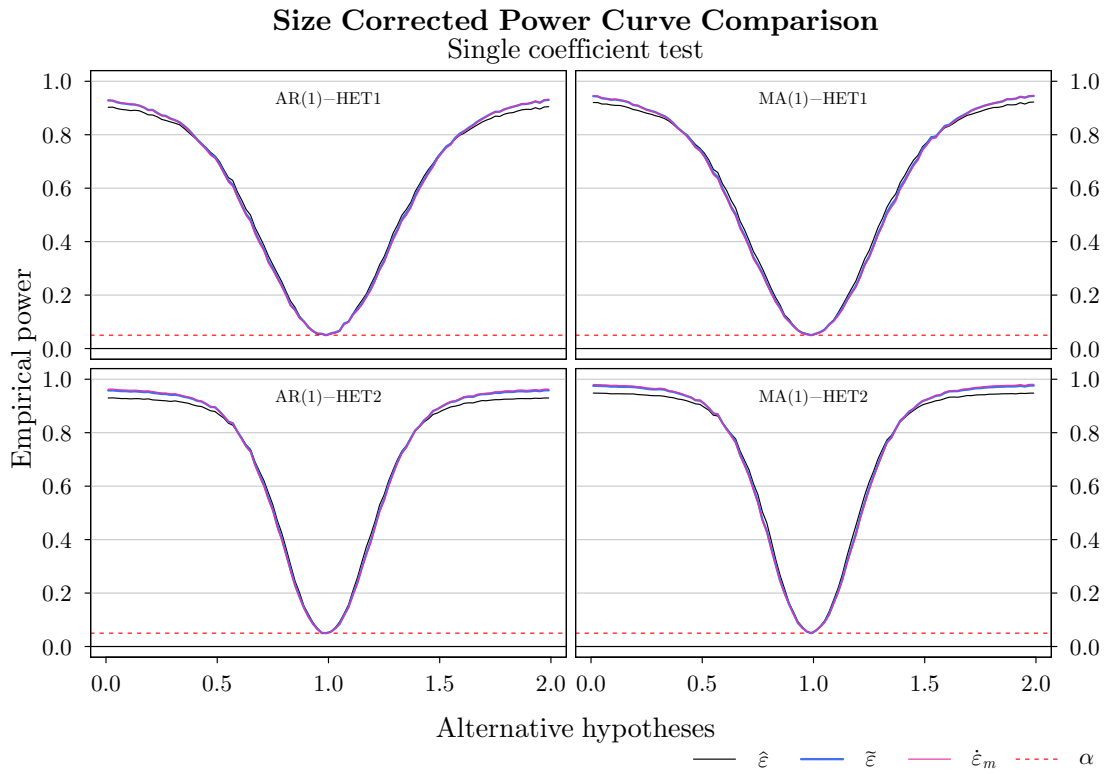


Figure A5: Student-t AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

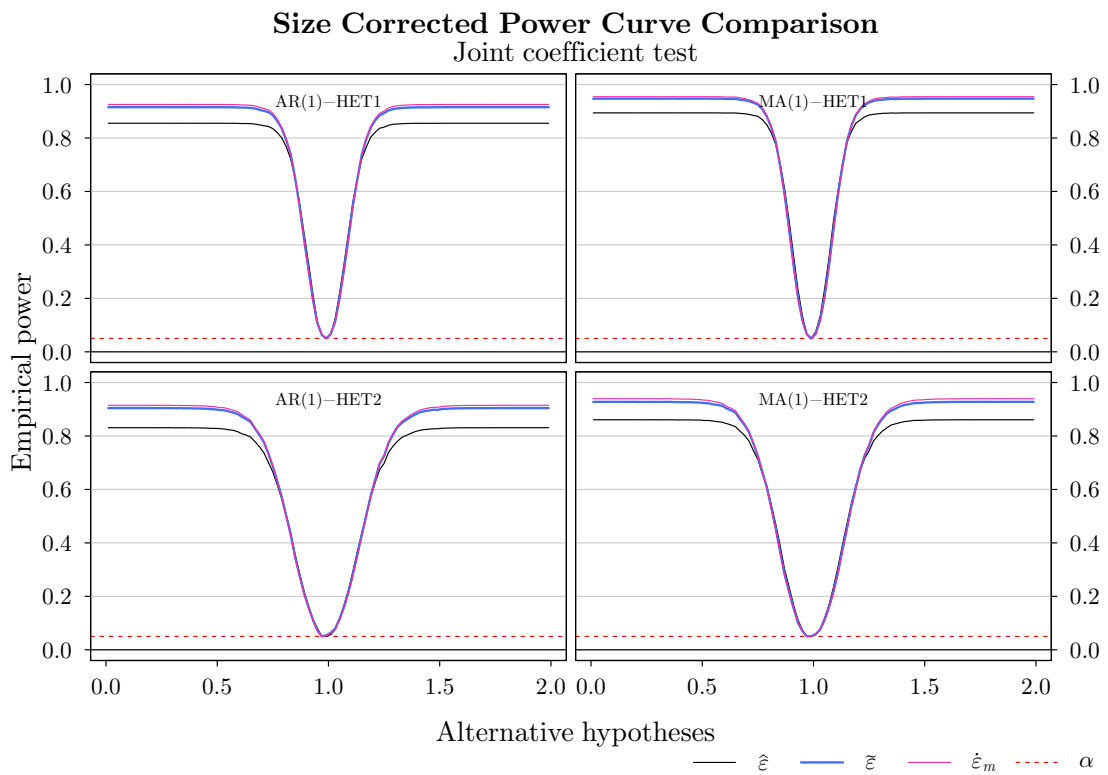


Figure A6: Student-t AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

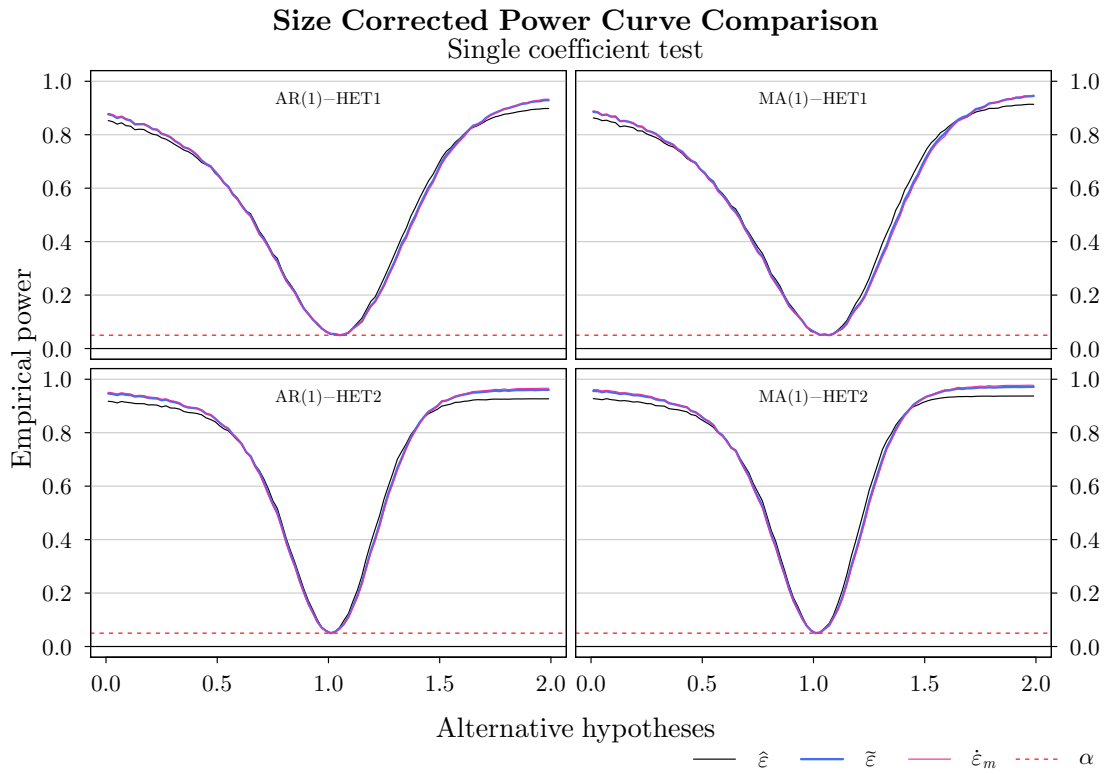


Figure A7: Chi-squared AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$

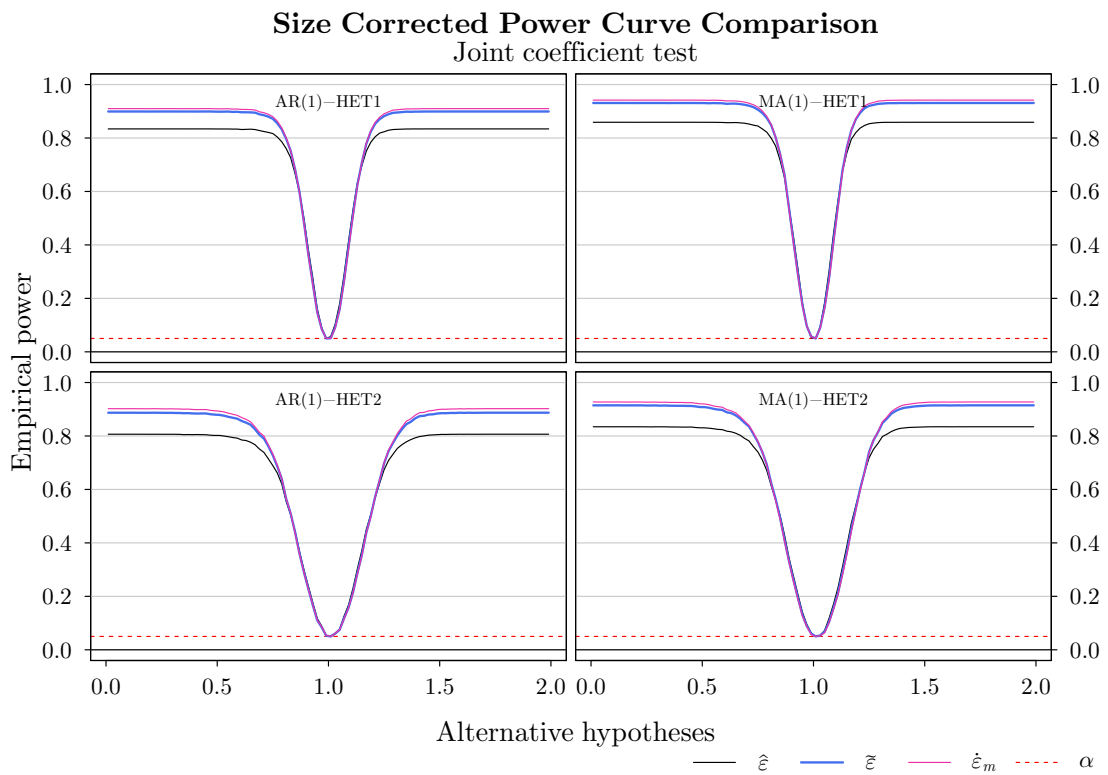


Figure A8: Chi-squared AR(1) & MA(1) Processes with  $\phi = 0.5$  &  $\theta = 0.5$