A NOTE ON
ROBUST HYPOTHESIS TESTING

CHUNG-MING KUAN

Institute of Economics
Academia Sinica

This version: November 27, 2004

Address for correspondence: Institute of Economics, Academia Sinica, Taipei 115, Taiwan.

E-mail: ckuan@econ.sinica.edu.tw; URL: www.sinica.edu.tw/as/ssrc/ckuan

1 Introduction

In this note we discuss a new approach to constructing an asymptotically valid test without estimating the asymptotic covariance matrix. In the context of linear regression, well known large sample tests, such as the Wald and LM tests, usually require estimating the asymptotic covariance matrix of the normalized OLS estimator. This estimation may be cumbersome when data have complex dynamic properties. Newey and West (1987) and Gallant (1987) suggested nonparametric kernel estimators that are consistent even when there are serial correlations and conditional heteroskedasticity of unknown forms. Such estimators were further elaborated by, among others, Andrews (1991), Andrews and Monahan (1992), Hansen (1992), and Newey and West (1994). These estimators are usually referred to as the Newey-West estimator in the econometrics literature. A drawback of the Newey-West estimator is that it is somewhat arbitrary, in that its performance depends on the choice of the kernel function and truncation lag. These choices affect the resulting test statistics and render testing results fragile.

Instead of estimating the asymptotic covariance matrix, Kiefer, Vogelsang, and Bunzel (2000) proposed using a normalizing matrix that, though *not* a consistent estimator for the asymptotic covariance matrix, is capable of eliminating the nuisance parameters (i.e., the matrix square root of the asymptotic covariance matrix) in the limit. This approach, hereafter the KVB approach, circumvents the problems arising from nonparametric kernel estimation and yields robust tests. Bunzel, Kiefer, and Vogelsang (2001) and Vogelsang (2002) applied this approach to tests based on the nonlinear least squares estimator and the generalized method of moments estimator; Lobato (2001) also derived a portmanteau test for serial correlations along the same line. Kuan and Lee (2004) further extend the KVB approach to the context of general m tests.

2 Hypothesis Testing in Linear Regression

For the linear specification $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}+e_t$, it is typical to require the data obeying a suitable weak law of large number (WLLN) and a functional central limit theorem (FCLT). In what follows, we let $[c]$ denote the integer part of c, $\xrightarrow{\mathbb{P}}$ convergence in probability, $\Rightarrow$ weak convergence (of associated probability measures), $\xrightarrow{D}$ convergence in distribution, $\stackrel{d}{=}$ equality in distribution, $\boldsymbol{W}_k$ a vector of k independent, standard Wiener processes, and $\boldsymbol{B}_k$ the Brownian bridge obtained from $\boldsymbol{W}_k$ such that $\boldsymbol{B}_k(r) = \boldsymbol{W}_k(r) - r\boldsymbol{W}_k(1)$, $0 \leq r \leq 1$. When $k = 1$, we simply write W and B.

[A1] $[Tr]^{-1}\sum_{t=1}^{[Tr]}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{M}_o$ uniformly in $r \in [0,1]$ such that $\boldsymbol{M}_o$ is nonsingular.

[A2] $T^{-1/2}\sum_{t=1}^{[Tr]}[\boldsymbol{x}_t e_t - \mathbb{E}(\boldsymbol{x}_t e_t)] \Rightarrow \boldsymbol{S}\boldsymbol{W}_k(r)$ for $r \in [0,1]$, where $\boldsymbol{S}$ is the nonsingular, matrix square root of $\lim_{T\to\infty}\operatorname{var}(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{x}_t e_t)$.

[A3] There exists $\boldsymbol{\beta}_o$ such that $\epsilon_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_o$ and $\mathbb{E}(\boldsymbol{x}_t\epsilon_t) = \boldsymbol{0}$.

Combining [A2] and [A3] we immediately have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\boldsymbol{x}_t\epsilon_t \Rightarrow \boldsymbol{S}\boldsymbol{W}_k(r), \quad r \in [0,1];$$

in particular, $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{x}_t\epsilon_t \xrightarrow{D} \boldsymbol{S}\boldsymbol{W}_k(1) \overset{d}{=} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{\Sigma}_o)$. Let $\boldsymbol{M}_T = T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$ and $\hat{\boldsymbol{\beta}}_T$ denote the OLS estimator of $\boldsymbol{\beta}$. We have

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o\big) = \boldsymbol{M}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{x}_t\epsilon_t \xrightarrow{D} \boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{W}_k(1), \tag{1}$$

which has the distribution $\mathcal{N}(\boldsymbol{0},\, \boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{S}'\boldsymbol{M}_o^{-1})$.

We are interested in testing the null hypothesis $H_0\colon \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ with $\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\beta}_o$ and $\boldsymbol{R}$ a $q \times k$ matrix with full row rank. The Wald test of this hypothesis is

$$\mathcal{W}_T = T\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big)'\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big), \tag{2}$$

where $\widehat{\boldsymbol{\Sigma}}_T$ is a consistent estimator of $\boldsymbol{\Sigma}_o$. It is well known that $\mathcal{W}_T$ converges in distribution to $\chi^2(q)$ under the null hypothesis but need not be so when $\widehat{\boldsymbol{\Sigma}}_T$ is not a consistent estimator.

The exact form of $\mathcal{W}_T$ depends on how the estimator $\widehat{\boldsymbol{\Sigma}}_T$ is computed. Suppose that $\boldsymbol{x}_t\epsilon_t$ are serially uncorrelated. Let $\mathcal{F}^t$ denote the σ-algebra generated by $\{(\boldsymbol{x}_i, \epsilon_i), i \le t\}$. Then when ϵ_t are conditionally homoskedastic, i.e., $\mathbb{E}(\epsilon_t^2 \mid \mathcal{F}^{t-1}) = \sigma_o^2$, we have $\boldsymbol{\Sigma}_o = \sigma_o^2\boldsymbol{M}_o$ which can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_T = \hat{\sigma}_T^2\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'\right),$$

with $\hat{\sigma}_T^2$ the OLS variance estimator. When ϵ_t are conditionally heteroskedastic, $\boldsymbol{\Sigma}_o = \lim_{T\to\infty}\sum_{t=1}^{T}\mathbb{E}(\epsilon_t^2\boldsymbol{x}_t\boldsymbol{x}_t')$. In this case, we can estimate $\boldsymbol{\Sigma}_o$ using White's heteroskedasticity-consistent estimator:

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t^2\boldsymbol{x}_t\boldsymbol{x}_t',$$

where $\hat{e}_t$ are the OLS residuals. Consistent estimation of $\boldsymbol{\Sigma}_o$ is more cumbersome when $\boldsymbol{x}_t\epsilon_t$ are both serially correlated and ϵ_t are conditionally heteroskedastic. The estimator of Newey and West (1987) is

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_T = {} & \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t^2\boldsymbol{x}_t\boldsymbol{x}_t' + \\ & \frac{1}{T}\sum_{\tau=1}^{T-1} w_{m(T)}(\tau)\sum_{t=\tau+1}^{T}\big(\boldsymbol{x}_{t-\tau}\hat{e}_{t-\tau}\hat{e}_t\boldsymbol{x}_t' + \boldsymbol{x}_t\hat{e}_t\hat{e}_{t-\tau}\boldsymbol{x}_{t-\tau}'\big),\end{aligned} \tag{3}$$

where $w_{m(T)}$ is the so-called Bartlett kernel function:

$$w_{m(T)}(\tau) = \begin{cases} 1 - \frac{\tau}{m(T)}, & \text{if } 0 \le \frac{\tau}{m(T)} \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

and $m(T)$, also know as the "truncation lag," is a number growing with T. In practice, $m(T)$ must be specified *a priori*, and it in effect determines the number of autocovariances that need to be computed. Other kernel functions, such as the Parzen kernel and quadratic spectral kernel, were also considered in the literature. It is well known that the performance of the Newey-West estimator $\widehat{\boldsymbol{\Sigma}}_T$ varies with the chosen kernel function and its truncation lag. These choices are somewhat arbitrary in applications, however.

Following the KVB approach, we define $\boldsymbol{\varphi}_t$ as the normalized partial sum of $\boldsymbol{x}_i\hat{e}_i$:

$$\boldsymbol{\varphi}_t = \frac{1}{\sqrt{T}}\sum_{i=1}^{t}\boldsymbol{x}_i\hat{e}_i = \frac{1}{\sqrt{T}}\sum_{i=1}^{t}\big[\boldsymbol{x}_i\epsilon_i - \boldsymbol{x}_i\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o)\big],$$

and $\widehat{\boldsymbol{C}}_T = T^{-1}\sum_{t=1}^{T}\boldsymbol{\varphi}_t\boldsymbol{\varphi}_t'$. It follows from [A1]–[A3] that

$$\begin{aligned}\boldsymbol{\varphi}_{[Tr]} &= \frac{1}{\sqrt{T}}\sum_{i=1}^{[Tr]}\boldsymbol{x}_i\epsilon_i - \frac{[Tr]}{T}\left(\frac{1}{[Tr]}\sum_{i=1}^{[Tr]}\boldsymbol{x}_i\boldsymbol{x}_i'\right)\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \\ &\Rightarrow \boldsymbol{S}\boldsymbol{W}_k(r) - r\boldsymbol{S}\boldsymbol{W}_k(1),\end{aligned}$$

which is just $\boldsymbol{S}\boldsymbol{B}_k(r)$, and hence

$$\widehat{\boldsymbol{C}}_T \Rightarrow \boldsymbol{S}\left(\int_0^1 \boldsymbol{B}_k(r)\boldsymbol{B}_k(r)'\,\mathrm{d}r\right)\boldsymbol{S}' =: \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{S}'.$$

In contrast with the Newey-West estimator (3) which has a nonstochastic limit, $\widehat{\boldsymbol{C}}_T$ has a random limit, depending on a functional of the Brownian bridge.

From (1) it is readily seen that

$$\sqrt{T}\boldsymbol{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{S}'\boldsymbol{M}_o^{-1}\boldsymbol{R}').$$

Letting $\boldsymbol{G}$ denote the matrix square root of $\boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{S}'\boldsymbol{M}_o^{-1}\boldsymbol{R}'$ we can write

$$\sqrt{T}\boldsymbol{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \overset{d}{=} \boldsymbol{G}\boldsymbol{W}_q(1). \tag{4}$$

Analogous to (2), consider the following statistic:

$$\mathcal{W}_T^* = T\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big)'\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big). \tag{5}$$

In view of (4), it is also easy to see that

$$\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}' \Rightarrow \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{S}'\boldsymbol{M}_o^{-1}\boldsymbol{R}'.$$

Noting that $\boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}\boldsymbol{W}_k(r) \overset{d}{=} \boldsymbol{G}\boldsymbol{W}_q(r)$, we immediately have

$$\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}' \Rightarrow \boldsymbol{G}\boldsymbol{P}_q\boldsymbol{G}'.$$

This result, together with (4), leads:

$$\mathcal{W}_T^* \Rightarrow \boldsymbol{W}_q(1)'\boldsymbol{G}'\big(\boldsymbol{G}\boldsymbol{P}_q\boldsymbol{G}'\big)^{-1}\boldsymbol{G}\boldsymbol{W}_q(1) = \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1). \tag{6}$$

Thus, $\mathcal{W}_T^*$ does not depend on the matrix of nuisance parameters, $\boldsymbol{G}$, in the limit and is asymptotically pivotal. Although its distribution is non-standard, it can be easily simulated. Lobato (2001) reported the critical values of this distribution for various q. Note that Kiefer et al. (2000) considered a slightly different statistic $F_T^* = \mathcal{W}_T^*/q$, analogous to the classical F test, with the null limit $\boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)/q$.

When the null hypothesis is $\beta_i = r$, a robust test analogous to the conventional t test is thus

$$t^* = \frac{\sqrt{T}\big(\hat{\beta}_{i,T} - r\big)}{\sqrt{\hat{\delta}_{ii}}} \xrightarrow{D} \frac{W(1)}{\big[\int_0^1 B(r)^2\,\mathrm{d}r\big]^{1/2}}, \tag{7}$$

where $\hat{\delta}_{ii}$ is the i th diagonal element of $\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}$. Some quantiles of this asymptotic distribution are taken from Kiefer et al. (2000) and summarized in the second row of Table 1. It can be seen that this distribution is also symmetric about zero but more disperse than the standard normal distribution.

Remarks:

1. Kiefer and Vogelsang (2002a) showed that $2\widehat{\boldsymbol{C}}_T$ is algebraically equivalent to the Newey-West estimator using the Bartlett kernel function without truncation, i.e., $\widehat{\boldsymbol{\Sigma}}_T$ computed as (3) with $m(T) = T$; see also Kiefer and Vogelsang (2002b). One

Table 1: The quantiles of the asymptotic distribution of the t^* and t^{**} tests.

prob.	1%	2.5%	5%	10%	90%	95%	97.5%	99%
t^*	−8.544	−6.811	−5.374	−3.890	3.890	5.374	6.811	8.544
t^{**}	−6.090	−4.771	−3.764	−2.740	2.740	3.764	4.771	6.090

may then compute the $\mathcal{W}_T^{**}$ test by replacing $\widehat{\boldsymbol{C}}_T$ in (5) with $2\widehat{\boldsymbol{C}}_T$ so that $\mathcal{W}_T^{**} = 0.5 \times \mathcal{W}_T^*$; the critical values for $\mathcal{W}_T^{**}$ are one half of those for $\mathcal{W}_T^*$. Similarly, we may compute the t^{**} test as $t^*/\sqrt{2}$; the critical values of t^{**} are summarized in the last row of Table 1).

2. Abadir and Paruolo (2002) showed that the distribution of the limit in (7) is the same as that analyzed in Abadir and Paruolo (1997) which also contains analytic formulae of its density function and moments.

References

Abadir, K. M. and P. Paruolo (1997). Two mixed normal densities from cointegration analysis, *Econometrica*, **65**, 671–680.

Abadir, K. M. and P. Paruolo (2002). Simple robust testing of regression hypothesis: A comment *Econometrica*, **70**, 2097–2099.

Bunzel, H., N. M. Kiefer and T. J. Vogelsang (2001). Simple robust testing of hypotheses in nonlinear models, *Journal of the American Statistical association*, **96**, 1088–1096.

Kiefer, N. M. and T. J. Vogelsang (2002a). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation, *Econometrica*, **70**, 2093–2095.

Kiefer, N. M. and T. J. Vogelsang (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size, *Econometric Theory*, **18**, 1350–1366.

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

Kuan, C.-M. and W.-M. Lee (2004). M tests without estimating asymptotic covariance matrix, Working paper, Institute of Economics, Academia Sinica.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated, *Journal of the American Statistical Association*, **96**, 1066–1076.

Vogelsang, T. J. (2002). Testing in GMM models without truncation, Working Paper, Department of Economics and Statistical Science, Cornell University.