EUROPEAN
UNIVERSITY AT
SAINT-PETERSBURG

Wojciech Charemza

Mikhail Lifshits

Svetlana Makarova

A simple test for unit root bilinearity

Working paper Ec-01/02

Department of Economics

St. Petersburg
2002

WOJCIECH CHAREMZA*, MIKHAIL LIFSHITS** and SVETLANA MAKAROVA***. A simple test for unit root bilinearity. – **European University at St.-Petersburg, Department of Economics. Working paper Ec-01/02.**

29 March, 2002

Abstract:

The paper introduces a *t*-ratio type test for detecting bilinearity in a stochastic unit root process. It appears that such process is a realistic approximation for many economic and financial time series. It is shown that, under the null of no bilinearity, the tests statistics are asymptotically normally distributed. Proofs of this asymptotic normality requires the Gihman and Skorohod theory for multivariate diffusion processes. Finite sample results describe speed of convergence, power of the tests and possible distortions to unit root testing which might appear due to the presence of bilinearity. It is concluded that the two-step testing procedure suggested here (the first step for the linear unit root and the second step for its bilinearity) is consistent in the sense that the size of step one test is not affected by the possible detection of bilinearity at step two.

Keywords: Time series econometrics, testing, nonstationary bilinear processes.

JEL classification numbers: C12, C22, G15.

European University at St. Petersburg

3 Gagarinskaia st., St. Petersburg, 191187 RUSSIA

e-mail: econ@eu.spb.ru

---

* University of Leicester, UK.

** St.Petersburg State University, St.Petersburg, Russia.

*** European University at St.Petersburg , St. Petersburg, Russia.

A Simple Test for Unit Root Bilinearity *)

Wojciech W. Charemza, Mikhail Lifshits and Svetlana Makarova

1. Introduction

Modelling of economic time series with the use of stochastic bilinear processes seems to be an attractive alternative to the usual linear modelling. Nevertheless, since the seminal Sabba Rao and Gabr (1974) and Granger and Andersen (1978) volumes and occasional further results regarding estimation and statistical inference (see Sabba Rao, 1981, Kim and Basava, 1990, Liu, 1990, Tong, 1990, Grahn, 1995, Brunner and Hess, 1995, Terdik, 1999) little has been done regarding economic applications of bilinear models. A notable exception here is the Peel and Davidson (1998) paper on the bilinear error correction model.

Usually the bilinear process used in economic applications is defined as:

---

*) Financial support of the INTAS99-00472 Project *Nonlinear Structural VAR Modelling of East European Economies* is gratefully acknowledged. We are indebted to Adriana Agapie and Derek Deadman for their comments on an earlier draft of the paper. We are solely responsible for the remaining deficiencies.

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \sum_{j=0}^{r} c_j \varepsilon_{t-j} + \sum_{l_1=1}^{m} \sum_{l_2=1}^{k} b_{l_1 l_2} y_{t-l_1} \varepsilon_{t-l_2} , \tag{1}$$

where ε_t is white noise. In compact notation, it is denoted as a $BL(p, r, m, k)$ process. This is clearly a wide family of processes, possibly too wide for a specific empirical inquiry. Its general nature is a possible explanation for a lack of interest in this type of modelling. So far, inference into this model has concentrated on its stationary case which, for economic implementations, is of a limited use.

In this paper attention is paid to a much narrower, nonstationary class of these processes, the $BL(1, 0, 1, 1)$ process, where $a_1 = 1$ and $c_0 = 1$. Such a process is called herein the unit root bilinear (*URB*) process. It seems to be of a particular interest to economists, since the linear unit root (*URL*) process is a straightforward and testable case of (1) with $BL(1, 0, 1, 1)$ and $b_{11} = 0$.

This paper develops a simple testing procedure in which the existence of the bilinear part in the unit root process can be detected by testing whether $b_{11} > 0$. Section 2 contains the general description of the problem and basics of the testing procedure. It is accompanied by Appendix A, describing the derivation of variance of the first difference of the analysed bilinear process. In Section 3 the main asymptotic results for the *URB* test statistics under the null hypothesis are given. They require utilisation of limit theorems for multivariate diffusion processes, which description is given in Appendix B.

The detailed proofs of limit distributions for test statistics under the null hypothesis of no bilinearity are given in Appendix C. Section 4 analyses the problem of possible size and power distortions related to the fact that the proposed test is conditioned on the validity of the testable hypothesis of the *URL* process. The corresponding finite sample results describing speed of convergence, approximated power of the test and possible distortions to unit root testing which might appear due to the presence of bilinearity and given in Section 5. Concluding remarks and suggestions for future research constitute Section 6.

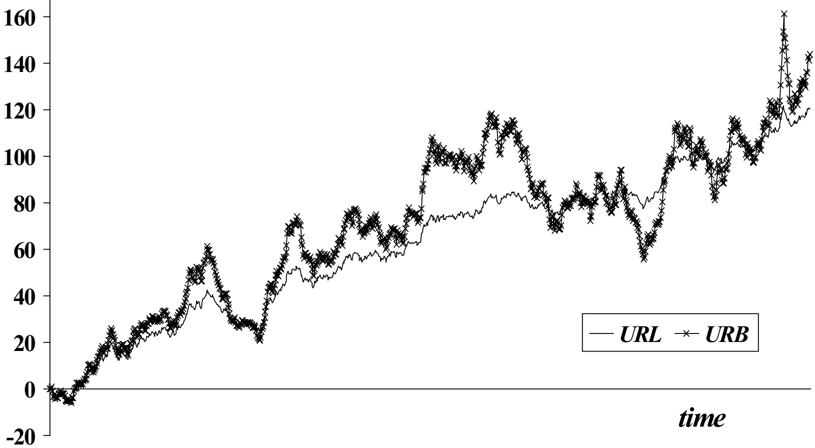
2. The general testing procedure

Consider the following *BL*(1, 0, 1, 1) process:

$$y_t = (a + b\varepsilon_{t-1})y_{t-1} + \varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim IID\,N(0\,,\sigma^2)$, t = 1, 2, … , T. For a = 1 this process is similar to the stochastic unit root (*STUR*) process introduced by Granger and Swanson (1997). The main difference between the *URB* and *STUR* processes is that the latter depends autoregressively on its own lagged values, while the *URB* explicitly relates the unit root dynamics to the lagged innovations. It seems that the *URB* formulation is more realistic. In fact some direct support for such a specification can be found in the economic literature related to speculative behaviour (see e.g. Diba and Grossman, 1988, Ikeda and Shibata, 1992). An illustration of differences between typical runs of the *URL* and *URB* processes is given by Figure 1, where simulated series of (2) with $a = 1$ for $T = 1{,}000$ are presented for $b = 0$ (that is, for the *URL* process) and for $b = 0.025$ (the *URB* process). It appears that the *URB* process, with its clearly visible periods of ups and downs, resembles a typical macroeconomic or financial time series better, than the *URL* process.

Figure 1: A comparison of simulated *URL* and *URB* processes



The stationarity condition for (2) is $a^2 + b^2\sigma^2 < 1$ (see e.g. Granger and Anderson, 1978). Evidently, if a = 1 and b = 0, the process (2) becomes a random walk. Since it is common for economic and financial time series exhibit a unit root, we concentrate on testing whether $b \neq 0$, assuming $a = 1$. A straightforward reparametrisation of (2) in this case is:

$$\Delta y_t = by_{t-1}\varepsilon_{t-1} + \varepsilon_t, \tag{3}$$

where Δ is the first difference operator. It can be noticed that, for $\varepsilon_0 = y_0 = 0$, $E(\Delta y_t) = b\sigma^2$ and $E(y_t) = b\sigma^2(t-1)$, which suggests that, for most economic and financial time series, $b \geq 0$. It is shown in Appendix A that the variance of Δy_t is:

$$Var(\Delta y_t) = \left(5\sigma^2 + b^2\gamma_\varepsilon\right)\left(1 + b^2\sigma^2\right)^{t-2} - 4tb^2\sigma^4 + 7b^2\sigma^4 - 4\sigma^2, \tag{4}$$

where $\gamma_\varepsilon = E\varepsilon^4$. It confirms that, unlike the *URL* process, the *ULB* process is not stationary in first differences.

In this paper we consider the problem of testing the null b = 0 given a = 1, using equation of the type (3) or similar against the alternative of $b > 0$ (further in Section 4 the consequences of the fact that the hypothesis a = 1 also has to be tested are discussed). Given ε_{t-1}, and for a = 1 the Student-*t* test (*t*-ratio) based on the ordinary least squares estimation of b in (2) using (3) can be derived. However, such formulae are not operational for $b \neq 0$, since in fact the term ε_{t-1} is not directly observed. Nevertheless, for small b's it can be noted that in (3) $\Delta y_t \approx \varepsilon_t$ and hence ε_{t-1} can be replaced by Δy_{t-1}[1]. It leads to the following statistic:

$$t_{\hat{b}} = \frac{\sum_{t=2}^{T} y_{t-1}\Delta y_{t-1}\Delta y_t}{\hat{\sigma} \cdot \sqrt{\sum_{t=2}^{T} y_{t-1}^2 \Delta y_{t-1}^2}} \ , \tag{5}$$

where $\hat{\sigma}$ is a consistent estimator of σ. In fact the statistic (5) is the Student-t statistic for $\hat{b}$ in the regression equation:

$$\Delta y_t = \hat{b} y_{t-1} \Delta y_{t-1} + e_t \ , \tag{6}$$

where e_t are the regression residuals and $\hat{\sigma}$ can be estimated using e_t. An analogous statistic can be formulated for a regression containing an intercept:

$$\Delta y_t = const + \hat{b} y_{t-1} \Delta y_{t-1} + e_t \ . \tag{7}$$

If the *URB* process contains a drift (intercept) μ, that is:

$$y_t = \mu + (1 + b\varepsilon_{t-1}) y_{t-1} + \varepsilon_t \ , \tag{8}$$

then the direct replacement of ε_{t-1} by Δy_{t-1} does not make sense, since for $b = 0$, $\Delta y_t = \mu + \varepsilon_t$. One of the possibilities here seems to be to

---

[1]It can be shown that for positive $b < 1/\sqrt{T}$ the ordinary least squares estimator of b and relevant statistics converge to well-defined random variables. See Charemza, Lifshits and Makarova (2002).

demean Δy_t that is, to replace Δy_t by $\Delta z_t = \Delta y_t - mean(\Delta y_t)$. Hence, the corresponding formulae becomes:

$$t_{\hat{b}} = \frac{(T-1)\sum_{t=2}^{T} y_{t-1}\Delta z_{t-1}\Delta y_t - \left(\sum_{t=2}^{T} y_{t-1}\Delta z_{t-1}\right)\left(\sum_{t=2}^{T}\Delta y_t\right)}{\hat{\sigma}\sqrt{T-1}\sqrt{(T-1)\sum_{t=2}^{T} y_{t-1}^2\Delta z_{t-1}^2 - \left(\sum_{t=2}^{T} y_{t-1}\Delta z_{t-1}\right)^2}} \quad . \tag{9}$$

The statistic (9) is in fact a Student-*t* statistic for $\hat{b}$ in the following regression:

$$\Delta y_t = const + \hat{b} y_{t-1}\Delta z_{t-1} + e_t \ . \tag{10}$$

Also, $\hat{\sigma}$ can be computed from the regression (10).

3. Asymptotic properties of the *URB* test statistics

In this section we present the limit properties of the statistics suggested in Section 2 under the null hypothesis that $b = 0$. In particular, we discuss two data generating processes (DGP's) and three test statistics:

DGP 1: The data generating process is:

$$y_t = y_{t-1} + \varepsilon_t \ , \tag{11}$$

where $\varepsilon_t \sim IID\, N(0\,,\sigma^2)$, $t = 1, 2, \dots, T$.

DGP 2: The data generating process is:

$$y_t = \mu + y_{t-1} + \varepsilon_t \ . \tag{12}$$

The test statistics are based on the regression without a constant, (6), with a constant, (7), and on the regression with a constant on the demeaned differences (10). These tests are denoted respectively as Test 1, Test 2 and Test 3. Limit distributions for DGP 1 and Tests 1, 2 and 3 are given by the following theorem:

Theorem 1. *Let the series* y_t *be generated by* (11). *For the regression models, either* (6) (Test 1), (7) (Test 2) or (10) (Test 3) and *under the null of b = 0, as* $T \to \infty$*:*

$$t_{\hat{b}} \Rightarrow \frac{\int_0^1 W_1(t)dW_2(t)}{\sqrt{\int_0^1 [W_1(t)]^2 \, dt}} \sim N(0,1) \quad , \tag{13}$$

where $\Rightarrow$ denotes weak convergence and W_1, W_2 are independent Wiener processes.

For DGP 2 and Tests 2 and 3, the following theorem holds:

Theorem 2. *Let the series* y_t *be generated by* (12). *For the regression model* (7) (Test 2), *and the regression model* (10) (Test 3), *under the null of b = 0, as* $T \to \infty$*:*

$$t_{\hat{b}} \Rightarrow \sqrt{3}\int_0^1 t dW(t) \sim N\left(0,1\right) \ . \tag{14}$$

As it is shown later, proofs of these theorems require convergence results for higher moments and nonlinear functions of random processes in (5) and (9). The usual lemmas of the univariate functional central limit theorem (the Donsker's theorem and its extensions, see e.g. Davidson, 1995, pp. 450-455 and Maddala and Kim, 1998, pp. 54-61) are not sufficient here. Therefore, it is necessarily to apply the more general Gihman and Skorohod (1979, pp. 200-208) multivariate limit theorem for diffusion processes. Its adaptation for discrete scheme and relevant lemmas and examples are described in Appendix B. Detailed proofs of Theorem 1 and Theorem 2 are given in Appendix C. They are based, among others, on a following:

<u>Statement</u>. *Let* $(\varepsilon_k)_{k\geq 1}$ *will be IID sequence with zero odd moments and variance* σ^2*, and the related random walk is defined by* $y_0 = 0$ *and* $y_k = \sum_{j=1}^{k} \varepsilon_j$, $k \geq 1$. *Then:*

1. $\dfrac{1}{\sigma^2 T}\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t \Rightarrow \int_0^1 W_1(t)dW_2(t)$, *where* W_1, W_2 *are independent Wiener processes;*

2. $\dfrac{1}{\sigma T}\sum_{t=2}^{T} y_{t-1}\varepsilon_t^3 \Rightarrow \int_0^1 W_K^{(1)}(t)dW_K^{(2)}(t)$, *where* $W_K^{(1)}, W_K^{(2)}$ *are components of two-dimensional Wiener process with covariance matrix K of the form:*

$$K = \begin{bmatrix} 1 & k_{12} \\ k_{12} & k_{22} \end{bmatrix}, \text{ where } k_{12} = \frac{1}{\sigma} E\varepsilon_t^4 \text{ and } k_{22} = E\varepsilon_t^6;$$

3. $\frac{1}{\sigma T^2} \sum_{t=2}^{T} t y_{t-1} \varepsilon_t^2 \Rightarrow \int_0^1 t W_1^{(1)}(t) dW_3(t)$, *where* $W_1(t)$, $W_3(t)$ *are independent Wiener processes.*

Proofs of the Statement and other results related to the proof of two main theorems, which follow from Gihman - Skorohod approach, are given in Appendix B.

4. Two-step testing

The test presented in previous section is conditional on the existence of unit root in the DGP, that is, it assumes the *URL* process under the null. In practice, however, the existence of a unit root is a testable hypothesis. Generally, the testing procedure consists in applying one or more linear unit root tests (with the *URL* processes at null or alternative hypothesis) at the *first step*, then, provided that the *URL* hypothesis is in some way confirmed, applying the second test, with the *URB* as the alternative, at the *second step*. Clearly, the question arises as to what extent the testing procedure explained above is valid if it is conditional on an earlier result of unit root testing. Generally, in this case one must be careful in distinguishing between the conditional and unconditional probabilities of confirming and rejecting the tested hypotheses.

We consider the case where, at the first step, the null hypothesis is that of a *URL* and the alternative is that of stationarity. Usually, this is a case of the Dickey-Fuller test and its extensions, e.g. that of Leybourne (1995) $DF_{\max}$ test. Let z_1 and z_2 be the statistics used respectively in the first and second step of the conditional testing that is, z_1 is a unit root statistic and z_2 is a bilinearity statistic (5) or (9). Let α be the nominal size of the test (for notational simplicity we are assuming that the nominal size of the test is identical in both steps). Let f_1 and f_2 be the density functions of the statistics z_1 and z_2 . Let us further denote:

$$\int_{z_1 \in \Omega_1^{\alpha}} f_1(z_1 \mid a = 1, b = 0) dz_1 = \alpha \ , \tag{15}$$

where Ω_1^{α} is the critical region for the statistic z_1 at the nominal level of significance α. Hence, α is the probability of making the type I error at step one, conditional on $a = 1$ and $b = 0$, that is, on the random walk assumptions. It can be noted that the probability of making at step one type I error conditional on $a = 1$ and $b > 0$, that is:

$$\int_{z_1 \in \Omega_1^{\alpha}} f_1(z_1 \mid a = 1, b > 0) dz_1 = \alpha_1(b) = \alpha_1 \ , \tag{16}$$

will not usually be equal to *a*. In step two we have:

$$\int_{z_2 \in \Omega_2^{\alpha}} f_2(z_2 \mid z_1 \notin \Omega_1^{\alpha} \ ; a = 1, b = 0) dz_2 = \alpha \ ,$$

where Ω_2^α is the critical region for the statistic z_2 at the nominal level of significance α. Hence, α is also the probability of making type I error at step two conditional on $a = 1$ and $b = 0$ and, additionally, on the non-rejection of the null hypothesis at step one. This will be called here the conditional probability of the type I error. The unconditional probability of the type I error at the nominal significance level α at step two is given by:

$$\int_{z_2 \in \Omega_2^\alpha} f_2(z_2 \,|\, a=1\,, b=0) dz_2 =$$
$$\int_{z_1 \notin \Omega_1^\alpha} f_1(z_1 \,|\, a=1\,, b=0) dz_1 \times \int_{z_2 \in \Omega_2^\alpha} f_2(z_2 \,|\, z_1 \notin \Omega_1^\alpha \;; a=1\,, b=0) dz_2$$
$$= (1-\alpha) \times \int_{z_2 \in \Omega_2^\alpha} f_2(z_2 \,|\, z_1 \notin \Omega_1^\alpha \;; a=1\,, b=0) dz_2 = \alpha_2 = (1-\alpha)\alpha \quad .$$

Next, consider the conditional (on the non-rejection of the null hypothesis at step one) power of the step two test, M_2^c:

$$M_2^c = \int_{z_2 \in \Omega_2^\alpha} f_2(z_2 \,|\, z_1 \notin \Omega_1^\alpha \;; a=1\,, b>0) dz_2 \;, \tag{17}$$

and the nominal unconditional and true unconditional power, denoted respectively as M_2^n and M_2^T. They are given as:

$$M_2^n = \iint_{(z_1, z_2) \in \bar{\Omega}_1^\alpha \times \Omega_2^\alpha} f_{12}(z_1, z_2 \,|\, a=1\,, b=0\,; a=1\,, b>0)\, dz_1 dz_2$$
$$= (1-\alpha) \int_{z_2 \in \Omega_2^\alpha} f_2(z_2 \,|\, z_1 \notin \Omega_1^\alpha \;; a=1\,, b>0) dz_2 = (1-\alpha) M_2^c$$

and:

$$M_2^T = \iint_{(z_1,z_2)\in\bar{\Omega}_1^\alpha\times\Omega_2^\alpha} f_{12}(z_1,z_2 \mid a=1, b>0)\, dz_1 dz_2$$

$$= \int_{z_1\notin\Omega_1^\alpha} f_1(z_1 \mid a=1\,, b>0) dz_1 \times \int_{z_2\in\Omega_2^\alpha} f_2(z_2 \mid a=1\,, b>0) dz_2$$

$$= (1-\alpha_1) \int_{z_2\in\Omega_2^\alpha} f_2(z_2 \mid z_1 \notin \Omega_1^\alpha\,;\, a=1\,, b>0) dz_2 = (1-\alpha_1) M_2^c\ ,$$

where f_{12} is a joint density function and $\bar{\Omega}_1^\alpha$ is a complement of Ω_1^α. It is clear that:

$$M_2^T = (\alpha-\alpha_1) M_2^c + M_2^n\ .$$

The above relation reveals an important practical problem related to the fact that testing at step one is performed at the nominal significance level *a* while, if in the second step the null hypothesis is rejected, the probability of rejecting the true null hypothesis at step one was in fact α_1. If, however, $\alpha_1 > \alpha$, the true unconditional power is going to decrease relatively to the nominal unconditional power since, at step one, the null hypothesis was rejected too often. In this case, the size of the first step test is distorted, but in such a way that it affects power, and not size, of the second step testing. If, however, $\alpha_1 < \alpha$ then, at step one, the null hypothesis was not rejected often enough and at the second step, the size of the test is distorted.

5. Some finite sample results

The previous section of this paper indicates that the entire testing procedure is dependent on the validity of the *URL* condition under the null of $b = 0$. This might be tested by a battery of well-known unit root tests (for a review of these tests see, for instance, Maddala and Kim, 1998). These tests are usually developed under the null hypothesis of the *URL* and their finite sample distributions are examined (mainly by Monte Carlo simulations) in order to establish the critical values of these tests, that is, defining the sets Ω_1^{α} in (15). This creates a potential problem in application of the unit root bilinearity test, since this is conditional on an appropriate unit root test not rejecting the null. But it was shown in Section 4 of this paper that the true probability of rejecting the true null at step one might not be equal to α if, at step two, the null of no bilinearity is rejected.

In order to evaluate the potential effect of this inequality, a series of Monte Carlo experiments have been conducted, in which the usual unit root tests have been applied for data generated by (2) with $a = 1$ and for various values of the b parameter. For each of 25 different sample sizes ranging from 15 to 1,000 and $b = 0.05, 0.10, 0.15$, there were 50,000 series generated. For each series the Dickey-Fuller and $DF_{\max}$ tests were applied and, for the nominal 5% significance level, the frequencies of the cases where the null hypothesis is rejected at the nominal significance level of 5%, was computed. This is in fact the numerical approximation of α_1, as defined by (16).

Figure 2: Numerical approximations of α_1, Dickey-Fuller and $DF_{\max}$ tests, $b = 0.05$

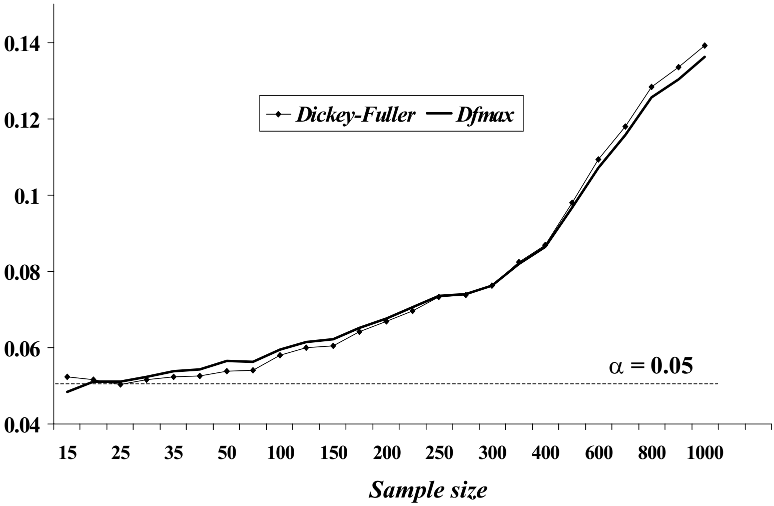


Figure 2 shows the frequencies of the cases where the null hypothesis was rejected at the 5% nominal level of significance in the case where $b = 0.05$ for the Dickey-Fuller and $DF_{\max}$ tests. It indicates a possible inconsistency of the $DF_{\max}$ test for very small sample sizes, since, for $T = 15$, the frequency of rejection is below 5%. However, for all other sample sizes, these frequencies are markedly greater than

5% which suggests that in fact $\alpha_1 > \alpha$ and hence the power, and not size, of the second step testing is affected.

Table 1: *p*-values of the Jarque-Bera statistics for the *URB* test (DGP 1)

Sample size	Test 1	Test 2	Test 3
50	0.0000	0.0000	0.0000
75	0.0053	0.0002	0.0709
100	0.0057	0.0008	0.1486
150	0.2602	0.2579	0.2051
200	0.1822	0.0999	0.5266
300	0.449	0.5891	0.6502
500	0.1466	0.1559	0.9175
700	0.3035	0.3513	0.1360
1000	0.7444	0.7441	0.1617

Although the results given in Section 3 of this paper confirm the convergence of the proposed test to the standard normal distribution, its finite sample properties, and in particular the speed of convergence to normality has to be investigated. Table 1 presents the *p*-values for the Jarque-Bera test for normality, obtained for 50,000 simulated values of the *URB* test statistics (DGP 1, Tests 1, 2 and 3) under the null hypothesis (that is, where $b = 0$). It shows their reasonably quick convergence to normality. In particular, the *URB* statistics became approximately normally distributed for sample sizes of 150 and more.

Table 2: Percentiles of distribution of the *URB* statistics (DGP 1)

T		1%	2.5%	5%	10.0%	90%	95%	97.5%	99%
	Test 1	-2.27	-1.91	-1.60	-1.23	1.23	1.57	1.90	2.28
50	**Test 3**	-2.36	-1.96	-1.64	-1.26	1.26	1.62	1.95	2.34
	Test 3	-2.27	-1.91	-1.59	-1.23	1.23	1.59	1.89	2.28
	Test 1	-2.29	-1.93	-1.61	-1.26	-1.23	1.60	1.91	2.28
100	**Test 2**	-2.34	-1.95	-1.63	-1.27	-1.25	1.62	1.94	2.32
	Test 3	-2.27	-1.93	-1.61	-1.26	1.25	1.61	1.92	2.29
	Test 1	-2.30	-1.93	-1.61	-1.26	1.26	1.63	1.95	2.29
200	**Test 2**	-2.32	-1.94	-1.63	-1.27	1.26	1.64	1.96	2.33
	Test 3	-2.28	-1.92	-1.62	-1.26	1.26	1.63	1.94	2.30
∞		-2.33	-1.96	-1.64	-1.28	1.28	1.64	1.96	2.33

Convergence is even faster for the demeaned test, which might be used for samples as small as 75-150. Table 2 presents percentiles of simulated distributions the test statistics for T = 50, 100 and 200 obtained for 50,000 replications. It confirms that in relatively small samples the deviation of percentiles of the *URB* statistics from the percentiles of the normal distribution is of a magnitude of less than 0.1. It appears that, for most empirical applications, percentiles of the normal distribution can be used as a sufficient approximation of true small sample critical values.

Figures 3, 4 and 5 show empirical frequencies of the rejection of the null hypothesis by the *URB* tests for sample sizes of $T = 50$, 200 and 1,000, where data are generated by (2) for different values of b, ranging from zero to 0.15. For each sample size and for every value of b number of replications of the series is set at 10,000. For $T = 50$, all these b's are 'small', that is at most equal to an inverse of a square root of a sample size. However, for $T = 200$ all b's greater than 0.07 should be regarded as 'large' and so are all b's greater than 0.03 for $T = 1,000$. For $b > 0$, the empirical frequencies of the rejection of the null are the numerical approximations of the step two conditional power M_2^c, as defined by (17).

As it might be expected, the conditional power of the tests does not rise uniformly with the increase in true value of b. Initially, for 'small' b's, the power rises monotonously. It stabilises before reaching its maximum 'small' value and then, for 'large' b's, it is falling fast. It should be noted that for 'small' values of b large enough to be close to the maximal 'small' value', power of the tests using Student-t statistics in Test 1 and Test 2 that is, from the regressions (6) and (7), is close to unity. However, power of the demeaned test (Test 3) applied for data generated by the DGP 1 that is, without an intercept, is lower than that of two other tests. This is quite understandable, since it uses more degrees of freedom in an overspecified model.

Figure 3: Evaluation of $M_2^c(b \mid T = 50)$

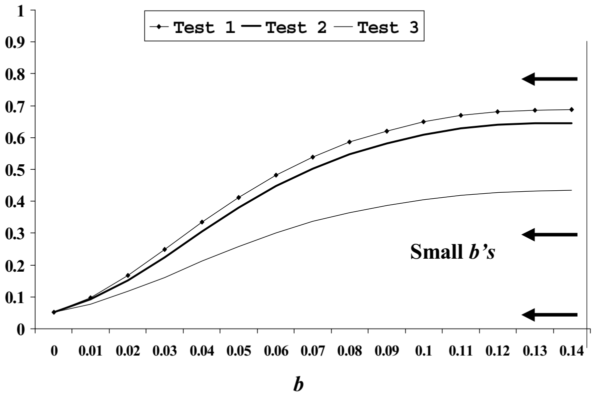


Figure 4: Evaluation of $M_2^c(b \mid T = 200)$

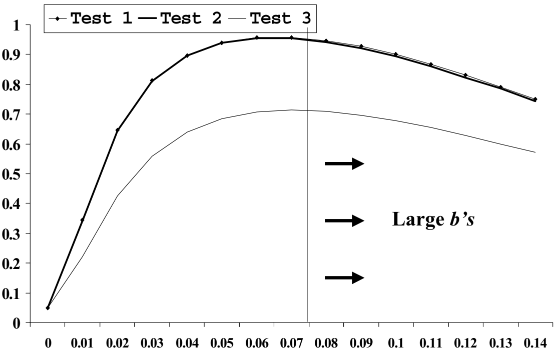
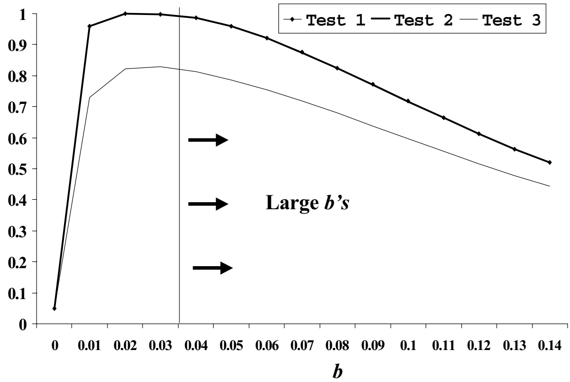


Figure 5: Evaluation of $M_2^c(b \mid T = 1{,}000)$



The finite sample analysis described above might be confusing, since for each b variances of the simulated series of Δy_T differ (see (4)) making direct comparison impossible. In another words, Figures 3, 4 and 5 show $M_2^c(b \mid T)$ for $b > 0$. Clearly, for different T, different variances of Δy_T might affect the probabilities of rejecting the null hypothesis. In order to evaluate potential effects of $Var(\Delta y_T)$ on M_2^c (conditional) and M_2^n (unconditional) powers, they are represented for different T but such that $Var(\Delta y_T) = const$.

Consequently, Figures 6 and 7 show the simulated (empirical) unconditional and conditional power of the *URL* Test 1 and $Var(\Delta y_T)$ = 1.05, 1.10 and 1.15 respectively (results for the Test 2 and Test 3 are nearly identical to these presented here) obtained for small b's (b = 0.006,…, 0.02) and 10,000 replications of (2) for each b. In order to keep variance of Δy_T constant, sample size varies from 207 (for b = 0.020) to 2,292 (for b = 0.06).

Figure 6: Evaluation of $M_2^n(b,T \mid Var(\Delta y_T = const)$, Test 1

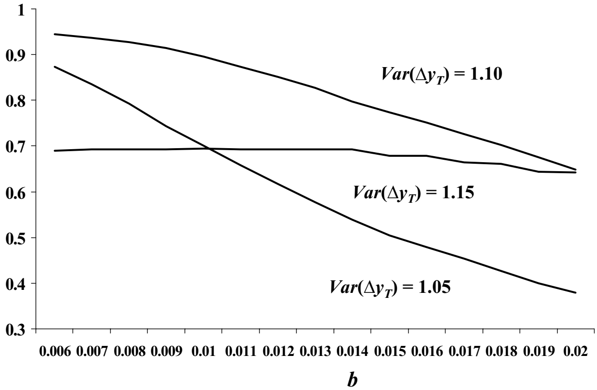
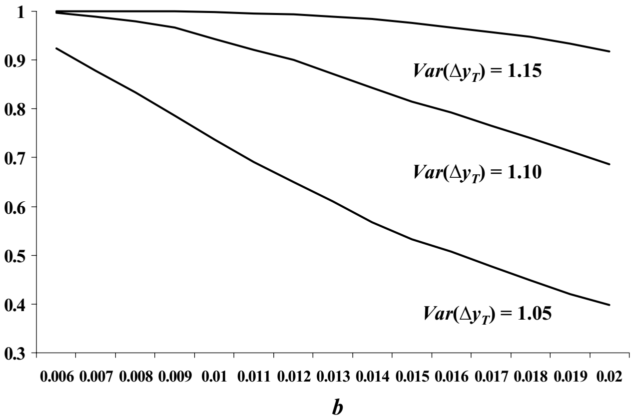


Figure 7: Evaluation of $M_2^c(b,T \mid Var(\Delta y_T = const)$, Test 1



It can be noticed that the conditional power of the test increases considerably with the increase in variance of Δy_T and decreases with the increase in b. This is, however, not the case for the unconditional power. With the increase in variance of Δy_T the unconditional power of the *URB* test becomes more invariant for b. For practitioners such invariance is encouraging and suggests the rationale of using the *URB* test in series which exhibit relatively large variability in first differences.

6. Conclusions

It appears that the proposed concept of unit root bilinearity and testing procedures might be applied in various areas of empirical macroeconomics. The concept of stochastic unit root can substantially enrich the analysis traditionally conducted within the linear unit root framework. In particular, the *URB* tests can be used for detecting speculative bubbles in financial time series, which are widely regarded as being not treatable by traditional linear unit root tests (see e.g. Evans, 1991). It is also possible to consider the bilinear unit root tests as an attractive alternative for unit root structural break tests. The asymptotic and finite sample properties of the tests and especially the asymptotic normality of the test statistics suggest robustness of the testing procedures proposed. The testing itself is simple and can easily be used without a need for developing of specialised software (a collection of procedures written in GAUSS for testing unit root bilinearity in empirical time series is available on request; see Charemza and Makarova, 2002).

It might be conjectured that further development of statistical analysis of economic time series will aim towards the relaxation of assumptions of linear and deterministic nonstationarity in more complicated multivariate models. Consequently, a natural way of future extensions of works presented here is likely to be into a generalisation for vector autoregressive processes and processes with multiple stochastic roots.

Appendix A: Derivation of $Var(\Delta y_t)$

Consider the process:

$$y_t = y_{t-1} + by_{t-1}\varepsilon_{t-1} + \varepsilon_t \quad t = 1, 2, ..., \quad y_0 = \varepsilon_0 = 0. \tag{A1}$$

Multiplying (A1) by ε_t and taking expectation we get:

$$Ey_t\varepsilon_t = \sigma^2, \tag{A2}$$

which implies, that $E\Delta y_t = b\sigma^2$. Moreover, from (A1) and (A2) it is clear that:

$$Ey_t = b\sigma^2(t-1), \quad t = 1, 2, ... \tag{A3}$$

Bearing in mind that $\Delta y_t = by_{t-1}\varepsilon_{t-1} + \varepsilon_t$ and $Cov(y_{t-1}\varepsilon_{t-1}, \varepsilon_t) = 0$ we get:

$$Var(\Delta y_t) = b^2 Var(y_{t-1}\varepsilon_{t-1}) + \sigma^2 = b^2 Ey_{t-1}^2\varepsilon_{t-1}^2 - b^2\sigma^4 + \sigma^2. \tag{A4}$$

From (A4), in order to recover $Var(\Delta y_t)$ it is enough to estimate $Ey_{t-1}^2\varepsilon_{t-1}^2$. Let us denote:

$$\alpha_t = Ey_t^2, \quad \beta_t = Ey_t^2\varepsilon_t^2 . \tag{A5}$$

In order to get expression for α_t, β_t let us derive the vector difference equation for vector $X_t = \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix}$.

Squaring (A1), multiplying the result by ε_t and ε_t^2, taking expectation and, considering (A3), we obtain:

$$\begin{cases} \alpha_t = \alpha_{t-1} + b^2\beta_{t-1} + \sigma_\varepsilon^2 + 4b^2\sigma_\varepsilon^4(t-2) \\ \beta_t = \sigma_\varepsilon^2\alpha_{t-1} + b^2\sigma_\varepsilon^2\beta_{t-1} + \gamma_\varepsilon + 4b^2\sigma_\varepsilon^6(t-2) \end{cases}, \tag{A6}$$

where $E\varepsilon^4 = \gamma_\varepsilon$. From (A6) we may conclude that 2-dimensional vector $X_t = \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix}$ satisfies the following difference equation:

$$X_t = AX_{t-1} + F_t, \qquad X_1 = \begin{pmatrix} \sigma^2 \\ \gamma_\varepsilon \end{pmatrix}, \qquad t = 2,3,\ldots \tag{A7}$$

where: $A = \begin{pmatrix} 1 & b^2 \\ \sigma^2 & b^2\sigma^2 \end{pmatrix}$, and:

$$F_t = 4b^2\sigma^4(t-2)\begin{pmatrix} 1 \\ \sigma^2 \end{pmatrix} + \begin{pmatrix} \sigma^2 \\ \gamma_\varepsilon \end{pmatrix} = 4b^2\sigma^4 t\begin{pmatrix} 1 \\ \sigma^2 \end{pmatrix} + \begin{pmatrix} \sigma_\varepsilon^2 - 8b^2\sigma^4 \\ \gamma_\varepsilon - 8b^2\sigma^6 \end{pmatrix}.$$

Multiplying (A7) by $\begin{pmatrix} b^2 & 1 \\ -1 & \sigma^2 \end{pmatrix}^{-1}$, where $\begin{pmatrix} b^2 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ \sigma^2 \end{pmatrix}$ are eigenvectors of matrix A and denoting

$$Y_t = \begin{pmatrix} b^2 & 1 \\ -1 & \sigma^2 \end{pmatrix}^{-1} X_t, \tag{A8}$$

we obtain the following equation:

$$\begin{aligned} Y_t &= \begin{pmatrix} 0 & 0 \\ 0 & 1 + b^2\sigma^2 \end{pmatrix} Y_{t-1} + G_t \\ Y_1 &= \begin{pmatrix} b^2 & 1 \\ -1 & \sigma^2 \end{pmatrix}^{-1} X_1 = \frac{1}{1 + b^2\sigma^2}\begin{pmatrix} \sigma^4 - \gamma_\varepsilon \\ \sigma^2 + b^2\gamma_\varepsilon \end{pmatrix} \end{aligned}, \tag{A9}$$

where:

$$G_t = \begin{pmatrix} b^2 & 1 \\ -1 & \sigma^2 \end{pmatrix}^{-1} F_t = 4b^2\sigma^4(t-2)\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{1+b^2\sigma^2}\begin{pmatrix} -2\sigma^4 \\ \sigma^2 + 3b^2\sigma^4 \end{pmatrix}. \tag{A10}$$

The solution of difference equation (A9)-(A10) can be written in the form:

$$Y_t = \begin{pmatrix} \dfrac{\sigma^4 - \gamma_\varepsilon}{1+b^2\sigma^2} \\ \dfrac{5\sigma^2 + b^2\gamma_\varepsilon}{b^2\sigma^2}\left(1+b^2\sigma^2\right)^{t-1} - 4t\sigma^2 - \dfrac{5\sigma^2 + b^2\gamma_\varepsilon - 4b^4\sigma^6}{b^2\sigma^2\left(1+b^2\sigma^2\right)} \end{pmatrix}. \tag{A11}$$

Recovering $Var(\Delta y_t)$ from (A4), (A5), (A8) and (A11), finally we obtain equation (4):

$$\begin{aligned} Var(\Delta y_t) &= b^2\beta_{t-1} - b^2\sigma^4 + \sigma^2 \\ &= \left(5\sigma^2 + b^2\gamma_\varepsilon\right)\left(1+b^2\sigma^2\right)^{t-2} - 4tb^2\sigma^4 + 7b^2\sigma^4 - 4\sigma^2. \end{aligned}$$

Appendix B: Gihman-Skorohod techniques in diffusion limit theorems

We describe here a simplified, but also, in some sense, extended version of technique for proving limit theorems for discrete schemes with convergence to a vector-valued diffusion process. The full version is available at Gihman and Skorohod (1979), Chapter 2, Section 3 (pp. 200 – 208, *in Russian edition pp. 267 – 275*).

1. Model and results

Consider a scheme of series of R^m - valued random vectors $X_{T,k}$, $T \in N$, $0 \le k < T$. Let's associate with each series a process:

$$X_T(t) = X_{T,k}, \qquad \frac{k}{T} \le t < \frac{k+1}{T}, \qquad 0 \le k < T \ ,$$

and consider the differences: $\Delta X_{T,k} = X_{T,k+1} - X_{T,k}$, $0 \le k < T$. Now specify the diffusion nature of *X*. Consider a martingale difference series $(\delta_{T,k})$, $T \in N$, $1 \le k \le T$, i.e. $E(\delta_{T,k} \mid X_{T,0},..., X_{T,k}) = 0$, such that

$$Cov(\delta_{T,k} \mid X_{T,0},..., X_{T,k}) = \frac{I_m}{T} \ , \qquad \text{(B1)}$$

where I_m means *m*-dimensional unit matrix. Let a function *h* be defined on $[0,T] \times R^m$, takes values in the space of $m \times m$ matrices and satisfy the conditions:

$\| h(t,x) \| \le C(1 + \| x \|)$, and $\| h(t,x) - h(t,y) \| \le C \| x - y \|$.

Let now $\Delta X_{T,k} = h(X_{T,k})\delta_{T,k}$. Assume also that sequences of martingales $V_T(t) = \sum_{j=1}^{k} \delta_{T,j}$, $\frac{k}{T} \le t < \frac{k+1}{T}$, weakly converge to the Wiener process W.

Under all conditions given above the Gihman Skorohod theorem (see Gihman and Skorohod (1979), p.207, Theorem 12) may be applied and in our particular case it will be obtained that X_T weakly converges to the solution of stochastic differential equation

$$dX(t) = h\left(t, X(t)\right) dW(t). \tag{B2}$$

We also need a minor extension of this result given in Lemma B below.

Lemma B If, instead of (B1), we have:

$$Cov(\delta_{T,k} \mid X_{T,0}, \ldots, X_{T,k}) = \frac{K}{T} , \tag{B3}$$

with some positively defined matrix *K*, then the limit process satisfies the equation

$$dX(t) = h\left(t, X(t)\right) dW_K(t) , \tag{B4}$$

where W_K stands for *m*-dimensional Wiener process with covariance *K*.

Proof of Lemma B. At first, let us choose the coordinate system where *K* has a diagonal form, say, with some diagonal elements $\ell_1^2 \ldots \ell_p^2, 0, \ldots 0$. Let L be the diagonal matrix with diagonal elements $\ell_1 \ldots \ell_p, 0, \ldots 0$. Construct a new martingale difference $\delta_{T,k}^*$ such that $\delta_{T,k} = L\delta_{T,k}^*$ and $Cov(\delta_{T,k}^*) = \frac{I_m}{T}$. Indeed, we can set $\delta_{T,k}^{*(j)} = \delta_{T,k}^{(j)} / \ell_j$ for $1 \le j \le p$ and choose $\delta_{T,k}^{*(j)}$ for $j > p$ to be random variables with zero mean, variances equal to $\frac{1}{T}$, which are mutually independent and independent on $\{\delta_{n,k}\}$. Both required properties are obviously satisfied. Next, let us consider equation $\Delta X_{T,k} = h(\frac{k}{T}, X_{T,k})\delta_{T,k}$. Since:

$$h(\frac{k}{T}, X_{T,k})\delta_{T,k} = h(\frac{k}{T}, X_{T,k})L\delta_{T,k}^* := h^*(\frac{k}{T}, X_{T,k})\delta_{T,k}^*,$$

where $h^*(\cdot,\cdot) := h(\cdot,\cdot)L$, and $Cov(\delta_{T,k}^*) = \frac{I_m}{T}$, our equation corresponds the standard Gihman-Skorohod scheme described above and the processes converge to the solution of equation:

$$\begin{aligned} dX(t) &= h^*(t, X(t))dW(t) = h(t, X(t))LdW(t) \\ &= h(t, X(t))d[LW(t)]. \end{aligned}$$

Finally, notice that $LW(t) = W_K(t)$ where W_K denotes a Wiener process with the covariance *K*. Hence the equation for the limit process is indeed, $dX(t) = h(t, X(t))dW_K(t)$, as required and the proof of Lemma B is completed.

2. Proof of Statement, Section 3.

Below we present three examples which follow from the scheme described above and which give us the proof of Statement from Section 3. These examples will later be used in Appendix C in order to prove the Theorems 1 and 2.

Let $(\varepsilon_k)_{k\geq 1}$ be *IID* sequence with zero odd moments and variance σ^2. We define the related random walk by $y_0 = 0$ and $y_k = \sum_{j=1}^{k} \varepsilon_j$, $k \geq 1$. Consider three Gihman-Skorohod 2-dimensional (*m*=2) constructions and investigate the limit behaviour of vector process $X_T(t) = \left(X_T^{(1)}(t), X_T^{(2)}(t)\right)$. Actually, we are really interested in the limit behaviour of a single random variable $X^{(2)}(1)$. In all examples we set: $X_{T,k}^{(1)} = \dfrac{y_k}{\sigma\sqrt{T}}$ and $\delta_{T,k}^{(1)} = \Delta X_{T,k}^{(1)} = \dfrac{\varepsilon_{k+1}}{\sigma\sqrt{T}}$, and, therefore, they will differ in the constructions of the second component and of matrix *h*.

<u>Example 1</u>. Let $h((x^{(1)}, x^{(2)})) = \begin{bmatrix} 1 & 0 \\ 0 & x^{(1)} \end{bmatrix}$ and $\delta_{T,k}^{(2)} = \dfrac{\varepsilon_k \varepsilon_{k+1}}{\sigma^2 \sqrt{T}}$. The solution of the equation (B2) may be represented in the form:

$$X(t) = \begin{pmatrix} W_1(t) \\ \int_0^t W_1(s)\,dW_2(s) \end{pmatrix},$$

where W_1, W_2 are independent Wiener processes. Hence the limit theorem provides:

$$X_T^{(2)}(1) = \frac{1}{\sigma^3 T}\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t \Rightarrow \int_0^1 W_1(t)dW_2(t). \tag{B5}$$

<u>Example 2</u>. Let $h\left((x^{(1)}, x^{(2)})\right) = \begin{bmatrix} 1 & 0 \\ 0 & x^{(1)} \end{bmatrix}$ and $\delta_{T,k}^{(2)} = \frac{\varepsilon_{k+1}^3}{\sqrt{T}}$. Now the components of $\delta_{T,k}$ are genuinely correlated as in (B3), and the corresponding covariance matrix is:

$$K = \begin{bmatrix} 1 & k_{12} \\ k_{12} & k_{22} \end{bmatrix}, \text{ where } k_{12} = \frac{1}{\sigma}E\varepsilon_t^4 \text{ and } k_{22} = E\varepsilon_t^6.$$

The solution of the equation (B4) may be represented in the form:

$$X(t) = \begin{pmatrix} W_K^{(1)}(t) \\ \int_0^t W_K^{(1)}(s)dW_K^{(2)}(s) \end{pmatrix}.$$

Hence the limit theorem provides:

$$X_T^{(2)}(1) = \frac{1}{\sigma T}\sum_{t=2}^{T} y_{t-1}\varepsilon_t^3 \Rightarrow \int_0^1 W_K^{(1)}(t)dW_K^{(2)}(t), \tag{B6}$$

where $(W_K^{(1)}, W_K^{(2)})$ is two-dimensional Wiener process with covariance *K*.

<u>Example 3</u>.: Let $h\left(t,(x^{(1)}(t),x^{(2)}(t)\right)=\begin{bmatrix}1 & 0\\ 0 & tx^{(1)}\end{bmatrix}$ and $\delta_{T,k}^{(2)}=\dfrac{\varepsilon_{k+1}^2-\sigma^2}{c\sqrt{T}}$,

where $c^2=E(\varepsilon_t^2-\sigma^2)^2=E\varepsilon_t^4-\sigma^4$. We make use of $E\varepsilon_t^3=0$ which provides the orthogonality of $\delta^{(1)}$ and $\delta^{(2)}$. The solution of the equation (B2) may be represented in the form:

$$X(t)=\begin{pmatrix} W_1(t) \\ \int_0^t sW_1(s)dW_3(s)\end{pmatrix},$$

where W_1,W_3 are independent Wiener processes. Hence the limit theorem provides:

$$X_T^{(2)}(1)=\frac{1}{c\sigma T^2}\sum_{t=2}^{T} ty_{t-1}(\varepsilon_t^2-\sigma^2)\Rightarrow\int_0^1 tW_1(t)dW_3(t)\ . \qquad \text{(B7)}$$

These three examples complete the proof of Statement.

Appendix C: Proofs of Theorem 1 and Theorem 2

Before giving the proofs of Theorem 1 and Theorem 2 we state the following auxiliary result.

Lemma C: *Consider IID sequence* $(\varepsilon_k)_{k\geq 1}$ *with zero odd moments and variance* σ^2 *and define the related random walk by* $y_0 = 0$ *and* $y_t = \sum_{j=1}^{t} \varepsilon_j$ *,* $t \geq 1$*. Then:*

1) $\frac{1}{T}\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t \Rightarrow \sigma^3 \int_0^1 W_1(t)dW_2(t)$, *where* W_1, W_2 *are independent Wiener processes;*

2) $X_T^{(2)}(1) = \frac{1}{\sigma T}\sum_{t=2}^{T} y_{t-1}\varepsilon_t^3 \Rightarrow \int_0^1 W_K^{(1)}(t)dW_K^{(2)}(t)$, *where* $(W_K^{(1)}, W_K^{(2)})$ *is two-dimensional Wiener process with covariance K of the form:*

$$K = \begin{bmatrix} 1 & k_{12} \\ k_{12} & k_{22} \end{bmatrix}, \text{ where } k_{12} = \frac{1}{\sigma} E\varepsilon_t^4 \text{ and } k_{22} = E\varepsilon_t^6;$$

3) $\sum_{t=2}^{T} y_{t-1}^2\left(\varepsilon_t^2 - \sigma^2\right) = O\left(T\sqrt{T}\right)$;

4) $\frac{1}{T^2}\sum_{t=1}^{T} y_t^2\varepsilon_t^2 \Rightarrow \sigma^4 \int_0^1 W_1^2(t)dt$;

5) $\frac{1}{T^2\sqrt{T}}\sum_{t=2}^{T} ty_{t-1}\varepsilon_t^2 \Rightarrow \sigma^3\int_0^1 tW_1(t)dt.$

Proof of Lemma C.

1) See (B5), Appendix B.

2) See (B6), Appendix B.

3) Denote $u_t=\varepsilon_t^2-\sigma^2$ and $Var(u_t)=c^2$. Obviously $Eu_t=0$, y_{t-1} and u_t are orthogonal. By using the orthogonality of the sequaence $\{y_{t-1}^2 u_t\}$ we get: $\|\sum_{t=2}^{T} y_{t-1}^2 u_t\|^2=c^2\sum_{t=2}^{T}\| y_{t-1}^2\|^2$. Bearing in mind that $\| y_{t-1}^2\|\le\sigma^2 t$, we obtain:

$$\|\sum_{t=2}^{T} y_{t-1}^2 u_t\|^2=c^2\sum_{t=2}^{T}\| y_{t-1}^2\|^2\le c^2\sigma^4\sum_{t=2}^{T}t^2=O\left(T^3\right).$$

4) Using *2)* and *3)* above we obtain:

$$\begin{aligned}\sum y_t^2\varepsilon_t^2 &= \sum y_{t-1}^2\varepsilon_t^2+2\sum y_{t-1}\varepsilon_t^3+\sum\varepsilon_t^4=\sum y_{t-1}^2\varepsilon_t^2+O(T)\\ &=\sigma^2\sum y_{t-1}^2+\sum y_{t-1}^2(\varepsilon_t^2-\sigma^2)+O(T)\\ &=\sigma^2\sum y_{t-1}^2+O\left(T\sqrt{T}\right).\end{aligned}$$

Hence: $\frac{1}{T^2}\sum_{t=1}^{T} y_t^2\varepsilon_t^2\Rightarrow\sigma^4\int_0^1 W_1^2(t)dt$.

5) $\sum_{t=2}^{T} ty_{t-1}\varepsilon_t^2=\sum_{t=2}^{T} ty_{t-1}\left(\varepsilon_t^2-\sigma^2\right)+\sigma^2\sum_{t=2}^{T} ty_{t-1}=\sum_1+\sum_2$. Applying (B7), Appendix B, we get: $\sum_1=O\left(T^2\right)$. Using the fact that

$\frac{1}{T^2\sqrt{T}}\sum_2 \Rightarrow \sigma^3\int_0^1 tW_1(t)dt$ (see *e.g.* Maddala and Kim 1998), we get statement *5)*. End of proof of Lemma C.

Proof of Theorem 1.

(i). Under the DGP 1, that is, (11), and the null of $b=0$ the *t*-ratio (5) for Test 1 given by (6) becomes:

$$t_{\hat{b}} = \frac{\sum_{t=2}^{T} y_{t-1}\Delta y_{y-1}\Delta y_t}{\hat{\sigma}_\varepsilon \cdot \sqrt{\sum_{t=2}^{T}\left(y_{t-1}\Delta y_{y-1}\right)^2}} = \frac{\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t}{\hat{\sigma}_\varepsilon \cdot \sqrt{\sum_{t=2}^{T} y_{t-1}^2\varepsilon_{t-1}^2}},$$

where $y_0 = 0$, $y_t = \sum_{j=1}^{t}\varepsilon_j$, $t \geq 1$ and $\left(\varepsilon_k\right)_{k\geq 1}$ is IID sequence with zero odd moments and variance σ^2.

Applying Lemma C, *1)* and *4)*, we get (13). *End of proof of (i).*

(ii). Under the null, the *t*–ratio (9) for Test 3 given by (10), becomes:

$$t_{\hat{b}} = \frac{(T-1)\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t - A_1}{\hat{\sigma}_\varepsilon\sqrt{T-1}\sqrt{(T-1)\sum_{t=2}^{T} y_{t-1}^2\varepsilon_{t-1}^2 + B_1}}, \qquad \text{(C1)}$$

where:

$$A_1 = (T-1)\bar{\varepsilon}\sum_{t=2}^{T} y_{t-1}\varepsilon_t + \left(\sum_{t=2}^{T} y_{t-1}(\varepsilon_{t-1}-\bar{\varepsilon})\right)\left(\sum_{t=2}^{T}\varepsilon_t\right),$$

$$B_1 = (T-1)\sum_{t=2}^{T} y_{t-1}^2(-2\bar{\varepsilon}\varepsilon_{t-1} + \bar{\varepsilon}^2) - \left(\sum_{t=2}^{T} y_{t-1}(\varepsilon_{t-1} - \bar{\varepsilon})\right)^2,$$

$\bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t$, and y_t and $(\varepsilon_k)_{k\geq 1}$ are defined as in *(i)*.

Bearing in mind the DGP 1 given by (11) consider expression A_1 in the numerator of (C1). It is clear that, $\bar{\varepsilon}\sum_{t=2}^{T} y_{t-1}\varepsilon_t = O\left(T\sqrt{T}\right)$, $\sum_{y=2}^{T} y_{t-1}(\varepsilon_{t-1} - \bar{\varepsilon}) = O(T)$ and $\sum_{t=2}^{T}\varepsilon_t = O\left(\sqrt{T}\right)$, which implies:

$$A_1 = O\left(T\sqrt{T}\right). \tag{C2}$$

Next, consider expression B_1 in the denominator of (C1). Analogously with A_1 we have:

$$B_1 = O\left(T^2\right). \tag{C3}$$

Applying *1)* and *4)* from Lemma C and substituting (C2) and (C3) into (C1), we get:

$$t_{\hat{b}} = \frac{\frac{T-1}{T^2}\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\varepsilon_t + \frac{1}{T^2}A_1}{\hat{\sigma}\sqrt{\frac{T-1}{T^3}\sum_{t=2}^{T} y_{t-1}^2\varepsilon_{t-1}^2 + \frac{1}{T^3}B_1}} \Rightarrow \frac{\int_0^1 W_1(t)dW_2(t)}{\sqrt{\int_0^1 [W_1(t)]^2\, dt}},$$

which gives (13). For Test 2, the proof is analogous.

End of proof of Theorem 1.

Proof of Theorem 2. Consider DGP 2 given by (12). For Test 3, under the null of $b = 0$ the t- ratio (9) becomes:

$$t_{\hat{b}} = \frac{(T-1)\sum_{t=1}^{T-1} y_t \varepsilon_t \varepsilon_{t+1} + A_2}{\hat{\sigma}_\varepsilon \sqrt{T-1}\sqrt{(T-1)\sum_{t=1}^{T-1} y_t^2 \varepsilon_t^2 + B_2}} \quad , \tag{C4}$$

where:

$$A_2 = -\left(\sum_{t=2}^{T} y_{t-1}\varepsilon_{t-1}\right)\left(\sum_{t=2}^{T}\varepsilon_t\right) - (T-1)\bar{\varepsilon}\sum_{t=2}^{T} y_{t-1}\varepsilon_t + \bar{\varepsilon}\left(\sum_{t=2}^{T} y_{t-1}\right)\left(\sum_{t=2}^{T}\varepsilon_t\right),$$

$$B_2 = (T-1)\sum_{t=2}^{T} y_{t-1}^2\left(-2\bar{\varepsilon}\varepsilon_{t-1} + \bar{\varepsilon}^2\right) - \left(\sum_{t=2}^{T} y_{t-1}\left(\varepsilon_{t-1} - \bar{\varepsilon}\right)\right)^2,$$

and $\bar{\varepsilon}$ is defined as in Theorem 1. Under the null of $b = 0$ and the DGP 2 (12) the process y_t may be written as $y_t = \mu t + S_t$, where $S_t = \sum_{k=1}^{t} \varepsilon_k$ and $\left(\varepsilon_k\right)_{k \geq 1}$ is IID sequence with zero odd moments and variance σ^2. Consider the numerator in (C4). Applying *1)* from Lemma C we may represent the first term as:

$$\begin{aligned}\sum_{t=1}^{T-1} y_t \varepsilon_t \varepsilon_{t+1} &= \mu \sum_{t=1}^{T-1} t \varepsilon_t \varepsilon_{t+1} + \sum_{t=1}^{T-1} S_{t-1}\varepsilon_t \varepsilon_{t+1} + \sum_{t=1}^{T-1} (\varepsilon_t^2 - \sigma^2)\varepsilon_{t+1} + \sigma^2 \sum_{t=1}^{T-1} \varepsilon_{t+1} \\ &= \mu \sum_{t=1}^{T-1} t \varepsilon_t \varepsilon_{t+1} + O\left(T\right).\end{aligned}$$

Hence:

$$\frac{1}{T\sqrt{T}}\sum_{t=1}^{T-1} y_t\varepsilon_t\varepsilon_{t+1} \Rightarrow \mu\sigma^2\int_0^1 t dW_2(t), \tag{C5}$$

Consider the term A_2 in the numerator of (C4). Analogously with the previous derivation we get:

$$A_2 = O\left(T^2\right) . \tag{C6}$$

Next, consider the denominator in (C4). Under DGP 2 (12) we have:

$$\sum_{t=1}^{T-1} y_t^2\varepsilon_t^2 = \mu^2\sum_{t=1}^{T-1} t^2\varepsilon_t^2 + 2\mu\sum_{t=1}^{T-1} S_t\varepsilon_t^2 + \sum_{t=1}^{T-1} S_t^2\varepsilon_t^2$$

$$= \mu^2\sigma^2\sum_{t=1}^{T-1} t^2 + \mu^2\sum_{t=1}^{T-1} t^2\left(\varepsilon_t^2 - \sigma^2\right) + 2\mu\sum_{t=1}^{T-1} tS_{t-1}\varepsilon_t^2 + \sum_{t=1}^{T-1} S_t^2\varepsilon_t^2 .$$

Applying *4)* and *5)* from Lemma C for the later formula we get: $\sum_{t=1}^{T-1} y_t^2\varepsilon_t^2 = \mu^2\sigma^2\sum_{t=1}^{T-1} t^2 + O\left(T^2\sqrt{T}\right)$ and hence:

$$\frac{1}{T^3}\sum_{t=2}^{T} y_{t-1}^2\varepsilon_{t-1}^2 \Rightarrow \frac{\mu^2\sigma^2}{3} . \tag{C7}$$

Analogously with the previous derivations we get that the term B_2 in the denominator of (C4) has the asymptotic:

$$B_2 = O\left(T^3\right) . \tag{C8}$$

Finally, substituting, (C5)-(C8) into (C4), we obtain:

$$t_{\hat{b}} = \frac{\dfrac{T-1}{T^2\sqrt{T}}\sum_{t=2}^{T-1} y_t \varepsilon_1 \varepsilon_{t+1} + \dfrac{1}{T^2\sqrt{T}} A_2}{\hat{\sigma}\sqrt{\dfrac{T-1}{T^4}\sum_{t=1}^{T-1} y_t^2 \varepsilon_t^2 + \dfrac{1}{T^4} B_2}} \Rightarrow \frac{\mu\sigma^2 \int_0^1 t dW_2(t)}{\dfrac{\mu\sigma^2}{\sqrt{3}}} = \sqrt{3}\int_0^1 t dW_2(t),$$

which gives (14). For Tests 1 and 2 the proofs are analogous.

End of proof of Theorem 2.

References

Charemza, W.W. and S. Makarova (2002), 'Blini, version 1.03: Collection of GAUSS procedures for linear and bilinear unit root analysis', University of Leicester, mimeo.

Charemza, W.W., M. Lifshits and S. Makarova (2002), 'Behaviour of statistical aggregates under small bilinearity', St. Petersburg State University, mimeo.

Brunner, A.D. and G.D. Hess (1995), 'Potential problems in estimating bilinear time series models', *Journal of Economic Dynamics and Control* **19**, pp. 663-681.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford.

Diba, B.T. and H.I. Grossman (1988), 'The theory of rational bubbles in stock prices', *Economic Journal* **98**, pp. 746-754.

Evans, G.W. (1991), 'Pitfalls in testing for explosive bubbles in asset pricing', *American Economic Review* **81**, pp. 922-930.

Granger, C.W.J. and A.P. Andersen (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Granger, C.W.J. and N. Swanson (1997), 'An introduction to stochastic unit-root processes', *Journal of Econometrics* **80**, pp. 35-62.

Grahn, T. (1995), 'A conditional least squares approach to bilinear time series estimation', *Journal of Time Series Analysis* **16**, pp. 509-529.

Gihman, I.I. and A.V. Skorohod (1979), *The Theory of Stochastic Processes III*, Springer Verlag, Berlin. Original Russian edition: *Teoria Sluchaynyh Processov*, vol. 3, Nauka, Moskva, 1975.

Ikeda, S. and A. Shibata (1992), 'Fundamentals-dependent bubbles in stock prices', *Journal of Monetary Economics* **30**, pp. 143-168.

Kim, W.K. and I.V. Basava (1990), 'Estimation for the first-order diagonal bilinear time series model', *Journal of Time Series Analysis* **11**, pp. 215-227.

Liu, J. (1990), 'Estimation of some bilinear time series', *Stochastic Models* **6**, pp. 649-665.

Maddala, G.S. and I-M. Kim (1998), *Unit Root, Cointegration and Structural Change*, Cambridge University Press, Cambridge.

Peel, D. and J. Davidson (1998), 'A non-linear error correction mechanism based on the bilinear model', *Economics Letters* **58**, pp. 165-170.

Sabba Rao, T. (1981), 'On the theory of bilinear time series model', *Journal of the Royal Statistics Society* **43**, *No.* 2, pp. 244-255.

Sabba Rao, T. and M.M. Gabr (1974), *An Introduction to Bispectral Analysis and Bilinear Time Series Models*', Series *Lecture Notes in Statistics* vol. 24, Springer, New York.

Terdik, G. (1999), *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis*, Series *Lecture Notes in Statistics* vol. 142, Springer, New York.

Tong, H. (1990), *Non-Linear Time Series. A Dynamical System Approach*, Clarendon Press, Oxford.