THE PERFORMANCE OF HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST TESTS: A MONTE CARLO STUDY WITH AN APPLICATION TO THE THREE-FACTOR FAMA-FRENCH ASSET PRICING MODEL

Surajit Ray[a] and N. E. Savin[b*]

[a] Bear Stearns Asset Management, 60 East 42nd Street, Suite 2544, New York, NY 10165

[b]Department of Economics, Tippie College of Business, University of Iowa, 108 John Pappajohn Bus. Bldg., Iowa City, IA 52242-1000

March 23, 2007

SUMMARY

This paper illustrates the pitfalls of the conventional heteroskedasticity and autocorrelation robust (HAR) Wald test and the advantages of new HAR tests developed by Kiefer and Vogelsang (2005) and Phillips, Sun and Jin (2003a, 2003b). The illustrations use the Fama-French (1993) three-factor model. The null that the intercepts are zero is tested for five-year, ten-year and longer sub-periods. The conventional HAR test with asymptotic *P*-values rejects the null for most five-year and ten-year sub-periods. By contrast, the null is not rejected by the new HAR tests. This conflict is explained by showing that the inferences based on the conventional HAR test are misleading for the sample sizes used in this application.

JEL classification: C32; G12.

Keywords: three-factor model, long-run variance, heteroskedastic and autocorrelation robust tests, heteroskedastic and autocorrelation consistent estimators, Bartlett and Parzen kernels.

Please send correspondence to: N. E. Savin, Tippie College of Business, Department of Economics, 108 John Pappajohn Bus. Bldg., Iowa City, IA 52242-1000, USA. E-mail: gene-savin@uiowa.edu. Fax: (319) 335-1956. Tel: (319) 335-0855.

1. INTRODUCTION

This paper illustrates the pitfalls of the conventional heteroskedasticity and autocorrelation robust (hereafter HAR (Phillips (2005)) Wald test and the advantages of new HAR tests developed by Kiefer and Vogelsang (2005, hereafter KV) and Phillips, Sun and Jin (2003a, 2003b, hereafter PSJ). The illustrations use the Fama-French (1993) three-factor model. This model provides an empirically relevant framework since it is one of the most popular multifactor models that now dominate empirical research on expected returns (Cochrane, (2005)). Fama-French (1996) present a summary of empirical findings with the three-factor model. See also Cochrane (1999) for an excellent survey of empirical results.

The three-factor model is formulated as a multivariate linear regression model where the dependent variables are the CRSP monthly excess returns of ten portfolios. The stocks are assigned to portfolios based on market equity (stock price times the number of shares). The explanatory variables are the Fama-French factors, namely the market excess return factor, the size factor and the book-to-market-equity factor. The null hypothesis is that the intercepts are zero. The explanation by Fama-French (1996) for testing the zero-intercept null is that if the intercepts are close to zero, then regressions that use the three-factors to absorb common time series variation in returns do a good job explaining the cross-section of average stock returns. In this paper, the three-factor model is estimated by least squares for five-year, ten-year and longer sub-periods.

The null is first tested using the conventional HAR Wald test. This test employs a heteroskedastic and autocorrelation consistent (HAC) estimator of the covariance matrix. We use the well-known HAC estimator proposed by Newey and West (1987, 1994) for the conventional HAR tests. The conventional test rejects the null for most five-year and ten-year sub-periods of monthly data using asymptotic P-values. There are two reasons for thinking that these results are problematic. The first is that the absolute values of most of the estimated intercepts are small in economic terms. Thus, there is a conflict between the economic significance and the statistical significance when statistical significance is judged by the conventional HAR test. The second is that the null is not rejected for most of the five-year and ten-year sub-periods when it is tested with new HAR tests. The conflict between the results produced by the conventional and new tests suggests that the results of the conventional test should not be accepted without further examination.

One potential explanation for the conflicting results is that the conventional test has high power compared to the new tests, assuming that conventional test has the correct Type I error in finite-samples, that is, the correct level. Another potential explanation for the conflict is that the actual finite-sample level of the conventional test is much larger than the nominal level when asymptotic critical values are used, or equivalently, the finite-sample *P*-value is substantially larger than the asymptotic *P*-value. Our study supports the second explanation.

The advantage of the new HAR tests over the conventional tests is better control over the level. This has been shown in simulation experiments by Kiefer and Vogelsang (2002a) and PSJ. Jansson (2004) and Phillips, Sun and Jin (2005) have explained the simulation results using higher-order asymptotic theory. The disadvantage is lower level-corrected power compared to the conventional HAR tests. The lower power is clearly shown in the simulations reported by KV and PSJ. Hence, the power of the new tests of the intercept vector is of interest. We simulated the level-corrected powers of the new HAR tests for the sample sizes used in this study. The simulated powers show that the new tests have good power against empirically relevant alternatives. This evidence suggests that low power is not the explanation for the non-rejections by the new tests.

The organization of the paper is the following: Section 2 reviews the classical Wald test and the conventional HAR Wald test. The new HAR tests are described in Section 3. HAR test results using asymptotic *P*-values are reported in Section 4, and results using finite-sample *P*-values are presented in Section 5. The effect of the number of intercepts tested on the level of the tests is considered in Section 6. Section 7 reports on the powers of the new robust tests. The concluding comments are in Section 8.

2. CLASSIC AND HAR WALD TESTS

This section considers the three-factor model as a classical multivariate linear regression model with random regressors and reviews the classic and conventional Wald tests for the intercept vector.

Define the variables $y_1,..., y_N$, where y_i is the excess return for the ith portfolio or asset, and the variables x_1, x_2, x_3, where x_1 is the market factor (the excess return on the market portfolio), x_2 is the size factor (the difference between the return on a portfolio of small capitalization stocks and the return on a portfolio of large capitalization stocks (SMB, small minus big)) and x_3 is the book-to-market factor (the difference between the return on a portfolio of high-book-to-market stocks and the return to a portfolio of low-book-to-market stocks (HML, high minus low)).

Suppose that the conditional expectation function is linear,

$$E(y \mid x) = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3\,, \tag{1}$$

and the conditional variances are constant,

$$V(y \mid x) = \Sigma, \tag{2}$$

where $y = (y_1,...,y_N)'$, $\beta_0 = (\beta_{01},...,\beta_{0N})'$, $\beta_1 = (\beta_{11},...,\beta_{1N})'$, $\beta_2 = (\beta_{21},...,\beta_{2N})'$, and $\beta_3 = (\beta_{31},...,\beta_{3N})'$. A nonzero value of the intercept is interpreted as saying that the model leaves an unexplained return, a mean excess return that is unexplained by the three factors.

Denote the *t*th observation on *y* by $y_{\bullet t} = (y_{1t},...,y_{Nt})'$ and on *x* by $x_{\bullet t} = (x_{1t}, x_{2t}, x_{3t})'$. In addition, suppose the population of *y* and *x* is randomly sampled, that is, the pairs $(y_{\bullet t}, x_{\bullet t})$ are independently and identically distributed (*iid*). Then random sampling from the above multivariate population supports the classical multivariate linear regression model with random regressors. The assumption of *iid* returns is, of course, highly unrealistic and will be relaxed below to allow for empirically relevant volatility clustering effects.

Following Greene (2003), the multivariate regression model can be restated as a seemingly unrelated regressions (SUR) model with identical regressors for the purpose of presenting the classic and conventional robust Wald tests. The SUR model is formulated using the *N* regression equations $y_{i\bullet} = X\theta_i + u_{i\bullet}$, $(i = 1,..., N)$, where $y_{i\bullet} = (y_{i1},..., y_{iT})'$, $X = [\iota, x_{1\bullet}, x_{2\bullet}, x_{3\bullet}]$, $\iota = (1,...,1)'$, $x_{j\bullet} = (x_{j1},..., x_{jT})'$, $(j = 1, 2, 3)$, $\theta_{\bullet i} = (\beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i})'$, and $u_{i\bullet} = (u_{i1},...,u_{iT})'$. Stacking the *N* regressions,

$$y_{\bullet\bullet} = (I \otimes X)\theta_{\bullet\bullet} + u_{\bullet\bullet} = Z\theta_{\bullet\bullet} + u_{\bullet\bullet},$$

where I is an $N\times N$ identity matrix, $\theta_{\bullet\bullet}=(\theta_1',...,\theta_N')'$, and $u_{\bullet\bullet}=(u_{1\bullet}',...,u_{N\bullet}')'$. The least squares estimator of $\theta_{\bullet\bullet}$ is obtained by regressing $y_{\bullet\bullet}$ on Z. This produces the estimator $\hat{\theta}_{\bullet\bullet}=(Z'Z)^{-1}Z'y_{\bullet\bullet}=\theta+(Z'Z)^{-1}Z'u_{\bullet\bullet}$.

The null hypothesis of interest is $H_0:\beta_0=0$, and the alternative is $H_1:\beta_0\neq 0$. The classic Wald statistic for testing H_0 is

$$W=c^{-1}\hat{\beta}_0'\hat{\Sigma}^{-1}\hat{\beta}_0,$$

where $c=(X'X)^{11}$ is the 1,1-th element of the inverse of $X'X$, $\hat{\beta}_0=(I\otimes(1,0,0,0))\hat{\theta}_{\bullet\bullet}=R\hat{\theta}_{\bullet\bullet}$, $\hat{\Sigma}=T^{-1}\sum_t\hat{u}_{\bullet t}\hat{u}_{\bullet t}'$ and $\hat{u}_{\bullet t}=(y_{\bullet t}-\hat{\beta}_0-x_{1t}\hat{\beta}_1-x_{2t}\hat{\beta}_2-x_{3t}\hat{\beta}_3)$. Under suitable regularity conditions, the statistic W has a limiting chi-square distribution with N degrees of freedom when H_0 is true.

In the case where the $y_{\bullet t}$ are independently distributed as $N(\beta_0+x_{1t}\beta_1+x_{2t}\beta_2+x_{3t}\beta_3,\Sigma)$, or equivalently, the $u_{\bullet t}$ are *iid* $N(0,\Sigma)$, the statistic $F=((T-N-3)/NT)W$ is unconditionally distributed as central F with N degrees of freedom in the numerator and (T-N-3) degrees of freedom in the denominator. This follows from the fact that $(T\text{-}4)W/T$ is a generalized *Hotelling's* T^{-2} *statistic* where $[(X'X)^{11}]^{-1/2}\hat{\beta}_0$ is distributed as $N(0,\Sigma)$ under H_0, and $T\hat{\Sigma}$ is independently distributed as a Wishart with parameters (T-4) and Σ; see Anderson (1958, Theorem 5.2.2, p. 106). For further treatment of testing in the normal case, see Stewart (1997).

Campbell, Lo and MacKinlay (1997, p. 208) review the ample evidence in finance on the failure of the classical assumptions to hold when tested with real data. In particular, assumptions (1) and (2) may not hold and/or the sampling scheme may not produce independent observations. Consider the standardized estimator

$$\sqrt{T}(\hat{\theta}_{\bullet\bullet}-\theta_{\bullet\bullet})=(T^{-1}Z'Z)^{-1}(T^{-1/2}Z'u_{\bullet\bullet})=\hat{Q}^{-1}T^{-1/2}\sum_{t=1}^{T}v_{\bullet t}, \tag{3}$$

where $\hat{Q}=T^{-1}Z'Z=(I\otimes(T^{-1}X'X))$ and $v_{\bullet t}=u_{\bullet t}\otimes(1,x_{1t},x_{2t},x_{3t})'$. Under general assumptions, for example, those given in KV (2002a) and Phillips, Sun and Jin (2005),

$$\sqrt{T}(\hat{\theta}_{\bullet\bullet}-\theta_{\bullet\bullet})\sim N(0,Q^{-1}\Omega Q^{-1}) \tag{4}$$

where $Q = (I \otimes \bar{Q})$, $\bar{Q} = p\lim T^{-1} X'X$, and Ω is the long run variance of $\nu_{\bullet t}$ in (3). The conventional approach to HAR testing relies on HAC estimation of Ω and hence on consistent estimation of the sandwich variance matrix $Q^{-1}\Omega Q^{-1}$. The term $\bar{Q}$ can be consistently estimated by $T^{-1}X'X$. The method proposed by Newey and West (1987, 1994) for obtaining an HAC estimator of Ω is commonly used in practice. We use this HAC estimator for the conventional HAR test in this paper. The estimation of Ω is discussed in more detail in Section 3. The conventional HAR statistic is

$$W_{\text{HAC}} = T\hat{\beta}_0' \left[R\hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}R' \right]^{-1} \hat{\beta}_0 , \tag{5}$$

where $\hat{\Omega}$ is an HAC estimator of Ω. When $H_0 : \beta_0 = 0$ is true, W_{HAC} is asymptotically distributed as a chi-square with N degrees of freedom; for details, see KV.

3. NEW WALD HAR TESTS

This section presents a brief review of the new HAR Wald tests. The new tests considered in this section are based on either the Bartlett or the Parzen kernel.

Following PSJ consider an n-vector stationary process $\{z_t\}_{t=1}^{T}$ with non-singular spectral density matrix $f_{zz}(\lambda)$, $\lambda \in [0, \pi]$. The long run variance matrix of z_t is defined as

$$\Omega = \gamma_0 + \sum_{h=1}^{\infty} (\gamma_h + \gamma_h') = 2\pi f_{zz}(0) ,$$

where $\gamma_h = E(z_t z_{t-h}')$. To estimate Ω, a $4N \times 4N$ matrix in the three-factor model, consider the following lag kernel estimator of $f_{zz}(0)$,

$$\hat{f}_{zz}(0) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} k_\rho\left(\frac{h}{bT}\right)\hat{\gamma}_h , \tag{6}$$

$$\hat{\gamma}_h = \begin{cases} T^{-1}\sum_{t=1}^{T-h} z_{t+h} z_t' & \text{for } h \geq 0, \\ T^{-1}\sum_{t=-h+1}^{T} z_{t+h} z_t' & \text{for } h < 0, \end{cases}$$

where $k_{\rho}(x)$ is the kernel $k(x)$ raised to some positive integer power ρ. Here, as in KV, the bandwidth M has been set equal to bT where $b \in (0,1]$ is a constant (remains fixed) as $T \to \infty$.

When $k(x)$ is the Bartlett kernel,

$$k_{\rho}(x) = \begin{cases} (1-|x|)^{\rho} & |x| \le 1, \\ 0 & |x| > 1, \end{cases}$$

where $k_{\rho}(x)$ is the usual Bartlett kernel when ρ = 1. In this case, given b =1, PSJ (2004) call $\hat{f}_{zz}(0)$ the sharp origin estimator. When $k(x)$ is the Parzen kernel,

$$k_{\rho}(x) = \begin{cases} (1-6x^2+6|x|^3)^{\rho} & \text{for } 0 \le |x| \le 1/2, \\ (2(1-|x|)^3)^{\rho} & \text{for } 1/2 \le |x| \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k_{\rho}(x)$ is the usual Parzen kernel when ρ = 1. In this case, given b = 1, PSJ call $\hat{f}_{zz}(0)$ a steep origin estimator. See PSJ for a detailed discussion of the properties of the kernels.

The HAR test statistic for testing the null $H_0 : \beta_0 = 0$ is

$$W^* = T\hat{\beta}_0'[R\hat{Q}^{-1}\hat{\Omega}_{\rho}(b)\hat{Q}^{-1}R']^{-1}\hat{\beta}_0, \tag{7}$$

where $\hat{\Omega}_{\rho}(b) = 2\pi\hat{f}_{vv}(0)$ is the long run variance of $v_{\bullet t}$ and $\hat{f}_{vv}(0)$ is defined similarly to (6). Setting ρ = 1, expression (7) gives the HAR test statistic considered by KV statistic, and, setting b = 1, gives the statistic considered by PSJ. In this paper, the test statistic W^* is referred to as the fixed-b test statistic when ρ = 1 and the fixed-ρ test statistic when b = 1.

The asymptotic theory for the estimator $\hat{\Omega}_1(b)$ is developed by KV and for $\hat{\Omega}_{\rho}(1)$ by PSJ. The asymptotic theory for $\hat{\Omega}_1(b)$ depends on the rate conditions on b and T, and for $\hat{\Omega}_{\rho}(1)$ on the rate conditions on ρ and T. For example, PSJ show that $\hat{\Omega}_{\rho}(1)$ based on the Bartlett kernel is consistent for Ω when $\rho \to \infty$ as $T \to \infty$ and similarly for the Parzen kernel. In this case, under the null, the limiting distribution of W^* is chi-square with N degrees of freedom. By contrast, the estimator $\hat{\Omega}_\rho(1)$ is inconsistent when $T\to\infty$ for a fixed ρ. Consider testing p restrictions when ρ is fixed. PSJ show under the null that asymptotically

$$W^* \sim W_p'(1)\left[\int_0^1\int_0^1 k_\rho(r-s)dV_p(r)dV_p'(s)\right]^{-1} W_p(1) \tag{8}$$

where $W_p(r)$, $r\in[0,1]$, is a ($p\times 1$) vector of independent standard Wiener processes and $V_p(r)=W_p(r)-rW_p(1)$ is a ($p\times 1$) vector of standard Brownian bridges. The nonstandard limiting distribution (8) arises from the random limit of $\hat{\Omega}_\rho(1)$ when ρ is fixed as $T\to\infty$.

The asymptotic critical values for the fixed-b and the fixed-ρ tests are needed to implement the tests. These critical values have not been available in the literature except in special cases. We computed the critical values numerically by approximating the Wiener processes and Brownian bridges using normalized partial sums of *iid* $N(0, 1)$ deviates. Using 1,000 deviates, the fixed-b critical values of a 5 percent level test for p = 1, 2,…,10 were computed using 50,000 replications for the Bartlett kernel. The fixed-ρ critical values of a 5 percent level test for p = 1, 2,…,10 were computed using 10,000 replications for the Bartlett and Parzen kernels. Table I reports the fixed-b asymptotic critical values for b = 0.00625, 0.0125, 0.025, 0.05, 0.10, 0.125, 0.20, 0.25, 0.40, 0.80, 1.00, and Table II reports the fixed-ρ asymptotic critical values for ρ = 1, 4, 8, 16, 32, 64, 128.

The tabled results essentially match those for the untruncated Bartlett kernel in Kiefer and Vogelsang (Table 2, 2002a) for the case ρ = 1 and p = 1, 5 and 10, and those in PSJ (Table 3, 2003a) for the case ρ = 1, 4, 8, 16 and p = 1. The time required for computing an approximation to the fixed-b or fixed-ρ asymptotic critical value depends on the kernel. In the case of the Bartlett kernel, the double integral in (8) can be replaced by a single integral; for example, see KV and PSJ. This replacement substantially reduces the computational burden, which explains why more replications are used in the case of Bartlett than Parzen. In the special case where p = 1, b = 1 and ρ = 1, namely the case of the t statistic, Abadir and Paruolo (2002) and Kiefer and Vogelsang (2000b) have obtained the analytical form of the density.

The accuracy of the asymptotic critical values in Tables I and II depends on the number of discrete steps, that is, the number of $N(0,1)$ deviates used and the number of replications. In the case of the Bartlett kernel, the critical value for $b = \rho = 1$ for $p = 10$ was recalculated using 4,000 *iid* $N(0, 1)$ deviates and 50,000 replications. The value is 512.4677 with 1,000 deviates (Table I) and 511.8423 with 4,000 deviates, a change in the third digit. The computation time for 4,000 deviates was about 10 times that for 1,000 deviates. Turning to Table II, the value is 510.60 with 1,000 deviates and 10, 000 replications, again a change in the third digit. These results suggest that 1,000 deviates and 10,000 replications provide two-digit accuracy.

The objective of introducing the fixed-b test is to obtain better control over the ERP while maintaining the power of the conventional test. Hence, the benchmark to beat is the ERP of the conventional test with asymptotic chi-square critical values. In the simulation experiments, Kiefer and Vogelsang (2002a) estimated the ERP of the fixed-b test with $b = 1$ (untruncated kernel) when it uses fixed-b asymptotic critical values. The ERP tended to be smaller for the fixed-b test with $b = 1$ than for the conventional test. Jansson (2004) has explained this result using higher-order asymptotic theory in the context of a simple Gaussian location model. He shows that the ERP of the fixed-b test is of a smaller order for sample size T than the corresponding ERP of the conventional test.

In light of the analytical and simulation results for untruncated kernels, PSJ proposed a new class of HAR tests based on untruncated kernels ($b = 1$) raised to some positive integer exponent ρ, namely, the fixed-ρ tests. In the simulation results reported by PSJ, the ERP was smaller for the fixed-ρ test with fixed-ρ asymptotic critical values than for the conventional test. Also the ERP was smaller for the Parzen-based tests than the Bartlett-based tests. Phillips, Sun and Jin (2005) use higher-order asymptotics comparable to Jansson (2004) to analyze the properties of the fixed-ρ test in a Gaussian location model and obtain similar results.

4. ASYMPTOTIC TEST RESULTS

This section reports test results for three-factor models using the conventional HAR test and the new HAR tests when the tests are based on asymptotic P-values. The asymptotic P-values are obtained from asymptotic chi-square distribution for the conventional test and the simulated nonstandard asymptotic distributions for the fixed-b and fixed-ρ tests.

The return data consists of CRSP monthly returns, including distributions, for ten (N =10) value-weighted portfolios with NYSE, AMEX and NASDAQ stocks. The stocks are assigned to the portfolios based on market value of equity and quarterly rebalanced. The data for the three factors is taken from the Fama and French section of the Wharton Research Data Services (WRDS) website. The sample extends from January 1965 through December 2004 (T = 480). The one-month Treasury bill as reported on the Fama and French section of the WRDS website is used as a measure of the risk-free return. The tests are performed for five-year, ten-year, thirty-year sub-periods and longer periods. The sub-periods include those used by CLM plus additional periods made possible by more recent data.

The asymptotic P-values for the conventional and new HAR tests are presented in Table III. The asymptotic P-values for the conventional test reject the null at the 5 percent significance level for almost all of the five-year and ten-year sub-periods. Turning to the thirty-year and longer sub-periods, the null is rejected for all six sub-periods. The Newey and West (1987, 1994) version of the conventional HAR test uses a HAC estimator based on the truncated Bartlett kernel. The bandwidth for the tabled results is M = 6. The result are not qualitatively changed by using M = 4. A well known guideline for choosing the bandwidth for the Bartlett kernel is $M = 0.75T^{1/3}$; see Andrews (1991).

By contrast, the asymptotic P-values of the new HAR tests do not reject the null at the 5 percent significance level for most of the five-year and ten-year sub-periods. The P-values for the fixed-b tests are calculated using the Bartlett kernel and b =1 and those for the fixed-ρ tests use the Parzen kernel and ρ = 32. The results are similar for values of b greater than 0.10 and for ρ = 16 and 64. The null is also not rejected by the asymptotic P-values for three out of six thirty-year and longer sub-periods.

As noted in the introduction, there are two reasons for thinking that results of the conventional HAR test are problematic. The first is the absolute values of most of the estimated intercepts are economically insignificant for the five-year and ten-year sub-periods. Table IV reports the estimated values of the intercepts for all the sub-periods.

For the eight five-year sub-periods and the four ten-year sub-periods the absolute values of the estimated intercepts are for the most part less than 0 .001 and hence less than 10 basis points monthly. This is considered small in the finance literature; see Fama and French (1996, p.57). There are some exceptions. The economically big intercepts are concentrated in the smallest capitalization portfolio, portfolio ten. Thus, we have a situation where most of the intercepts are small in economic terms but the null is rejected when the conventional HAR test is applied. The second reason is that the asymptotic *P*-values of the new HAR tests do not reject the null for most of the five-year and ten-year sub-periods.

There are two possible explanations for the conflict between economic and statistical significance and the conflict between the conventional and new HAR test results. One is a power story. This says that the conventional test has high power against economically insignificant intercepts. An alternate version of the power story is that rejection is caused by the few economically significant intercepts. In either case, the assumption here is that the conventional test based on asymptotic critical values has close to the correct Type I error or level. The other is a level story or ERP story. This says that true level of the test is substantially larger than the nominal level or equivalently, the finite-sample *P*-value is substantially larger than the asymptotic *P*-value. The finite sample results reported in section 5 below provide support for the ERP story.

5. FINITE SAMPLE TEST RESULTS

This section reports simulated finite-sample *P*-values of the conventional and the new HAR tests. The *P*-values are calculated for the three forms of the HAR test in four different experiments. The four experiments are conducted for each of the sub-periods.

The purpose of the simulation experiments is to provide examples of the finite sample performance of the tests. The null that the intercepts are zero is a composite hypothesis because the values of the nuisance parameters are unknown in practice. The nuisance parameters include not only the slope parameters but also those that specify the process generating the factors and the errors. In our experiments, the values of the nuisance parameters are set equal to estimates based on the sample data. The level of the tests refers to the probability of a Type I error, not the size where the latter is defined as the maximum level over all admissible values of the nuisance parameters. In this paper, the simulated finite sample P-values are treated as exact, meaning that they are conditional on the values of the nuisance parameters used in the designs. This should be borne in mind when reviewing the discussion of the test results.

The experiments are now described for the January 1965 to 1969 sub-period. The value of $y_{\bullet t}$ is simulated using the constrained least squares estimate of the conditional expectation function (1) under the null:

$$y_{\bullet t}^{*} = x_{1t}^{*}\tilde{\beta}_1 + x_{2t}^{*}\tilde{\beta}_2 + x_{3t}^{*}\tilde{\beta}_3 + u_{\bullet t}^{*} \quad (t = 1, ..., T), \tag{9}$$

where $y_{\bullet t}^{*}$, $x_{\bullet t}^{*}$, $u_{\bullet t}^{*}$ are the simulated values of $y_{\bullet t}$, $x_{\bullet t}$, $u_{\bullet t}$ and $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3$, the constrained least squares estimates of the slope vectors calculated from the sample data for the sub-period.

Normal-Normal (NN) *P*-value Experiment. This experiment produces data that satisfies the assumptions of the classical normal SUR model with normally distributed regressors. The P-value simulation procedure consists of five steps:

S1. Generate a sample of T = 60 $x_{\bullet t}^{*}$ vectors by randomly sampling the $N(\bar{x}, S)$ distribution where $\bar{x} = T^{-1}\sum_t x_{\bullet t}$ and $S = T^{-1}\sum_t (x_{\bullet t} - \bar{x})(x_{\bullet t} - \bar{x})'$ are calculated from sample data for the sub-period.

S2. Generate a sample of T = 60 $u_{\bullet t}^{*}$ vectors independently of $x_{\bullet t}^{*}$ by randomly sampling the $N(0, \tilde{\Sigma})$ distribution where $\tilde{\Sigma} = T^{-1}\sum_t (\tilde{u}_{\bullet t} - \bar{\tilde{u}}_{\bullet})(\tilde{u}_{\bullet t} - \bar{\tilde{u}}_{\bullet})'$ and $\bar{\tilde{u}}_{\bullet} = T^{-1}\sum_t \tilde{u}_{\bullet t}$ are calculated from the constrained residual vectors $\tilde{u}_{\bullet t} = (y_{\bullet t} - x_{1t}\tilde{\beta}_1 - x_{2t}\tilde{\beta}_2 - x_{3t}\tilde{\beta}_3)$ for the sub-period.

S3. Generate a sample of T = 60 $y_{\bullet t}^{*}$ vectors from (9) using the $x_{\bullet t}^{*}$ vectors from S1, the $u_{\bullet t}^{*}$ vectors from S2 and the constrained least squares estimates as the values for the slope parameters.

S4. Compute the three forms of the HAR test statistic from the simulated dataset of size T = 60.

S5. Repeat steps S1, S2, S3 and S4 10, 000 times. Compute the P-value for each form of the HAR test statistic from the empirical distribution of the test statistic.

Resample-Resample (RR) *P*-value Experiment. This experiment captures the non-normality present in the data. In this and the remaining experiments, only one or both of the first two steps differ from those in the NN experiment.

S1′. Generate a sample of $T = 60$ $x^*_{\bullet t}$ vectors by randomly sampling with replacement the observations $x_{\bullet t}$.

S2′. Generate a sample of $T = 60$ $u^*_{\bullet t}$ vectors independently of $x^*_{\bullet t}$ by randomly sampling with replacement the demeaned constrained least squares residuals $\tilde{u}_{\bullet t} - \bar{\tilde{u}}_{\bullet}$.

Normal-VAR (NV) *P*-value Experiment. This experiment introduces serial correlation in the errors. The first step is the same as in the NN experiment.

S2″. Generate a sample of $T = 60$ $u^*_{\bullet t}$ vectors independently of $x^*_{\bullet t}$ using a Gaussian VAR(1) process

$$u^*_{\bullet t} = \tilde{\Phi} u^*_{\bullet t-1} + \eta^*_{\bullet t},$$

where $\tilde{\Phi}$ is a 10×10 matrix of autoregressive coefficients. The autoregressive matrix $\tilde{\Phi}$ is obtained by a least squares regression of $\tilde{u}_{\bullet t}$ on $\tilde{u}_{\bullet t-1}$ using the constrained least square residuals for the sub-period. The vector $\eta^*_{\bullet t}$ is randomly sampled from the $N(0, \tilde{\Sigma}_\eta)$ distribution, where $\tilde{\Sigma}_\eta = T^{-1}\sum_{t=1}^{T}(\tilde{\eta}_{\bullet t} - \bar{\tilde{\eta}}_{\bullet})(\tilde{\eta}_{\bullet t} - \bar{\tilde{\eta}}_{\bullet})'$ and $\bar{\tilde{\eta}}_{\bullet} = T^{-1}\sum_t \tilde{\eta}_{\bullet t}$ are calculated from the VAR residuals. The conditions for covariance-stationarity are checked by calculating the roots of the $\tilde{\Phi}$ matrix. In each replication, the initial values of $u^*_{\bullet t-1}$ in the VAR (1) are set equal to zero, and the first 200 draws are discarded in order make the results independent of the initial values.

Resample-Block (RB) *P*-value Experiment. This experiment allows for volatility clustering of the returns. The first step is the same as in the RR experiment.

S2‴. Generate a sample of $T = 60$ $u^*_{\bullet t}$ vectors independently of $x^*_{\bullet t}$ by randomly sampling with replacement the demeaned constrained least squares residuals $\tilde{u}_{\bullet t} - \bar{\tilde{u}}_{\bullet}$ in consecutive fixed-length non-overlapping blocks where the block length is six months.

The NN and RR experiments provide evidence on how the tests perform when the multivariate *iid* assumption holds with and without normality. If the tests exhibit poor performance under this assumption, it is unlikely that they will perform well in the presence of autocorrelation or volatility clustering.

The rationale for the NV and RB experiments is a substantial body of evidence in finance documenting departures from the *iid* assumption. As a preliminary check for heteroskedasticity and autocorrelation, the Breusch and Pagan (1979) test for heteroskedasticity as modified by Cook and Weisberg (1983) and the Breusch-Godfrey test for autocorrelation (Godfrey (1988)) were performed equation-by-equation for each sub-period. The tests tended to reject for the smaller capitalization portfolios at the usual levels, especially for the ten-year and longer sub-periods.

There is considerable evidence that asset return volatility is both time-varying and predictable; for example, see Bollerslev (1986) and Bollerslev, Engle and Nelson (1995). As a preliminary check for autoregressive conditional heteroskedasticity, a GARCH (1, 1) model for the errors was estimated equation-by-equation for each sub-period. The maximization of the pseudo-log likelihood frequently failed for the five-year sub-periods and for some of the ten-year sub-periods. Estimates of the ARCH and GARCH coefficients were obtained for the thirty-year and longer sub-periods, and these were often significantly different from zero.

In the simulation experiments, it was not feasible to generate the errors for each period using an estimated multivariate GARCH model. Instead, we use a procedure that is employed in bootstrap sampling with dependent data. The procedure is to divide the residual vectors for each sub-period into blocks, and then randomly resample the blocks with replacement. In the RB experiments, six-month length blocks were chosen because this is approximately the half-life of an estimated univariate GARCH process for monthly stock returns; for example, see French, Schwert and Stambaugh, (1987) for estimates for the period 1928-1984.

More generally, the RB experiments capture dependence in the errors. There are other processes that may be generating dependence in addition to autoregressive conditional heteroskedasticity. These include ARMA models and also models that produce non-martingale difference sequences such as nonlinear moving average and bilinear models. Consequently, the results of the RB experiments cannot be interpreted as only due to volatility clustering, although this may be the dominant effect.

Table V presents the simulated finite-sample *P*-values for the conventional and new HAR tests. The message from this table is that the null is not rejected at the five percent level by the simulated finite-sample *P*-values for most of the five-year and ten-year sub-periods for the three forms of the HAR test.

For the conventional test, the differences between the asymptotic and simulated finite-sample *P*-values for the five-year sub-periods are quite large in all four experiments. This suggests that the conventional test based on asymptotic *P*-values produces misleading inferences when testing the three-factor model. For these sub-periods, the conflict between the economic and statistical significance is much reduced when the conventional HAR test is based on finite-sample *P*-values. As expected, the difference between the asymptotic and finite-sample *P*-values decreases as the length of the sample period increases with the consequence that the rejections for the thirty-year and longer sub-periods are the same for both asymptotic and finite-sample *P*-values with the exception of one sub-period.

For the fixed-*b* and the fixed-ρ tests, there is almost no conflict between the inferences based on the asymptotic and simulated finite-sample *P*-values for the five-year and ten-year periods. The same is true for the six thirty-year and longer sub-periods. Hence, the null is not rejected for three out of six thirty-year and longer sub-periods based on asymptotic and finite-sample *P*-values.

Two more general features of the results merit attention. The first is that the difference between the asymptotic and finite-sample *P*-values is much smaller for the new HAR tests than for the conventional test. This is consistent with the simulation results of KV and PSJ and the theoretical results obtained by Jansson (2004) and PSJ (2005). Second, the differences between the asymptotic and finite-sample *P*-values tend to be smaller for the fixed-ρ test than for the fixed-*b* test.

6. MULTIVARIATE COMPLICATIONS

This section reports the effect of increasing the number of intercepts tested on the rejection probabilities of the conventional and new HAR tests. The rejection probabilities for the three forms of the HAR test with asymptotic critical values are simulated in four different experiments.

Consider a three-factor model with i equations and hence i intercepts. Model i is

$$y_{\bullet t}^{i} = \beta_0^i + x_{1t}\beta_1^i + x_{2t}\beta_2^i + x_{3t}\beta_3^i + u_{\bullet t}^i \quad (i = 1,...,10,\ t = 1,...,T), \tag{10}$$

where $y_{\bullet t}^{i} = (y_{1t},..., y_{it})'$, $\beta_0^i = (\beta_{01},...,\beta_{0i})$, $\beta_1^i = (\beta_{11},...,\beta_{1i})'$, $\beta_2^i = (\beta_{21},...,\beta_{2i})'$, $\beta_3^i = (\beta_{31},...,\beta_{3i})'$ and $u_{\bullet t}^i = (u_{it},...,u_{it})'$. The ordering of the models and equations makes use of the fact that the portfolios are ordered by market equity. The ith intercept is the intercept of the equation for the ith portfolio of stocks.

For the ith model, the null hypothesis of interest is $H_0^i : \beta_0^i = 0$, and the alternative is $H_1^i : \beta_0^i \neq 0$. The null H_0^i is tested for the ith model using the conventional, fixed-b and fixed-ρ tests with five percent asymptotic critical values. The finite-sample levels of the tests for the ith model are obtained by simulation. In the simulation experiments, the null $H_0^i : \beta_0^i = 0$ is imposed. In the ith model, the value of $y_{\bullet t}$ is simulated using

$$y_{\bullet t}^{i*} = x_{1t}^{*}\tilde{\beta}_1^i + x_{2t}^{*}\tilde{\beta}_2^i + x_{3t}^{*}\tilde{\beta}_3^i + u_{\bullet t}^{i*} \quad (i = 1,...,10,\ t = 1,...,T),$$

where $y_{\bullet t}^{*i}$, $x_{\bullet t}^{*}$, $u_{\bullet t}^{*i}$ are the simulated values of $y_{\bullet t}^{i}$, $x_{\bullet t}$, $u_{\bullet t}^{i}$ and $\tilde{\beta}_1^i, \tilde{\beta}_2^i, \tilde{\beta}_3^i$ are the constrained least squares estimates of the slope vectors. The slope estimates are calculated using the data for January 1965 through December 2004. The rejection probabilities are simulated for T = 60, 120 and 240.

Normal-Normal (NN) Rejection Probability Experiments. For T = 60, the simulation procedure for the rejection probabilities consists of five steps:

S1. Generate a sample of T = 60 $x_{\bullet t}^{*}$ vectors by randomly sampling the $N(\bar{x}, S)$ distribution where $\bar{x} = T^{-1}\sum_t x_{\bullet t}$ and $S = T^{-1}\sum_t (x_{\bullet t} - \bar{x})(x_{\bullet t} - \bar{x})'$ are calculated from the entire sample.

S2. Generate a sample of T = 60 $u_{\bullet t}^{i*}$ vectors independently of $x_{\bullet t}^{*}$ by randomly sampling the $N(0, \tilde{\Sigma}^i)$ distribution, where $\tilde{\Sigma}^i = T^{-1}\sum_t (\tilde{u}_{\bullet t}^i - \bar{\tilde{u}}_{\bullet}^i)(\tilde{u}_{\bullet t}^i - \bar{\tilde{u}}_{\bullet}^i)'$ and

$\bar{\tilde{u}}^i_{\bullet} = T^{-1}\sum_t \tilde{u}^i_{\bullet t}$ are calculated from the constrained residual vectors

$\tilde{u}^i_{\bullet t} = (y^i_{\bullet t} - x_{1t}\tilde{\beta}^i_1 - x_{2t}\tilde{\beta}^i_2 - x_{3t}\tilde{\beta}^i_3)$ from the entire sample.

S3. Generate a sample of T= 60 $y^{*i}_{\bullet t}$ for the ith model using the $x^*_{\bullet t}$ vectors from S1, the $u^{i*}_{\bullet t}$ vectors from S2 and the constrained least squares estimates as the values for the slope parameters.

S4. Compute the three forms of the HAR test statistic from the simulated dataset of size T = 60.

S5. Repeat steps S1, S2, S3 and S4 10, 000 times. Compute the rejection probabilities for each form of the HAR test statistic from the empirical distributions using five percent asymptotic critical values.

The steps are similar for T = 120 and 240. The steps in the RR, NV and RB rejection probability experiments are obtained by making the analogous changes to the RR, NV and RB *P*-value simulation experiments.

Table VI reports the results for the NN experiments. The number of intercept parameters has a very strong effect on the simulated levels for the Bartlett-based conventional robust test with M = 6. The results for T = 60 show that the ERP is about 4.5 percent for the one equation model and 70 percent for the ten equation model, about a fourteen-fold increase in the ERP as the number of intercept parameters tested is increased from one to ten. Given T = 120, the ERP is about 3 percent for the one equation model and about 35 percent for all ten equations. Although the ERP is not large for one parameter, it is very substantial for ten parameters. A point worth stressing is that the distortion in the level in the NN experiments is not due to a violation of the *iid* assumption or the normality assumption.

Procedures such as prewhitening have been proposed for improving the performance of conventional robust Wald tests, for example, Andrews (1991) and Andrews and Monahan (1992). We note that in the NN and RR experiments performance will not be significantly improved by prewhitening since these experiments generate *iid* data.

Table VI reports the effect of the number of intercepts on the level of the fixed-b and fixed-ρ test when the tests use five percent asymptotic critical values. For the fixed-b

test, the effect of the number of intercepts is much reduced, although not eliminated. The number of intercepts has a small effect on the Parzen-based test with $\rho = 32$. For $T = 60$, the ERP is about 1 percent for the one equation model and about 2 or 3 percent for the ten equation model. For $T = 120$ and 240, the ERP tends to be less than 1 percent for all ten equations.

The increase in level distortion of as the number of intercepts increases is roughly the same in the RR, NV and RB experiments. See the supplemental Appendix on the JAE website. Experiments were also carried out when the equation ordering is reversed. In this case, the excess return of the first equation is the excess return of the tenth portfolio (the portfolio of stocks with the smallest market value of equity), and the excess return of the tenth equation is that of the first portfolio (the portfolio of stocks with the largest market value of equity).

7. POWER OF NEW HAR TESTS

This section reports simulated level-corrected powers of the conventional and the new HAR tests. The level-corrected powers are calculated for the three forms of the HAR test in four different experiments. The four experiments are conducted for each of the sub-periods.

The simulated powers are estimates of the true level-corrected powers conditional on the experimental design. The design specifies the vector of intercepts under the alternative, the nuisance parameters including the slope vectors and the long run variance matrix and the process generating the factors as well the errors.

The powers are calculated for a test of H_0 against the alternative $H_1 : \beta_0 = c\iota(0.0005), |c| > 0$. Here the alternative intercept vector β_0 is proportional to a vector of ones, ι, where c is a scalar. With this setup, a unit increase in c translates into an increase in the monthly excess return of 5 basis points. As noted earlier, in finance a monthly excess return of 10 basis points ($c = 2$) is considered small. On the other hand, a monthly excess return of 50 basis points ($c = 10$) is considered large by traditional benchmarks. One benchmark is the equity premium. This is about 6 percent per annum, which translates into a monthly excess return of 50 basis points. Another is the monthly excess return on the market portfolio, which is between 80 and 100 basis points. Hence, this setup provides a natural metric for interpreting the power, which is often absent in power studies.

The power experiments are now described for the January 1965 to 1969 sub-period. The value of $y_{\bullet t}$ is simulated using

$$y^*_{\bullet t} = \beta_0 + x^*_{1t}\tilde{\beta}_1 + x^*_{2t}\tilde{\beta}_2 + x^*_{3t}\tilde{\beta}_3 + \tilde{u}^*_{\bullet t} \quad (t = 1,...,T), \tag{11}$$

where $y^*_{\bullet t}$, $x^*_{\bullet t}$, $u^*_{\bullet t}$ are the simulated values of $y_{\bullet t}$, $x_{\bullet t}$, $u_{\bullet t}$. The intercept vector β_0 is known constant given by the alternative H_1. The slopes $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3$ are obtained by running a constrained least squares regression of $y_{\bullet t}$ on $x_{\bullet t}$ for the sample data where the constraint is $\beta_0 = 0$.

Normal-Normal (NN) Power Experiment. The power simulation procedure consists of five steps for each value of *c*. For *c* = 0, steps S1, S2, S3 and S4 are the same as in the *P*-value simulation procedure. The fifth step is:

S5. Repeat steps S1, S2 S3 10, 000 times. Compute the 5 percent critical value for each form of the HAR test statistic from the empirical distribution of the test statistic under H_0 ($c = 0$).

For *c* = 1, steps S1, S2, S3, S4 are the same as the *P*-value simulation procedure. The fifth step is:

S5. Repeat steps S1, S2, S3 and S4 10,000 times. Compute the power for each form of the HAR test statistic from the empirical distribution of the test statistic using the simulated five percent critical value obtained from the *c* = 0 experiment.

For *c* > 1, the power experiments are similar to those for *c* = 1.

The steps in the RR, NV and RB power simulation experiments are obtained by making the analogous changes to the RR, NV and RB *P*-value simulation experiments.

The powers for the NN experiments are reported in Table VII. The powers are reported only for positive values of *c* since the power curves are symmetric in *c*. The results show that the tests tend to have high level-corrected power against empirically relevant departures from the null. The level-corrected powers tend to be close to one at *c* = 2 (monthly excess return of 10 basis points) for *T* = 60 and close to one at *c* =1.5 (monthly excess return of 7.5 basis points) for *T* = 120. The power results support the conclusion that the non-rejections by the fixed-*b* tests and fixed-ρ tests are due to small intercepts rather than low power. The same conclusion is supported by the results from the RR, NV and RB power experiments. Again see the supplemental Appendix on JAE website.

Table VII shows that the powers do depend on the kernel and hence on choice of the HAR test. Additional simulations show that the powers of the fixed-b tests tend to increase as b decreases and the powers of the fixed-ρ tests tend to increase as ρ increases. These results are consistent with the findings in KV and PSJ. However, this does not imply that a small b should be chosen for the fixed-b test or a large ρ for the fixed-ρ test. This is because as b decreases the ERP of the fixed-b test increases and as ρ increases the ERP of the fixed-ρ test increases. The trade-off between the ERP and power is analyzed in detail in Phillips, Sun and Jin (2005).

8. CONCLUDING COMMENTS

This paper reports the results of testing the Fama and French three-factor model using the conventional and the new HAR tests developed by Kiefer and Vogelsang (2005) and Phillips, Sun and Jin (2003a, 2003b). The zero intercept null is rejected for most sub-periods by the conventional HAR test with the asymptotic *P*-values. This occurs even though the absolute values of the estimated intercepts are considered small in economic terms. Simulation results show that the conventional test with asymptotic *P*-values over-rejects. The null is not rejected for most sub-periods by the new tests with asymptotic *P*-values. The new tests produce similar results with finite-sample *P*-values. The power results show that the new tests have high power against empirically relevant alternatives. The conflict between the conventional test and the new tests for the five-year and ten-year sub-periods tends to disappear when the tests are based on simulated finite-sample *P*-values.

We conclude with two remarks. One is that our experimental results are consistent with the theory provided by Jansson (2004) for fixed b-asymptotics and Phillips, Sun and Jin (2005) for fixed-ρ asymptotics. However, these authors have focused on testing a single parameter in a Gaussian location model. Additional theory is needed for the more general class of multivariate models considered in this paper. The second is that Gonclaves and Vogelsang (2006) as well as Kiefer and Vogelsang (2005) apply the naïve block bootstrap to HAR tests in simulation studies. The studies suggest that the naïve block bootstrap works much better than predicted by the existing literature. Goncalves and Vogelsang (2006) establish that the naïve block bootstrap yields an approximation to the distributions of HAR test statistics that is as accurate as the approximation obtained by fixed-*b* asymptotic theory. A challenging research project is to investigate whether the naïve block bootstrap provides an asymptotic refinement relative to fixed-*b* asymptotics.

ACKNOWLEDGEMENTS

The authors are grateful to Tim Bollerslev and two anonymous referees for their helpful and constructive comments, which led to improvements in both the content and presentation of an earlier version of this paper. We are also grateful to John Geweke, John Nankervis, Ashish Tiwari and Paul Weller for useful suggestions.

REFERENCES

Abadir, K. M. and Paruolo, P., 2002. Simple robust testing of regression hypotheses: a comment. *Econometrica* 70: 2097-2099.

Anderson, T. W., 1958. *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons: New York.

Andrews, D.W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817-858

Andrews, D. W. K., Monahan, J. C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60: 953-966.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31: 307-327.

Bollerslev, T., Engle, R. R., Nelson, D. B., 1994. Arch Models. In: *Handbook of Econometrics*, vol. 4, Engle, R. R., McFadden, D. L. (eds). North-Holland: Amsterdam.

Breusch, T., Pagan, A., 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47: 1287-1294.

Campbell, J.Y., Lo, A. W., MacKinlay, A. C., 1997. *The Econometrics of Financial Markets*. Princeton University Press: Princeton, New Jersey.

Cochrane, J. H., 1999. New facts in finance. *Econometric Perspectives* XXIII (3): (Federal Reserve Bank of Chicago) 36-58.

Cochrane, J. H., 2005. *Asset Pricing*, revised edition. Princeton University Press: Princeton, New Jersey.

Cook, R.D., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. *Biometrika* 70: 159-178.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3-56.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. *The Journal of Finance* 51: 55-83.

French, K.R., G. W. Schwert, Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19: 3-30.

Godfrey, L. G., 1988. *Misspecification Tests in Econometrics*. Econometric Society Monographs, No. 16. Cambridge University Press: Cambridge

Goldberger, A. S., 1991. *A Course in Econometrics*. Harvard University Press: Cambridge.

Goncalves, S., Vogelsang, T.J., 2006. Block bootstrap HAC robust test: the sophistication of the naïve bootstrap. Working paper, Department of Economics, Cornell University.

Greene, W. H., 2003. *Econometric Analysis*, fifth edition. Prentice Hall: New Jersey.

Jansson, M., 2004. The error in rejection probability of simple autocorrelation robust tests. *Econometrica* 72: 937-946.

Kiefer, N. M., Vogelsang, T. J., Bunzel, H., 2000. Simple robust testing of regression hypotheses. *Econometrica* 68: 695-714.

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

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

Kiefer, N. M., Vogelsang, T. J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21: 1130-1164.

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

Newey, W.K, West, K.D., 1994. Automatic lag selection in covariance estimation. *Review of Economics Studies* 61: 631-654.

Phillips, P.C. B., 2005. HAC estimation by automated regression. *Econometric Theory* 21: 116-142.

Phillips, P.C.B., Sun, Y., Jin, S., 2003a. Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. Cowles Foundation Discussion Paper No. 1407, Yale University.

Phillips, P.C. B., Sun, Y., Jin, S., 2003b. Long run variance estimation using steep origin kernels without truncation. Cowles Foundation Discussion Paper No. 1437, Yale University.

Phillips, P.C. B., Sun, Y., Jin, S., 2005. Improved HAR inference using power kernels without truncation. Cowles Foundation Discussion Paper No. 1513, Yale University.

Stewart, K. G., 1997. Exact testing in multivariate regression. *Econometric Reviews* 16: 321-352.

Table I. Fixed-*b* asymptotic 5 percent critical values using Bartlett kernels

Power Parameter	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
$b = 0.00625$	4.06	6.38	8.38	10.18	11.97	13.69	15.32	17.02	18.71	20.31
$b = 0.0125$	4.17	6.58	8.60	10.53	12.50	14.28	16.05	18.09	19.78	21.79
$b = 0.025$	4.11	6.54	8.77	11.02	12.96	15.18	17.12	19.41	21.69	24.08
$b = 0.05$	4.40	7.27	9.76	12.67	15.27	18.26	21.36	24.81	27.95	31.96
$b = 0.10$	5.01	8.67	12.48	16.61	21.22	26.19	32.40	38.90	45.89	54.23
$b = 0.125$	5.24	9.47	13.86	18.99	24.85	31.75	39.36	48.10	57.31	68.36
$b = 0.20$	6.46	12.47	19.41	27.78	37.87	49.85	63.15	78.45	94.13	112.65
$b = 0.25$	7.36	14.41	23.79	34.55	47.74	62.96	80.19	98.89	119.10	141.43
$b = 0.40$	9.89	21.79	36.75	55.67	78.03	101.34	127.70	157.10	190.83	223.06
$b = 0.50$	11.97	27.25	45.97	67.92	95.52	125.77	157.95	193.50	232.12	274.00
$b = 0.80$	18.39	41.95	71.08	106.32	146.57	192.82	241.42	294.96	356.23	417.62
$b = 1.00$	22.57	52.94	87.24	129.56	180.07	234.56	291.15	360.25	431.30	512.47

Note: The Brownian motion and Brownian bridge processes were approximated using normalized partial sums of $T = 1000$ *iid* $N(0,1)$ random variables. The simulations used 50,000 replications.

Table II. Fixed-ρ asymptotic 5 percent critical values using Bartlett and Parzen kernels

Power	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
Bartlett										
$\rho = 1$	22.27	52.39	87.44	127.24	177.07	226.99	293.00	351.95	428.96	510.60
$\rho = 4$	8.55	16.47	24.79	35.92	48.18	61.77	76.61	91.36	113.92	131.40
$\rho = 8$	6.11	11.13	15.98	22.19	28.55	35.16	42.67	51.68	60.16	70.74
$\rho = 16$	5.12	8.20	11.61	15.41	19.19	23.15	26.80	32.05	36.61	41.26
$\rho = 32$	4.47	7.07	9.91	12.41	14.67	17.64	20.28	22.69	25.85	29.02
$\rho = 64$	4.06	6.55	8.60	11.04	12.78	14.82	16.79	19.03	21.17	23.35
$\rho = 128$	3.95	6.16	8.21	10.16	11.84	13.67	15.46	17.28	19.18	20.58
$\rho = O(T^{2/3})$	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51	16.92	18.31
Parzen										
$\rho = 1$	31.28	147.01	535.10	1369.90	2781.40	4747.50	7896.16	12626.82	18618.29	27975.88
$\rho = 4$	11.56	28.93	64.13	150.96	309.72	599.02	1080.68	1851.99	3018.91	4450.50
$\rho = 8$	8.52	18.36	34.83	64.76	110.01	191.28	333.62	570.04	959.41	1476.88
$\rho = 16$	6.62	12.76	21.93	34.57	51.84	77.87	116.65	176.11	279.85	398.61
$\rho = 32$	5.70	10.38	15.99	22.49	30.27	41.71	56.87	72.99	98.32	135.53
$\rho = 64$	5.01	8.67	12.52	17.52	22.23	27.53	34.91	43.60	52.59	66.08
$\rho = 128$	4.50	7.53	10.61	13.85	17.87	21.80	25.53	30.24	35.66	42.74
$\rho = O(T^{2/3})$	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Note: The Brownian motion and Brownian bridge processes were approximated using normalized partial sums of T =1000 *iid* $N(0,1)$ random variables. The simulations used 10,000 replications.

Table III. Asymptotic P-values (%) for HAR tests

Sub-Period	Conventional: Barlett		Fixed-b: Barlett		Fixed-ρ: Parzen	
	W_{HAC} $M = 6$	P-value	W^* $b = 1$	P-value	W^* $\rho = 32$	P-value
Five-Year						
1/65-12/69	56.59	0.00	527.14	4.35	82.67	16.12
1/70-12/74	22.17	1.43	208.95	45.10	44.72	42.86
1/75-12/79	49.32	0.00	441.65	8.30	88.38	14.08
1/80-12/84	54.16	0.00	358.77	15.07	77.29	18.31
1/85-12/89	23.10	1.04	239.92	36.33	27.61	67.94
1/90-12/94	33.28	0.02	253.19	32.93	94.99	12.21
1/95-12/99	58.71	0.00	543.50	3.98	374.2	0.22
1/00-12/04	22.33	1.35	201.57	47.50	28.67	66.25
Ten-Year						
1/65-12/74	28.82	0.13	343.79	16.66	43.746	43.97
1/75-12/84	41.73	0.00	522.24	4.54	174.02	2.34
1/85-12/94	21.88	1.57	360.43	14.90	51.276	36.25
1/95-12/04	24.11	0.73	416.63	10.03	124.79	6.23
Thirty-Year						
1/65-12/94	43.33	0.00	979.70	0.18	195.75	1.78
1/70-12/99	25.80	0.40	536.55	4.25	163.27	2.93
1/75-12/04	22.73	1.18	390.49	12.45	89.30	13.84
More Years						
1/65-12/99	33.59	0.02	692.47	1.20	109.78	8.57
1/70-12/04	25.35	0.47	417.83	10.10	135.84	4.98
1/65-12/04	33.33	0.00	481.45	6.22	151.78	3.61

Note: The tabled asymptotic P-values are computed by simulation using 10,000 replications of each experiment

Table IV. Estimated intercepts for the ten portfolios sorted by market equity

Sub-Period	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5	Portfolio 6	Portfolio 7	Portfolio 8	Portfolio 9	Portfolio 10
Five-Year										
1/65-12/69	3.822E-04	-2.362E-04	1.320E-03	-4.658E-04	9.617E-04	2.176E-03	-2.221E-03	1.103E-04	-2.349E-03	-9.040E-04
1/70-12/74	4.281E-04	-6.098E-04	1.246E-03	2.588E-04	5.778E-04	-1.536E-04	-3.791E-04	-2.646E-04	-2.208E-03	-1.007E-03
1/75-12/79	-1.161E-04	-7.023E-04	1.686E-03	1.139E-03	9.320E-04	1.804E-03	2.582E-04	4.974E-04	-2.404E-03	-4.070E-03
1/80-12/84	1.317E-03	-3.560E-05	3.682E-04	-7.732E-05	-9.035E-04	7.401E-05	-1.182E-03	-1.141E-03	-2.044E-03	-6.015E-03
1/85-12/89	-2.149E-05	4.837E-04	7.247E-04	1.646E-03	1.053E-03	9.673E-04	6.799E-04	9.882E-04	-1.879E-03	-4.430E-03
1/90-12/94	2.761E-04	1.547E-04	3.490E-04	5.801E-04	2.195E-03	1.154E-03	8.797E-04	-7.023E-04	-4.065E-04	-5.559E-04
1/95-12/99	2.010E-03	-2.419E-03	-1.826E-03	-1.514E-03	-5.114E-03	-8.584E-04	-6.192E-04	4.593E-05	1.072E-03	-4.102E-04
1/00-12/04	-4.784E-04	9.241E-04	-2.439E-04	1.290E-04	-5.197E-04	8.961E-04	-1.492E-03	1.821E-03	3.585E-03	1.278E-02
Ten-Year										
1/65-12/74	4.470E-04	-6.370E-04	9.400E-04	0.000E+00	4.560E-04	3.360E-04	-1.283E-03	-6.180E-04	-2.657E-03	-1.199E-03
1/75-12/84	7.680E-04	-6.780E-04	4.760E-04	-6.800E-05	-3.320E-04	6.840E-04	-1.105E-03	-1.051E-03	-2.858E-03	-6.063E-03
1/85-12/94	2.270E-04	3.790E-04	5.530E-04	9.790E-04	1.520E-03	9.510E-04	8.930E-04	2.660E-04	-8.390E-04	-2.094E-03
1/95-12/04	1.221E-03	-5.370E-04	-1.263E-03	-1.091E-03	-3.128E-03	-7.050E-04	-1.794E-03	-1.210E-04	8.360E-04	3.373E-03
Thirty-Year										
1/65-12/94	3.980E-04	-1.960E-04	6.350E-04	3.900E-04	6.230E-04	8.630E-04	-3.270E-04	-3.000E-04	-1.925E-03	-2.675E-03
1/70-12/99	7.100E-04	-7.790E-04	1.110E-04	-8.400E-05	-2.630E-04	6.640E-04	-2.560E-04	-2.640E-04	-1.368E-03	-2.781E-03
1/75/12/04	7.650E-04	-2.170E-04	-9.600E-05	1.900E-05	-5.580E-04	3.590E-04	-6.810E-04	-2.700E-04	-9.260E-04	-1.589E-03
More Years										
1/65-12/99	6.960E-04	-6.540E-04	2.640E-04	-4.400E-05	-9.600E-05	6.280E-04	-5.230E-04	-3.110E-04	-1.462E-03	-2.262E-03
1/70-12/04	7.260E-04	-3.370E-04	9.400E-05	-1.700E-05	-4.460E-04	4.180E-04	-5.390E-04	-2.900E-04	-1.248E-03	-2.000E-03
1/65/12/04	6.590E-04	-3.030E-04	2.540E-04	2.300E-05	-2.380E-04	5.090E-04	-6.850E-04	-2.510E-04	-1.254E-03	-1.533E-03

Note: The first portfolio is the portfolio of stocks with the largest market value of equity and the tenth portfolio is the one with the smallest market value of equity.

Table V. Simulated finite-sample P-values (%) for HAR tests

Sub-Period	Conventional test: Bartlett $M = 6$				Fixed-b: Bartlett $b = \rho = 1$				Fixed-ρ : Parzen $\rho = 32$			
Five-Year	NN	RR	NV	RB	NN	RR	NV	RB	NN	RR	NV	RB
1/65-12/69	16.07	18.74	11.63	21.53	10.77	13.20	7.56	16.17	22.99	25.13	20.28	35.70
1/70-12/74	66.61	68.84	62.64	80.86	56.95	60.09	52.92	75.45	50.17	52.97	48.21	72.63
1/75-12/79	22.17	25.30	18.64	25.76	17.81	20.73	14.47	20.91	20.77	23.01	19.01	29.61
1/80-12/84	17.78	20.19	13.94	32.07	26.87	29.67	21.30	40.66	25.12	27.25	22.95	44.64
1/85-12/89	64.42	68.49	57.73	64.81	49.03	53.55	42.27	51.51	74.13	76.86	70.92	78.70
1/90-12/94	39.70	46.08	31.06	40.21	43.17	48.94	34.01	42.75	14.92	19.08	12.31	24.64
1/95-12/99	15.18	16.88	13.55	14.79	10.41	11.98	9.36	10.27	0.51	0.62	0.42	1.46
1/00-12/04	69.07	70.94	61.85	58.36	62.84	65.13	55.07	50.36	74.19	76.31	71.68	72.99
Ten-Year												
1/65-12/74	12.45	13.70	10.90	23.43	19.68	21.04	17.56	30.25	45.70	46.03	45.11	56.55
1/75-12/84	3.62	3.75	3.13	7.53	6.32	6.37	5.79	10.43	2.97	2.95	3.02	7.75
1/85-12/94	27.65	28.40	26.89	30.31	18.51	18.81	17.89	21.17	37.66	38.39	37.81	45.15
1/95-12/04	23.00	24.51	19.31	21.86	13.51	14.32	11.04	12.50	7.39	7.97	7.23	11.23
Thirty-Year												
1/65-12/94	0.04	0.03	0.07	0.16	0.17	0.17	0.20	0.32	1.90	1.81	1.59	2.07
1/70-12/99	2.58	2.76	2.24	4.35	4.64	4.44	4.13	6.05	3.18	3.00	2.90	3.67
1/75-12/04	5.39	5.66	4.13	7.03	13.19	13.04	12.06	14.62	13.64	14.39	14.06	15.92
More Years												
1/65-12/99	0.38	0.31	0.23	0.76	1.59	1.51	1.20	1.69	8.89	9.07	8.21	9.71
1/70-12/04	2.24	2.56	2.20	3.22	10.40	10.41	9.33	12.14	4.39	5.25	4.90	5.82
1/65-12/04	16.07	18.74	11.63	21.53	10.77	13.20	7.56	16.17	22.99	25.13	20.28	35.70

Note: The tabled finite-sample rejection probabilities are computed by simulation using 10,000 replications of each experiment.

Table VI. Simulated rejection probabilities (%) of nominal 5 percent HAR tests by equation subsets for the NN experiments

Eqs	Conventional Bartlett $M = 6$	Fixed-b: Bartlett $b = 0.10$	Fixed-b: Bartlett $b = \rho = 1$	Fixed-ρ: Parzen $\rho = 32$
		$T = 60$		
1	9.67	6.22	5.82	5.42
1-2	14.18	7.05	5.71	5.23
1-3	20.11	7.96	6.15	6.05
1-4	26.33	8.84	6.97	6.49
1-5	34.07	9.30	6.53	6.52
1-6	42.91	11.05	7.71	6.97
1-7	50.43	11.20	7.98	6.26
1-8	59.00	13.31	9.00	7.94
1-9	68.25	14.70	9.50	7.86
1-10	74.97	15.12	9.37	6.83
		$T = 120$		
1	7.35	5.70	5.52	5.24
1-2	9.08	5.94	5.07	4.61
1-3	11.37	6.28	5.31	4.86
1-4	13.49	6.44	5.62	5.49
1-5	17.10	7.40	5.78	5.66
1-6	21.03	7.84	5.91	5.87
1-7	24.61	7.92	6.08	5.16
1-8	29.41	8.37	6.34	6.34
1-9	34.42	9.00	6.20	6.01
1-10	39.80	9.05	6.20	5.61
		$T = 240$		
1	6.18	5.46	5.57	5.12
1-2	6.90	5.46	4.76	4.57
1-3	7.90	5.72	5.15	4.82
1-4	9.16	5.65	5.44	5.22
1-5	10.31	6.05	5.25	5.66
1-6	11.68	6.09	5.33	5.41
1-7	13.33	6.96	5.84	5.37
1-8	15.32	6.87	5.52	5.72
1-9	17.10	7.07	5.49	5.67
1-10	19.29	7.15	5.85	5.12

Note: The tabled rejection probabilities are computed by simulation using 10,000 replications of each experiment. The fixed-b and fixed-ρ asymptotic critical values are from Tables I and II.

Table VII. Simulated Power (%) of level-corrected 5 percent new HAR tests for the NN experiments

Sub-Period	Fixed-*b*				Fixed-ρ			
	Bartlett Kernel, $b = \rho = 1$				Parzen Kernel, $\rho = 32$, $b = 1$			
Five-Year	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$
1/65-12/69	45.2	96.4	99.9	100.0	32.0	86.9	99.4	100.0
1/70-12/74	35.2	91.6	99.7	100.0	25.3	78.3	98.0	99.9
1/75-12/79	15.2	56.6	90.3	98.6	11.9	40.9	75.5	93.5
1/80-12/84	18.8	64.7	94.1	99.3	14.8	47.9	82.8	96.6
1/85-12/89	59.7	98.9	100.0	100.0	42.7	94.4	99.9	100.0
1/90-12/94	39.6	93.5	99.8	100.0	26.8	81.2	98.5	99.9
1/95-12/99	17.0	58.9	91.1	98.9	12.8	42.4	76.6	94.1
1/00-12/04	10.6	33.2	67.7	90.1	8.2	23.7	49.4	75.0
Ten-Year								
1/65-12/74	70.5	99.7	100.0	100.0	51.5	97.6	100.0	100.0
1/75-12/84	21.4	72.6	96.6	99.7	15.8	53.6	87.7	98.3
1/85-12/94	56.0	98.5	100.0	100.0	37.8	92.5	99.8	100.0
1/95-12/04	20.6	72.2	96.6	99.7	15.6	52.9	86.9	98.1
Thirty-Year								
1/65-12/94	70.8	99.7	100.0	100.0	50.2	97.6	100.0	100.0
1/70-12/99	65.2	99.4	100.0	100.0	45.5	96.0	100.0	100.0
1/75-12/04	51.7	98.1	99.9	100.0	35.6	90.9	99.8	100.0
More Years								
1/65-12/99	77.2	99.8	100.0	100.0	57.0	99.0	100.0	100.0
1/70-12/04	64.3	99.3	100.0	100.0	46.9	96.5	100.0	100.0
1/65-12/04	73.0	99.8	100.0	100.0	52.5	98.2	100.0	100.0

Note: The tabled finite-sample powers are computed by simulation using 10,000 replications of each experiment