The robustness of modified unit root tests in the presence of GARCH

Cook, Steven

Empfohlene Zitierung / Suggested Citation:
Cook, S. (2006). The robustness of modified unit root tests in the presence of GARCH. *Quantitative Finance*, *6*(4), 359-363. https://doi.org/10.1080/14697680600702045

The robustness of modified unit root tests in the presence of GARCH.

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2005-0100.R1
Manuscript Category:	Research Paper
Date Submitted by the Author:	18-Mar-2006
Complete List of Authors:	Cook, S; University of Wales Swansea, Economics
Keywords:	GARCH models, Time Series Economics, Numerical Simulation, Quantitative Finance
JEL Code:	C12 - Hypothesis Testing < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, C15 - Statistical Simulation Methods|Monte Carlo Methods < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

QFrev2.tex

The robustness of modified unit root tests in the presence of GARCH*

Steven Cook†
Department of Economics
University of Wales Swansea

March 18, 2006

ABSTRACT

The research of Kim and Schmidt (1993) is extended to examine the properties of modified Dickey-Fuller unit root tests in the presence of generalised autoregressive conditional heteroskedasticity (GARCH). Using Monte Carlo simulation, the properties of the tests are examined for a range of GARCH processes over alternative sample sizes. Oversizing is observed for all tests, with the extent of size distortion driven by the volatility, rather than the persistence, of the underlying GARCH process. While the original Dickey-Fuller test is found to exhibit greater size distortion than the modified tests, the modified tests are found to be substantially oversized when the GARCH process exhibits a high degree of volatility, even for large samples.

Keywords: GARCH; Unit root tests; Size distortion.

* I am grateful to two anonymous referees for suggestions which have lead to improvements in the content and presentation of the present paper.

† Dr Steven Cook, Department of Economics, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP. Tel: (01792) 602106. E-mail: s.cook@swan.ac.uk.

1 Introduction

The notion that many financial time series are characterised by volatility clustering has a long history in finance theory. To allow formal examination of this phenomenon, Engle (1982) introduced the notion of autoregressive conditional heteroskedasticity (ARCH) which permits modelling of the temporal dependency of conditional variances. This approach was subsequently extended by the independent research of Bollerslev (1986) and Taylor (1986) which proposed the generalised autoregressive conditional heteroskedasticity (GARCH) model. This model and its various modifications are now a central component of empirical finance, receiving widespread application and proving to provide accurate forecasts (see Anderson and Bollerslev 1998). Given the prominence of GARCH in the economics and finance literatures it is unsurprising that Kim and Schmidt (1993) have examined its impact upon unit root testing, a topic which itself constitutes a central component of empirical research. However, while the research of Kim and Schmidt (1993) is to be welcomed due to its obvious significance for the practitioner, their analysis considered the seminal unit root test of Dickey and Fuller (1979) alone. In this paper, the research of Kim and Schmidt (1993) is extended to consider the impact of GARCH upon more recently proposed modified Dickey-Fuller tests. The modified tests considered are those of Park and Fuller (1995), Shin and So (2001), Elliott *et al.* (1996), Leybourne (1995) and Granger and Hallman (1991) which employ weighted symmetric estimation, recursive mean adjustment, local-to-unity detrending, forward and reverse regressions and rank-based estimation respectively. These tests, and their underlying properties, are of obvious importance to the applied researcher as they have been shown to possess either greater power or exhibit greater robustness than the seminal Dickey-Fuller test. These particular modified Dickey-Fuller tests have been selected for analysis as they have recently received much attention in the literature, being the subject of both increased application in empirical research and theoretical analysis (see Leybourne *et al.* 2005).[1] The results of the present paper will allow it to be seen whether the improved properties of the modified tests relative to the original Dickey-Fuller test in the presence of white noise errors are offset when the analysis is extended to consider empirically relevant GARCH error processes. It should be noted that the current study complements a number of areas of recent research. In particular, two alternative themes have emerged in the analysis of unit root testing in the presence of changing variances. First, a body of research has evolved examining the behaviour of unit root tests in the presence of single, abrupt breaks in variance (see, *inter alia*, Hamori and Tokihisa 1999; Cook 2002; Kim *et al.* 2003). Second, a num-

[1] As noted by a referee, the present analysis could be extended further to consider alternative unit root tests and tests of stationarity.

ber authors have advocated an alternative approach to the testing the unit root hypothesis in the presence of GARCH, involving the joint estimation of a unit root testing equation and a GARCH specification. Examples of this latter research are provided by Seo (1999), Boswijk (2001) and Li *et al.* (2002). The current analysis is therefore related to both of the above strands of research. However, despite the similarities, there are also clear differences between the present study and previous research. While the current analysis does consider changing variances, the changes here are driven by GARCH rather than one-off changes as in the first body of research. With regard to the second body of research, the analysis here does not consider the joint estimation of a unit root testing equation along with a GARCH specification, but instead considers the impact of employing modified unit root tests when the possible GARCH behaviour is neglected. This possibility is considered as it reflects standard practice in empirical finance where GARCH is ignored when conducting unit root tests.

This paper proceeds as follows. In section [2] the alternative unit root tests to be examined are presented. Section [3] contains the Monte Carlo experimental design employed and simulation results obtained. Section [4] provides some concluding remarks.

2 Alternative unit root tests

In this section the original Dickey-Fuller test and the five modified tests to be considered are presented.[2]

2.1 The Dickey-Fuller test

Given a series $\{y_t\}_{t=0}^{T}$, the familiar Dickey-Fuller (DF) τ_μ test examines the unit root hypothesis ($\mathrm{H}_0 : \phi < 0$) via the t-ratio of $\widehat{\phi}$ in the following regression:

$$\Delta y_t = \mu + \phi y_{t-1} + \varepsilon_t \tag{1}$$

2.2 The weighted symmetric Dickey-Fuller test

The weighted symmetric DF test of Park and Fuller (1995) results from the application of a double length regression, with the weighted symmetric estimator of the autoregressive parameter, denoted as $\widehat{\rho}_{ws}$, given as the value minimising:

$$Q_{ws}(\rho) = \sum w_t (y_t - \rho y_{t-1})^2 + \sum (1 - w_t)(y_t - \rho y_{t+1})^2 \tag{2}$$

[2] In this paper unit root tests are considered in their 'with intercept' forms, as the recursively mean-adjusted Dickey-Fuller test of Shin and So (2001) and the rank-based Dickey-Fuller test of Granger and Hallman (1991) are available in this form alone.

where $w_t = (t-1)/T$. The unit root hypothesis is then tested using the statistic τ_{ws}:

$$\tau_{ws} = \sigma_{ws}^{-1}\left(\widehat{\rho}_{ws} - 1\right)\left(\sum_{t=2}^{T-1} y_t^2 + T^{-1}\sum_{t=1}^{T} y_t^2\right)^{\frac{1}{2}} \tag{3}$$

where $\sigma_{ws}^2 = (T-2)^{-1} Q_{ws}\left(\widehat{\rho}_{ws}\right)$.

2.3 The recursive mean adjusted Dickey-Fuller test

In recent research, Shin and So (2001) have proposed a recursively mean-adjusted DF test. Shin and So (2001) note that the use of mean-adjusted observations $(y_t - \overline{y})$ in the following regression results in correlation between the regressor $(y_{t-1} - \overline{y})$ and the error (ϵ_t):

$$y_t - \overline{y} = \gamma\left(y_{t-1} - \overline{y}\right) + \epsilon_t \tag{4}$$

The resulting bias of the ordinary least squares estimator $\widehat{\gamma}$ has been calculated by Tanaka (1984) and Shaman and Stine (1988) as:

$$\mathsf{E}\left(\widehat{\gamma} - \gamma\right) = -T^{-1}\left(1 + 3\rho\right) + o\left(T^{-1}\right) \tag{5}$$

Consequently, Shin and So (2001) propose the use of recursively mean-adjusted observations to overcome this correlation, with the recursive mean $(\overline{y}_t)$ calculated as:

$$\overline{y}_t = t^{-1}\sum_{i=1}^{t} y_i \tag{6}$$

The recursively mean-adjusted DF test, denoted as τ_{rec}, is then given as the t-test of $\gamma_0 = 1$ in the following regression:

$$y_t - \overline{y}_{t-1} = \gamma_0\left(y_{t-1} - \overline{y}_{t-1}\right) + \epsilon_t \tag{7}$$

2.4 The GLS detrended Dickey-Fuller test

To increase the power of the τ_μ test, Elliott *et al.* (1996) propose local-to-unity detrending via generalised least squares (GLS) or quasi-differencing. To implement this test, GLS-transformed data are derived as:

$$\begin{aligned} y_{\overline{\alpha}} &= \left[y_1, y_2 - \overline{\alpha}y_1, ..., y_T - \overline{\alpha}y_{T-1}\right]' \\ z_{\overline{\alpha}} &= \left[z_1, z_2 - \overline{\alpha}z_1, ..., z_T - \overline{\alpha}z_{T-1}\right]' \end{aligned}$$

where z_t denotes the deterministic terms considered, $\overline{\alpha} = 1 + \overline{c}T^{-1}$ and $\overline{c}$ is a constant determining the extent of local-to-unity detrending. When an intercept term is employed, as in the present analysis, $z_t = 1$ and $\overline{c} = -7$. The GLS demeaned series y_t^α is then derived as $y_t^\alpha = y_t - \widehat{\beta}_0$ where

$\widehat{\beta}_0$ denotes the coefficient obtained from the regression of $y_{\overline{\alpha}}$ upon $z_{\overline{\alpha}}$. The resulting unit root test, denoted as τ_{gls}, is then given as the t-ratio of $\widehat{\delta}$ in the following regression:

$$\Delta \widetilde{y}_t = \delta \widetilde{y}_{t-1} + \varepsilon_t \tag{8}$$

2.5 The maximum Dickey-Fuller test

A further modification proposed to increase the power of the DF test is provided by the maximum DF test of Leybourne (1995) which requires the joint application of forward and reverse regressions. Given a series of interest $\{y_t\}_{t=0}^{T}$, the DF test of (1) is applied to both $\{y_t\}$ and $\{z_t\}$, where $z_t = y_{T-t}$ for $t = 0, ..., T$. The maximum DF test, denoted here as τ_{max}, is then simply the maximum (less negative) of the two test statistics obtained.

2.6 The rank-based Dickey-Fuller test

The rank-based DF test proposed by Granger and Hallman (1991) simply involves replacing y_t with r_t, where r_t is the rank of y_t in $y_0, ..., y_T$. Testing of the unit root hypothesis is then achieved via examination of the null hypothesis $\mathrm{H}_0 : \beta^* = 0$ in the model below:

$$\Delta r_t = \alpha^* + \beta^* r_{t-1} + \xi_t^* \qquad t = 1, ..., T \tag{9}$$

The above rank-based test is of interest as in addition to possessing greater power than the τ_μ test (see Granger and Hallman 1991, p.219), the use of ranked data might be expected to result in robust inference.

3 Monte Carlo experimentation

3.1 Monte Carlo design

To examine the properties of the above unit root tests in the presence of GARCH(1,1) errors, the following data generation process (DGP) is employed:

$$\begin{aligned} y_t &= y_{t-1} + w_t \qquad t = 1, ..., T & (10) \\ h_t^2 &= \phi_0 + \phi_1 w_{t-1}^2 + \phi_2 h_{t-1}^2 & (11) \\ w_t &= h_t v_t & (12) \\ v_t &\sim \mathrm{N}(0, 1) & (13) \end{aligned}$$

The above DGP therefore closely follows that of Kim and Schmidt (1993). In this paper, neither of the extreme cases of degenerate GARCH ($\phi_0 = 0$) nor integrated GARCH ($\phi_1 + \phi_2 = 1$) are considered. Instead, more realistic GARCH processes are generated for a range of values of

$\{\phi_1, \phi_2\}$ corresponding to near integration, with $\phi_0 = 1 - \phi_1 - \phi_2$ in all cases. The precise values of the volatility parameter (ϕ_1) and the degree of persistence ($\phi_1 + \phi_2$) employed are based upon values observed in empirical research. In particular, a range of values are chosen to reflect the typically differing values of $\{\phi_1, \phi_2\}$ noted at alternative data frequencies (see Drost and Nijman 1993; Engle and Patton 2001). In addition to considering GARCH parameter values reported in previous empirical research, the selected values of ϕ_1 and ϕ_2 are informed by analysis of current data. For example, while examination of quarterly data on UK macroeconomic variables (GDP, imports, exports and consumers' expenditure) produced smaller values for ϕ_1 in the range 0.07 to 0.15, analysis of long term and short interest rates for the US produced intermediate to higher values for ϕ_1 in the range 0.18 to 0.37. For both the macroeconomic and interest rate series, near integration ($\phi_1 + \phi_2 \simeq 1$) is noted.[3] The initial value of the conditional variance is set equal to one ($h_0 = 1$) without loss of generality, while the initial value of y_t is set to zero ($y_0 = 0$). The innovation series $\{v_t\}$ is generated using pseudo *i.i.d.* $\mathsf{N}(0, 1)$ random numbers from the RNDNS procedure in the GAUSS, with the same random numbers used for the different experimental designs to allow comparison. All experiments are performed over 25,000 simulations with four sample sizes considered: $T = \{100, 250, 500, 1000\}$. To observe the empirical sizes, rejection frequencies for the alternative tests are calculated at the 5% level of significance.[4]

3.2 Monte Carlo experimentation

The results of the Monte Carlo simulation are reported in Table One. From inspection of the results, a number of features are apparent. It is clear that for all tests it is the degree of volatility of the GARCH process, as measured by ϕ_1, rather than the degree of persistence, as measured by $\phi_1 + \phi_2$, which has the greater impact upon the sizes of the tests. For example, for $\{\phi_1, \phi_2, T\} = \{0.32, 0.66, 100\}$, the τ_μ test has an empirical size of 12.97% at the 5% nominal level of significance. This represents size distortion in excess of 250%. In contrast, consider the alternative design $\{\phi_1, \phi_2, T\} = \{0.04, 0.95, 100\}$. While this design employs the same sample size it represents a substantially reduced degree of volatility but slightly increased persistence. The resulting size of the τ_μ test for this design is a near nominal 5.63%. It can also be seen that for the τ_μ test, size distortion decreases slightly as the sample size is increased. However, this is not true for the modified tests as they all exhibit some evidence of a greater degree of size distortion for larger samples, although this pattern is not consistent across all increases in sample size. The presence

[3] Further information on these empirical results is available upon request.

[4] Critical values for the alternative tests are drawn from the seminal studies wherever possible. For instances where critical values are not available for a given test at a specific sample size, these are derived using Monte Carlo experimentation using a standard data generation process with normally distributed *i.i.d.* errors. Details are available upon request.

of size distortion in larger samples, as noted by Kim and Schmidt (1993), may result from the accumulation of the GARCH variance under the unit root hypothesis and is an issue which itself warrants further research. However, the most apparent feature of the simulation results overall is the robustness of the modified tests relative to the seminal τ_μ test for the smaller sample sizes considered. However, it is also apparent that as the sample size is increased the difference in oversizing between the original and modified tests is reduced. To illustrate this, consider the results for higher degrees of volatility where size distortion is most apparent. While the modified tests are clearly more robust than the τ_μ for smaller samples of 100 and 250 observations, the difference in empirical size is reduced when the larger samples are observed. For example, for the design $\{\phi_1, \phi_2, T\} = \{0.32, 0.66, 1000\}$ the tests can be ranked in descending order of oversizing as follows: τ_μ, τ_{rec}, τ_{max}, τ_{ws}, τ_{rank}, τ_{gls}, with empirical sizes 10.11%, 9.30%, 9.12%, 9.11%, 8.92% and 7.54% respectively. The difference between the original DF test and the τ_{rec}, τ_{max}, τ_{ws} tests in particular is small, with the observed rejection frequencies representing substantial oversizing for a relatively large sample size.

TABLE ONE ABOUT HERE

4 Conclusion

In this paper the research of Kim and Schmidt (1993) has been extended to examine the properties of modified Dickey-Fuller unit root tests in the presence of GARCH. Given the importance and prevalence of GARCH and unit root testing in the economics and finance literatures, their interaction clearly warrants close attention. Using Monte Carlo simulation the properties of a number of popular, modified Dickey-Fuller tests have been examined in the presence of a range of GARCH processes over alternative sample sizes. In general it was found that the oversizing of all tests was driven by the volatility rather than the persistence of the underlying GARCH process, with the original Dickey-Fuller test exhibiting greater size distortion than the modified tests. Interestingly, while a variation in the oversizing of the modified tests was noted, it was apparent that as the sample size was increased the difference between the oversizing of the seminal Dickey-Fuller test and the modified tests was reduced. Indeed, for large values of the volatility parameter of the GARCH model, modified unit root tests were found to exhibit substantial oversizing (the empirical size being approximately double the nominal size) even for a large sample of observations ($T = 1000$). The results therefore indicate that in the presence of GARCH, the original and modified Dickey-Fuller tests may spuriously reject the unit null hypothesis.

References

[1] Bollerslev, T. (1986) 'Generalised autoregressive conditional heteroscedasticity', *Journal of Econometrics*, **31**, 307-327.

[2] Boswijk, P. (2001) 'Testing for unit roots with near-integrated volatility', Tinbergen Institute Discussion Papers 01-077/4, Tinbergen Institute.

[3] Cook, S. (2002) 'Unit root testing in the presence of innovation variance breaks: A simple solution with increased power', *Journal of Applied Mathematics*, **5**, 233-240.

[4] Dickey, D. and Fuller, W. (1979) 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**, 427-431.

[5] Drost, F. and Nijman, T. (1993) 'Temporal aggregation of GARCH processes', *Econometrica*, **61**, 909-927.

[6] Elliott, G., Rothenberg, T. and Stock, J. (1996) 'Efficient tests for an autoregressive unit root', *Econometrica* **64**, 813-836.

[7] Engle, R. (1982) 'Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation', *Econometrica*, **50**, 987-1008.

[8] Engle, R. and Patton, A. (2001) 'What makes a good volatility model?', *Quantitative Finance*, **1**, 237-245.

[9] Granger, C. and Hallman, J. (1991) 'Nonlinear transformations of integrated time series', *Journal of Time Series Analysis*, **12**, 207-218.

[10] Hamori, S. and Tokihisa, A. (1999) 'Testing for a unit root in the presence of a variance shift', *Economics Letters*, **57**, 245-253.

[11] Kim, K. and Schmidt, P. (1993) 'Unit root tests with conditional heteroskedasticity' *Journal of Econometrics*, **59**, 287-300.

[12] Kim, T., Leybourne, S. and Newbold, P. (2003) 'Unit root tests with a break in variance', *Journal of Econometrics*, **109**, 365-387.

[13] Leybourne, S. (1995) 'Testing for unit roots using forward and reverse Dickey-Fuller regressions', *Oxford Bulletin of Economics and Statistics* **57**, 559-571.

[14] Leybourne, S., Kim, T. and Newbold, P. (2005) 'Examination of some more powerful modifications of the Dickey-Fuller test', *Journal of Time Series Analysis*, **26**, 355-369.

[15] Li, W., Ling, S. and McAleer, M. (2002) 'Recent theoretical results for time series models with GARCH error', *Journal of Economic Surveys*, **16**, 245-269.

[16] Pantula, S., Gonzalez-Farias, G. and Fuller, W. (1994) 'A comparison of unit root test criteria', *Journal of Business and Economic Statistics* **12**, 449-459.

[17] Park, H. and W. Fuller (1995) 'Alternative estimators and unit root tests for the autoregressive process', *Journal of Time Series Analysis* **16**, 415-429.

[18] Seo, B. (1999) 'Distribution theory for unit root tests with conditional heteroskedasticity', *Journal of Econometrics*, **91**, 113-144.

[19] Shaman, P. and Stine, R. (1988) 'The bias of autoregressive coefficient estimators', *Journal of the American Statistical Association* **83**, 842–848.

[20] Shin, D. and So, B. (2001) 'Recursive mean adjustment for unit root tests', *Journal of Time Series Analysis* **22**, 595–612.

[21] Tanaka, K. (1984) 'An asymptotic expansion associated with maximum likelihood estimators in ARMA models', *Journal of the Royal Statistical Society* **B46**, 58–67.

[22] Taylor, S. (1986) *Modelling Financial Time Series*, New York: Wiley.

Table One: Empirical test sizes in the presence of GARCH$(1,1)$ errors

	(ϕ_1,ϕ_2)						
	$(0.32,0.66)$	$(0.26,0.72)$	$(0.20,0.75)$	$(0.16,0.80)$	$(0.12,0.85)$	$(0.08,0.90)$	$(0.04,0.95)$
				$T=100$			
τ_μ	12.97	11.54	7.95	7.57	7.02	6.39	5.63
τ_{ws}	8.04	7.47	6.47	6.14	5.82	5.46	5.16
τ_{rec}	7.48	6.96	6.60	6.14	5.76	5.47	5.22
τ_{gls}	7.46	6.98	6.43	6.11	5.74	5.46	5.15
τ_{max}	7.82	7.11	6.25	5.92	5.57	5.18	5.06
τ_{rank}	8.93	8.13	6.67	6.44	6.03	5.77	5.30
				$T=250$			
τ_μ	12.19	10.89	7.29	7.07	6.73	6.25	5.68
τ_{ws}	8.65	8.15	6.64	6.36	6.02	5.64	5.28
τ_{rec}	8.93	8.36	7.10	6.81	6.38	5.98	5.40
τ_{gls}	7.48	7.04	6.21	5.97	5.85	5.54	5.20
τ_{max}	8.99	8.38	6.89	6.60	6.24	5.82	5.40
τ_{rank}	8.95	8.46	6.48	6.31	6.08	5.76	5.20
				$T=500$			
τ_μ	11.33	10.44	7.05	6.96	6.61	6.24	5.82
τ_{ws}	8.65	8.07	6.40	6.22	6.07	5.77	5.45
τ_{rec}	9.22	8.74	6.83	6.56	6.40	6.07	5.66
τ_{gls}	7.24	7.04	5.82	5.71	5.65	5.40	5.12
τ_{max}	8.94	8.39	6.54	6.40	6.13	5.97	5.55
τ_{rank}	9.15	8.81	6.65	6.52	6.28	5.99	5.57
				$T=1000$			
τ_μ	10.11	9.38	6.42	6.33	6.15	5.94	5.52
τ_{ws}	9.11	8.64	6.40	6.41	6.12	5.96	5.58
τ_{rec}	9.30	8.75	6.52	6.37	6.20	6.01	5.62
τ_{gls}	7.54	7.20	5.90	5.86	5.77	5.63	5.31
τ_{max}	9.12	8.67	6.45	6.37	6.06	5.80	5.55
τ_{rank}	8.92	8.68	6.39	6.35	6.28	6.01	5.56

Notes: The reported results represent empirical rejection frequencies of the unit root hypothesis, measured in percentage terms, for the alternative tests calculated using the DGP of (10)-(13) calculated over 25,000 simulations at the 5% nominal level of significance.