Improved HAC Covariance Matrix Estimation Based on Forecast Errors

Chung-Ming Kuan*
Institute of Economics
Academia Sinica

Yu-Wei Hsieh
Department of Economics
National Taiwan University

April 19, 2007

Abstract

We propose computing HAC covariance matrix estimators based on one-step-ahead forecasting errors. It is shown that this estimator is consistent and has smaller bias than other HAC estimators. Moreover, the tests that rely on this estimator have more accurate sizes without sacrificing its power.

Keywords: forecast error; HAC estimator; kernel estimator, recursive residual; robust test

JEL classification: C12; C22

*Author for correspondence: C.-M. Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan. Tel: +886 2 2782-2791 ext. 646; Email: ckuan@econ.sinica.edu.tw

1 Introduction

Consistent estimation of asymptotic covariance matrix is crucial in constructing tests of parameters. A leading class of consistent estimators for covariance matrix, also known as the heteroskedasticity and autocorrelation consistent estimator (henceforth HAC estimator), is the nonparameteric kernel estimator advocated by Newey and West (1987) in the econometrics literature. It has been found that the kernel HAC estimator is negatively biased, and hence the resulting test is typically over-sized in finite samples.

To improve on the performance of the kernel estimator, much research effort has been devoted to determine a better kernel function or its bandwidth (truncation lag); see, e.g., Andrews (1991), Newey and West (1994), and Phillips et al. (2005). It has also been found that the kernel estimator with pre-whitened estimands performs quite well (Andrews and Monahan, 1992). Yet, the potential benefit of using other residuals has not been investigated. This paper proposes computing HAC estimators based on one-step-ahead forecast errors (FEs), instead of the OLS residuals, so as to mitigate the estimation effect in residuals. We show that the FE-based HAC estimator is consistent but has smaller bias than other HAC estimators. As a result, the tests that depend on this estimator have more accurate finite-sample sizes, but it does not suffer from power loss, in contrast with the robust tests of Kiefer et al. (2000, henceforth KVB).

This paper proceeds as follows. We study the FE-based HAC estimator in Section 2 and present the simulation results in Section 3. Section 4 concludes.

2 The HAC-FE estimator

Given the linear specification $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_o + \varepsilon_t$, $t = 1, ..., T$, and the least square estimator $\hat{\boldsymbol{\beta}}_T$, the asymptotic covariance matrix of $T^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o)$ is $\boldsymbol{M}_{xx}^{-1}\boldsymbol{V}_o\boldsymbol{M}_{xx}^{-1}$, where $\boldsymbol{M}_{xx}$ is the probability limit of $T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$ and $\boldsymbol{V}_o = \lim_{T\to\infty}\mathrm{var}\big(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{x}_t\varepsilon_t\big)$. When $\boldsymbol{v}_t = \boldsymbol{x}_t\varepsilon_t$ are weakly dependent and heterogeneously distributed, a leading consistent estimator of $\boldsymbol{V}_o$ is the kernel HAC estimator:

$$\widehat{\boldsymbol{V}}_T = \sum_{j=-T}^{T}\kappa\left(\frac{j}{S_T}\right)\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t\hat{e}_t\hat{e}_{t+j}\boldsymbol{x}_{t+j}'\right],$$

where κ is a kernel function, S_T is the bandwidth, and $\hat{e}_t$ is the OLS residual. Typically, as this estimator is negatively biased, the resulting tests are over-sized in finite samples. To improve on the performance of $\widehat{\boldsymbol{V}}_T$, much research interest has been devoted to the choice of kernel function and its bandwidth.

Other than the kernel function and its bandwidth, note that conceivable that the estimation effect in $\hat{e}_t$ is also responsible for the bias of $\widehat{\boldsymbol{V}}_T$. In fact, while $\boldsymbol{x}_t\hat{e}_t$ must sum to zero, a consequence of OLS estimation, $\boldsymbol{x}_t\varepsilon_t$ are not subject to this constraint. This motivates us to consider other residuals that can circumvent such restriction. A candidate is the recursive residual proposed by Brown, Durbin, and Evans (1975):

$$\hat{u}_t = \big(y_t - \boldsymbol{x}_t'\tilde{\boldsymbol{\beta}}_{t-1}\big)/\sqrt{1 + \boldsymbol{x}_t'\big(\textstyle\sum_{i=1}^{t-1}\boldsymbol{x}_i\boldsymbol{x}_i'\big)^{-1}\boldsymbol{x}_t},$$

where $\tilde{\boldsymbol{\beta}}_{t-1}$ is the recursive OLS estimator based on the subsample of first $t-1$ obser-vatons. Note that $\sum_{t=1}^{T}\boldsymbol{x}_t\hat{u}_t \neq \boldsymbol{0}$. Moreover, $\hat{u}_t$ can be interpreted as the martingale transformation of Khmaladze (1981) that eliminates the estimation effect (Bai, 2003). However, this interpretation is valid only under certain conditions (e.g., i.i.d. normality of error term). Also, the complex normalizing factor in the recursive residuals may have an adverse effect on the resulting HAC estimator, as will be seen in our simulations. We thus propose using the "non-standardized" version of recursive residuals, namely, the one-step-ahead FEs: $\tilde{e}_t = y_t - \boldsymbol{x}_t'\tilde{\boldsymbol{\beta}}_{t-1}$. Clearly, $\sum_{t=1}^{T}\boldsymbol{x}_t\tilde{e}_t \neq \boldsymbol{0}$. In what follows, the HAC estimator based on $\tilde{e}_t$, denoted as $\widetilde{\boldsymbol{V}}_T$, will be referred to as the HAC-FE estimator, and the HAC estimators based on the OLS and recursive residuals will be referred to as the HAC-OLS and HAC-BDE estimators, respectively.

We now briefly discuss the consistency of $\widetilde{\boldsymbol{V}}_T$. Similar to the OLS-residual-based $\widehat{\boldsymbol{V}}_T$, the kernel function κ for $\widetilde{\boldsymbol{V}}_T$ is assumed to be continuous such that $\kappa(0) = 1$, $|\kappa(x)| \leq 1$ for all x in $\mathbf{R}$, $\kappa(x) = \kappa(-x)$, and $\int_{\mathbf{R}}|\kappa(x)|dx < \infty$. Moreover, the bandwidth S_T is required to diverge with T such that for some $d \in (2,4]$, $S_T = o(T^{1/2-1/d})$. These requirements are standard in the literature of HAC estimation; see e.g., de Jong (2000). Let $\xrightarrow{\mathbb{P}}$ denote convergence in probability and $[c]$ denote the integer part of the number c. Given some regularity conditions on $\boldsymbol{v}_t$ and $\boldsymbol{x}_t$ (e.g., Hansen, 1992), suppose further that the recursive OLS estimators are consistent uniformly in the following sense: $\hat{\boldsymbol{\beta}}_{[Tr]} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}_o$ uniformly in $r \in (0,1]$. Then, it is not difficult to show that $\widetilde{\boldsymbol{V}}_T \xrightarrow{\mathbb{P}} \boldsymbol{V}_o$. The proof is similar to that of Hansen (1992); we omit the details.

3 Monte Carlo Simulation

We redo the simulations in Andrews (1991) and KVB. We focus on the Wald tests based on various HAC estimators and the robust test based on KVB's normalizing matrix. The HAC estimators we consider are the HAC-FE, HAC-BDE, HAC-OLS, and HAC-PW estimators with the quadratic spectral kernel and the bandwidth selection method of Andrews (1991), where HAC-PW is the pre-whitened HAC estimator of Andrews and

Monahan (1992). For the HAC-FE and HAC-BDE estimators, the first 10% of FEs and recursive residuals are dropped. In all simulations, the sample is $T = 128$, the number of replications is 2000, and the nominal size is 5%.

For size simulations, the data generating process (DGP) is $y_t = \varepsilon_t$. The DGP is denoted as AR1-HOMO when ε_t is autocorrelated and homoskedastic: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ with v_t i.i.d. $\mathcal{N}(0, 1 - \rho^2)$ and $\rho = 0, \pm 0.3, \pm 0.5, \pm 0.7$, and it is denoted as AR1-HET1 when ε_t is generated as in AR1-HOMO but is multiplied by $|x_{1,t}|$. We estimte a linear regression model with a constant and four stochastic regressors; the i-th regressor is generated as: $x_{i,t} = \rho x_{i,t-1} + u_{i,t}$, where for a given i, $u_{i,t}$ are i.i.d. $\mathcal{N}(0, 1 - \rho^2)$ with the same ρ as in the generation of ε_t, and $\{u_{i,t}\}$ and $\{u_{j,t}\}$ are independent for $i \neq j$. We test two hypotheses: (i) $\beta_1 = 0$ and (ii) $\beta_i = 0$, $i = 1, \ldots, 3$. The results of size simulations are reported in Table 1.

We find that the HAC-FE estimator performs quite well. For testing a single hypothesis, the test with the HAC-FE estimator is properly sized when ρ is not too large (for AR1-HOMO, $|\rho| \leq 0.5$; for AR1-HET1, $|\rho| \leq 0.3$). Its empirical sizes are comparable with KVB's robust test under AR1-HOMO, and are more accurate than KVB's test under AR1-HET1, except when $|\rho|$ is large. By contrast, the tests with other HAC estimators have larger size distortions. In particular, HAC-OLS yields the worst test sizes under AR1-HOMO, and both HAC-OLS and HAC-BDE perform poorly under AR1-HET1.

It can also be seen that the size performances of these tests all deteriorate when the number of hypotheses increases. For the case of 3 hypotheses, the test with the HAC-FE estimator is properly sized only under AR1-HOMO with $|\rho| \leq 0.3$ and is in general oversized under AR1-HET1. Nonetheless, its empirical sizes are comparable with or better than those of KVB's test under both DGPs with $|\rho| \leq 0.3$. Again, the test with the HAC-FE estimator outperforms the tests with other HAC estimators, where those with the HAC-BDE and HAC-OLS estimators perform similarly poor and are outperformed by the test with the HAC-PW estimator. It is quite remarkable that the HAC-FE estimator yields better size performance than does the HAC-PW estimator in most cases we considered. This shows that the pre-whitening regression which requires a user-chosen parameter (lag order) may be avoided in practice. In other simulations (not reported here), we find that the HAC-PW estimator may result in even larger size distortion than does the HAC-OLS estimator when nonlinear dependence (e.g., bilinearity, SETAR, and Markov switching) is present, but the HAC-FE estimator still performs better.

In Figures 1 (Figure 2), we plot the bias, variance, and mean squared error (MSE) of the HAC-OLS, HAC-FE, and HAC-BDE estimators with different truncation lags

($S_T = 1, \ldots, 30$), under AR1-HOMO (AR1-HET1), where the results of $\rho = 0.3$ and $\rho = 0.5$ are in the top and bottom rows, respectively. We can see that, for all truncation lags, the HAC-FE estimator has the smallest bias (may even have a positive bias) in all cases, whereas HAC-OLS has the largest bias under AR1-HOMO. Bias performances of HAC-OLS and HAC-BDE are similar under AR1-HET1. In terms of MSE, the HAC-OLS estimator performs the best, yet the HAC-FE estimator has the largest MSE. These results, together with Table 1, suggest that bias reduction, rather than MSE minimization, of the HAC estimator is more important for accurate test size, as argued by Simonoff (1993).

For the power simulations, the DGP is $y_t = \beta_o x_t + \varepsilon_t$, where x_t is ARMA(1,1): $x_t = 0.5x_{t-1} + u_t + 0.2u_{t-1}$ with u_t i.i.d. $\mathcal{N}(0,1)$, ε_t is AR(1)-GARCH(1,1): $\varepsilon_t = 0.3\varepsilon_{t-1} + \eta_t$ with $\eta_t = \sqrt{h_t} u_t$ and $h_t = 1 + 0.1\eta_{t-1}^2 + 0.8h_{t-1}$, and $\beta_o = 0.2, 0.4, \ldots, 2$. We estimate the model with a constant and one regressor x_t and test $H_0 : \beta_o = 0$. The size-corrected power curves of the tests with various HAC estimators and KVB's normalizing matrix are plotted in Figure 3, with β_o on the horizontal axis. While the power curves of the tests with HAC estimators are quite close to each other (the test with HAC-FE has only negligible power loss), they all dominate that of KVB's robust test. Hence, the HAC-FE estimator yields more accurate finite-sample test size without losing testing power, as opposed to the robust test of KVB.

4 Conclusion

In this paper we show that the HAC estimator based on one-step-ahead FEs remains consistent and compares favorably with other HAC estimators, in that it has a smaller bias and delivers better test sizes. Compared with KVB's robust test, the test with the HAC-FE estimator has similar size performance without sacrificing its power. Our result also suggests that bias reduction, rather than MSE minimization, is more important for improving on the performance of the HAC estimator.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–854.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimatior, *Econometrica*, **60**, 953–966.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models, *Review of Economics and Statistics*, **85**, 531–549.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society* B, **37**, 149–192.

de Jong, R. M. (2000). A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators, *Econometric Theory*, **16**, 262–268.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests, *Theory of Probability and Its Applications*, **26**, 240–265.

Kiefer, N. M., , T. J. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses, *Econometrica*, **68**, 695–714.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance estimation, *Review of Economic Studies*, **61**, 631–654.

Phillips, P. C. B., Y. Sun, and S. Jin (2005). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation, *International Economic Review*, forthcoming.

Simonoff, J. (1993). The relative importance of bias and variability in the estimation of the variance of a statistic, *The Statistician*, **42**, 3–7.

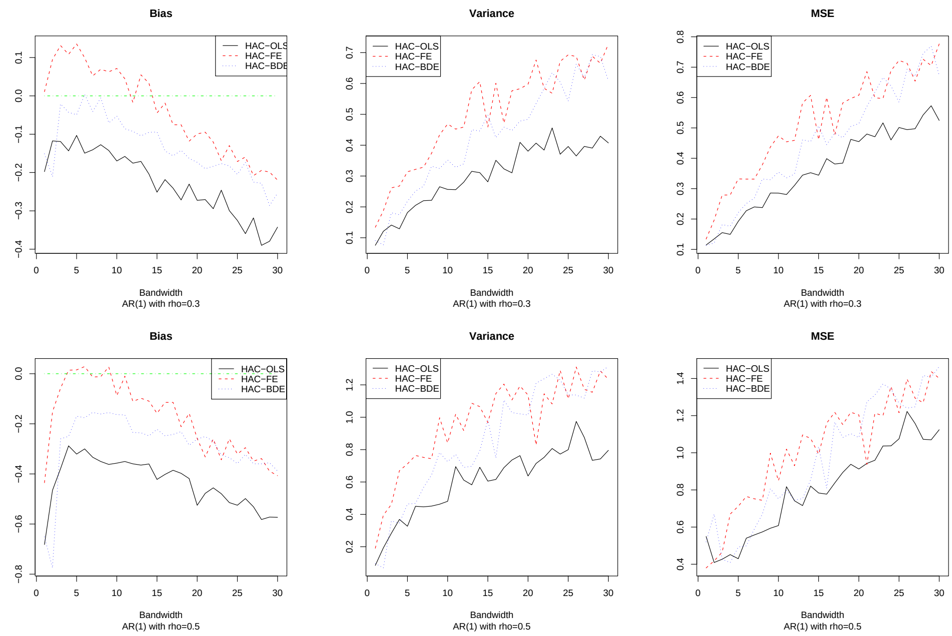

Figure 1: Bias, variance and MSE of HAC-OLS, HAC-FE, and HAC-BDE: AR1-HOMO.

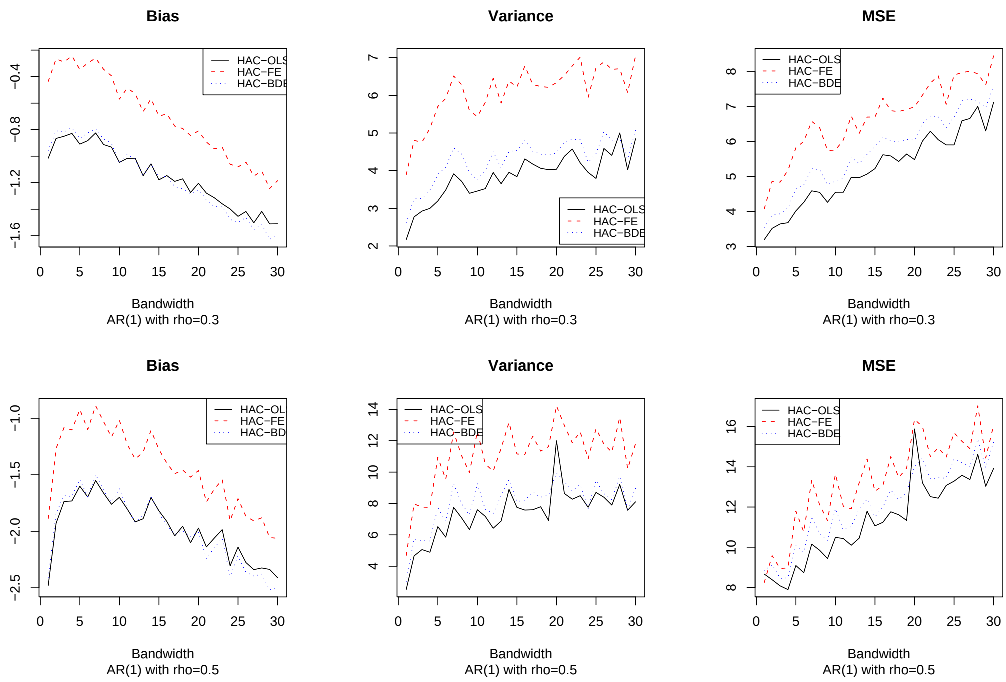

Figure 2: Bias, variance and MSE of HAC-OLS, HAC-FE, and HAC-BDE: AR1-HET1.

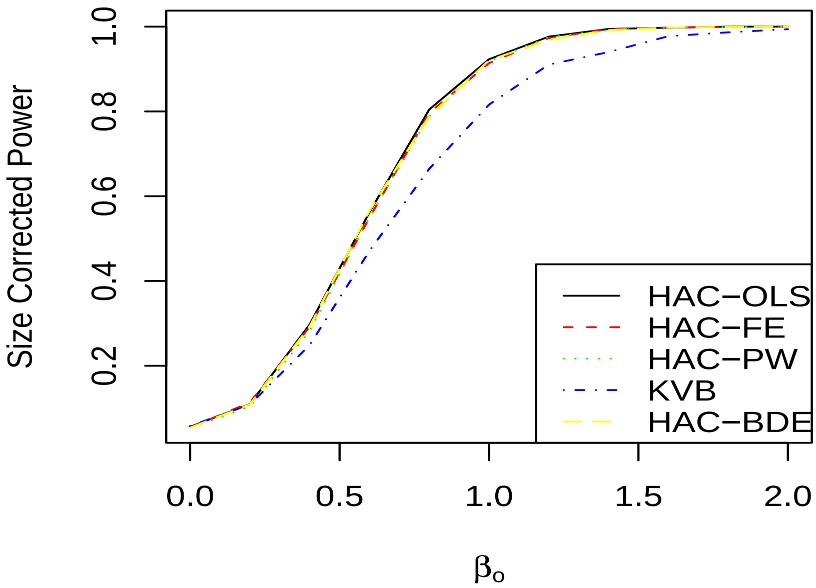

Figure 3: Empirical powers of the tests with HAC estimators and KVB's normalizing matrix.

Table 1: Empirical Size of AR1 Model

Model	ρ	HAC-OLS	HAC-FE	HAC-BDE	HAC-PW	KVB
$H_0 : \beta_1 = 0$						
AR1-HOMO	-0.7	0.116	0.0875	0.1065	0.096	0.088
	-0.5	0.0815	0.064	0.081	0.069	0.0595
	-0.3	0.066	0.0445	0.07	0.063	0.063
	0.0	0.066	0.046	0.0515	0.0675	0.0575
	0.3	0.0685	0.046	0.0595	0.0645	0.064
	0.5	0.0905	0.0665	0.0775	0.076	0.075
	0.7	0.1345	0.103	0.11	0.099	0.1025
AR1-HET1	-0.7	0.1415	0.115	0.139	0.115	0.0925
	-0.5	0.101	0.081	0.094	0.085	0.074
	-0.3	0.0765	0.0605	0.0815	0.0735	0.0655
	0.0	0.0595	0.0455	0.066	0.06	0.053
	0.3	0.0745	0.052	0.083	0.067	0.068
	0.5	0.095	0.0745	0.099	0.081	0.077
	0.7	0.14	0.1185	0.146	0.104	0.0855
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$						
AR1-HOMO	-0.7	0.2045	0.167	0.2045	0.165	0.126
	-0.5	0.146	0.108	0.123	0.1215	0.09
	-0.3	0.0965	0.059	0.0975	0.0965	0.068
	0.0	0.0845	0.0555	0.066	0.0975	0.067
	0.3	0.1035	0.065	0.1085	0.097	0.0615
	0.5	0.1545	0.119	0.145	0.118	0.088
	0.7	0.281	0.2105	0.2105	0.194	0.14
AR1-HET1	-0.7	0.203	0.178	0.2225	0.172	0.116
	-0.5	0.1495	0.1145	0.1405	0.1335	0.0845
	-0.3	0.104	0.074	0.1065	0.109	0.08
	0.0	0.0845	0.0645	0.077	0.0945	0.072
	0.3	0.1045	0.065	0.1005	0.1085	0.073
	0.5	0.1605	0.124	0.139	0.137	0.1055
	0.7	0.2335	0.1875	0.2285	0.178	0.126