A class of stochastic unit-root bilinear processes: mixing properties and unit-root test*

Christian Francq†, Svetlana Makarova ‡
and
Jean-Michel Zakoïan§

Abstract

A class of stochastic unit-root bilinear processes, allowing for GARCH-type effects with asymmetries, is studied. The volatility is not bounded away from zero and is minimum for non zero innovations, which are important differences with the standard GARCH. Necessary and sufficient conditions for the strict and second-order stationarity of the error process are given. The strictly stationary solution is shown to be strongly mixing under mild additional assumptions. It follows that, in this model, the standard (non-stochastic) unit-root tests of Phillips-Perron and Dickey-Fuller are asymptotically valid to detect the presence of a (stochastic) unit-root. The finite sample properties of these tests are studied via Monte Carlo experiments.

JEL classification: C22, C12, C52.

Keywords: Augmented Dickey-Fuller test, Bilinear processes, GARCH, Mixing, Phillips-Perron test, Stochastic unit-roots.

*This work was supported by INTAS: research project number 03-51-3714.

†Université Lille III, GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: christian.francq@univ-lille3.fr

‡European University at St. Petersburg, Russia and the National Bank of Poland. E-mail: makarova@eu.spb.ru

§Corresponding author. GREMARS and CREST, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr

1 Introduction

It is now recognized that many economic series display nonstationarities and nonlinearities. Empirical researchers often find standard linear models, *i.e.* with iid (independent and identically distributed) innovations, inappropriate for differenced series. For this reason, recent papers dealing with unit roots have been concerned with modeling the error term of the linear dynamics as a non-iid process. Results on estimating and testing unit roots with non-iid errors can be found in Phillips (1987), Kim and Schmidt (1993), Seo (1999), Ling and Li (2003), Ling (2004), Rodrigues and Rubia (2005) and the references therein.

Charemza, Lifshits and Makarova (2005) showed that unit-roots models with bilinear errors have interesting economic interpretations, and are empirically relevant. Following this paper, we also consider a unit-root model with bilinear errors, but our specification is different. Our model allows for stationary increments, contrary to the model by Charemza et al (2005). Further differences will be discussed below. It will be seen that our model is closely connected to the class of GARCH processes. Indeed, the solution of this model displays conditional heteroskedasticity, including the so-called leverage effect. Contrary to the standard GARCH, however, our specification does not constrain the coefficients to be positive, which is convenient for statistical purposes. Interestingly, the volatility is not bounded away from 0, and is minimum for non zero innovations.

A natural practice, followed by Charemza et al (2005), is to test for the presence of unit roots in a first step, and then to perform specifications tests on the noise dynamics in a second step. Caution is needed, however, in the blind application of standard unit root tests in the framework of non-iid errors. Rodrigues and Rubia (2005) present numerical experiments showing that non-iid errors may cause severe distortions in conventional unit-root tests. Ling (2004) provided an example of a unit-root model with non-iid errors, namely the so-called double-autoregressive model, in which the LS estimator of the AR coefficient does not converge in law to the standard Dickey-Fuller (DF) distribution. For such models, the most commonly used unit-root tests, *i.e.* the Phillips-Perron and augmented DF tests, may not have the correct asymptotic size.

An important issue for linear models with non-iid errors thus concerns the validity of those unit-root tests. Phillips (1987) and Phillips and Perron (1988) showed that, under moment and mixing conditions on the noise process, the unit-root hypothesis can be tested using the standard DF asymptotic distribution. The main goal of this paper is to establish the validity of those standard unit-root tests for the bilinear model under consideration. This requires analyzing in detail the probability structure of the model, in particular its mixing properties.[1] Apart from the unit-root testing problem, these properties have of course independent interest.

The rest of the paper is organized as follows. The general model is presented in Section 2 and interpretations are given. In Section 3 we study the existence of strictly stationary and second-order stationary solutions. Under a mild additional assumption on the distribution of the iid process, the strictly stationary solution is shown to be strongly mixing in Section

[1] Mixing, which will be defined precisely below, is one way to characterize the decrease of dependence when the variables become sufficiently far apart (see *e.g.* Davidson, 1994).

4. Section 5 is devoted to examining the validity of the Phillips-Perron and augmented DF unit-root tests in our framework. Monte Carlo experiments are presented in Section 6. Concluding remarks are given in Section 7.

2 First-order models

For ease of presentation we only discuss, in this section, a simple sub-class of a more general model considered further. To motivate our specification, we first discuss the properties of a unit-root model with bilinear innovations, which was recently introduced in the literature.

2.1 The model of Charemza et al

Charemza, Lifshits and Makarova (2005) used a bilinear process of the form

$$\Delta y_t := y_t - y_{t-1} = \phi y_{t-1} + u_t, \quad t = 1, 2, \ldots, \tag{1}$$

where

$$u_t = \epsilon_t + b\epsilon_{t-1}y_{t-1}, \quad \epsilon_t \text{ iid } (0, \sigma_\epsilon^2), \quad y_0 = \epsilon_0 = 0. \tag{2}$$

This model has received an economic interpretation as being derived from a model of speculative behavior. In their paper Charemza et al (2005) were mostly concerned by testing the assumption that $b = 0$, giving rise to the so-called "b-test". When $b = 0$ the model has the form of an AR(1), which may ($\phi = 0$) or may not ($\phi \neq 0$) contain a unit-root. When $b \neq 0$ the error term is not stochastically stable (in particular, as demonstrated by the authors, the variance of u_t tends to infinity). Therefore the specification (1)-(2) is not suitable for the so-called integrated of order one $I(1)$ series, which may be found undesirable for many economic series.

Model (1)-(2) can be seen as an element of the class of the so-called stochastic unit-root processes, defined by

$$y_t = \rho_t y_{t-1} + v_t, \quad t = 1, 2, \ldots, \tag{3}$$

where $E\rho_t = 1$ (or more generally $E\varphi(\rho_t) = \varphi(1)$ for some non-trivial function φ). Taking $\rho_t = 1 + \phi + b\epsilon_{t-1}$ and $v_t = \epsilon_t$ yields the model (1)-(2). When $\phi = 0$, the random variable ρ_t has unit expectation justifying the name of stochastic unit-root. Other specifications have been suggested, e.g. by taking $\rho_t = e^{\alpha_t}$ where $E(\alpha_t) = 0$ as in Granger and Swanson (1997). See also Leybourne, McCabe and Tremayne (1996).

2.2 An alternative

Unfortunately, the moment conditions required to apply Phillips and Perron (1988)'s result are not satisfied in the specification (2) because the variance of u_t increases to infinity with t. Instead, a model for (u_t) of the form

$$u_t = \epsilon_t + b\epsilon_t u_{t-1}, \quad \epsilon_t \text{ iid } (0, \sigma_\epsilon^2), \quad y_0 = \epsilon_0 = 0 \tag{4}$$

will be shown to satisfy the conditions of the Phillips-Perron theorem. Model (4) is the first-order version of the class of this paper. When $b = 0$ (but $|b|$ not too large) the model coincide with the one of the previous section. When $b \neq 0$ the error term (u_t) will be shown to be stochastically stable (contrary to the one of the previous section).

It is worth noting that (4) is a bilinear extension of the strong white noise model, obtained for $b = 0$. Bilinear models have been studied by Granger and Andersen (1978), who introduced them in the time series literature, by Subba Rao and Gabr (1984) and by many others. Among the nonlinear models, the bilinear class is one of the most attractive in terms of generality.

Model (1)–(4) can be interpreted as a stochastic unit-root model when $\phi = 0$. Indeed, the representation (3) holds, in which $\rho_t = 1 + \phi + b\epsilon_t$ has mean 1 when $\phi = 0$, and $v_t = \{1 - b(1-\phi)y_{t-2}\}\epsilon_t$ is an error term which is uncorrelated with the y_{t-i} for $i > 0$.

Another interesting feature of Model (4) is its ARCH-type interpretation. The first two conditional moments of u_t are given by

$$E(u_t \mid u_{t-1}, \ldots) = 0, \quad \mathrm{Var}(u_t \mid u_{t-1}, \ldots) = (1 + bu_{t-1})^2\sigma_\epsilon^2.$$

This form of conditional variance is a particular case of the quadratic ARCH model introduced by Sentana (1995). It is seen that the conditional variance is asymmetric: for instance when $b < 0$, a negative shock u_{t-1} increases the conditional variance by a larger amount than a positive shock of the same magnitude. This so-called leverage-effect property is often described as one of the main stylized facts of financial time series (see e.g. Nelson (1991), Zakoïan (1994)). It can be visualized in the so-called news impact curve displayed in Figure 1. Another interesting feature of the model, which is transparent on this figure, is that the volatility is not minimal at zero. In other word, an increase of small positive returns may lower volatility. One can imagine that the volatility is minimal when the returns correspond to the free-risk return ($-1/b$ on the figure). This interpretation, as well as the leverage effect, of course requires $b < 0$. Finally, the volatility is not bounded away from 0, as in the case in the other GARCH models.

In the sequel we consider models of the form (1), and their augmented versions, with bilinear noise specifications including the model in (4).

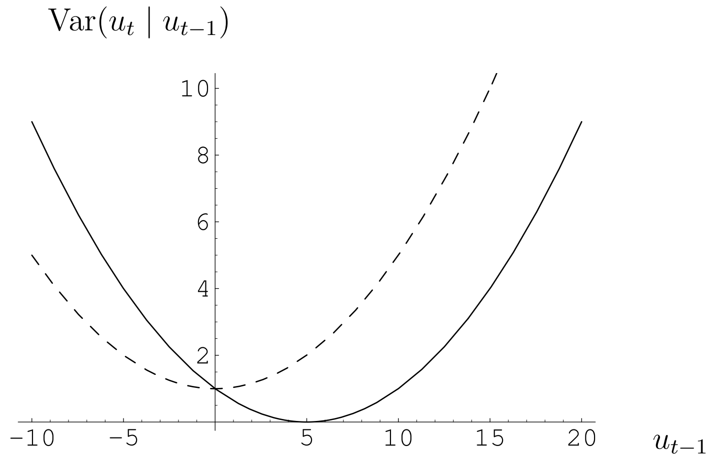

Figure 1: News impact curve of Model (4) with $b = -0.2$ and $\sigma_\epsilon = 1$ (full line) compared with the news impact curve of the ARCH(1) model $u_t = \sqrt{1 + b^2 u_{t-1}^2}\epsilon_t$ (dotted line).

3 A more general specification

We consider the following model

$$\begin{cases} \Delta y_t = \phi y_{t-1} + \psi_1 \Delta y_{t-1} + \cdots + \psi_p \Delta y_{t-p} + u_t, \\ \\ u_t = (1 + b_1 u_{t-1} + \cdots + b_q u_{t-q})\epsilon_t, \qquad \epsilon_t \text{ iid } (0, \sigma_\epsilon^2) \end{cases} \tag{5}$$

where $\phi, \psi_1, \ldots, \psi_p, b_1, \ldots, b_q$ are real coefficients and $\sigma_\epsilon^2 > 0$. The adjunction of higher-order autoregressive terms in the equation of Δy_t can be motivated by the necessity to control for serial correlation, as in the augmented DF test. Similarly, the introduction of several coefficients b_i allows for more persistence in the conditional variance.

As already noted in the case $q = 1$, the process (u_t) is bilinear in the sense it involves products of the variable ϵ_t and u_t past values. However, (u_t) does not strictly speaking belong to the general bilinear class extensively studied in the literature, e.g. by Granger and Andersen (1978).[2] Thus we cannot rely on general results on bilinear processes to study the stationary properties of our class.

The process (u_t) also belongs to the class of Linear ARCH (LARCH), introduced by Robinson (1991) and recently studied by Giraitis, Robinson and Surgailis (2000), Giraitis and Surgailis (2002). The main interest of this class, in which an infinite sequence of coefficients b_j is considered, is to allow for long-memory properties.

3.1 Strict stationarity

We first give a condition for the existence of a strictly stationary white noise solution (u_t). For the reason just mentioned we cannot directly use existing results, e.g. those established by Liu and Brockwell (1988). Moreover, using the approach of Bougerol and Picard (1992a, 1992b), as will be done, gives sharper results.

Let $\underline{u}_t = (u_t, \dots, u_{t-q+1})' \in \mathbb{R}^q$ and $\underline{c}_t = (\epsilon_t, 0, \dots, 0)' \in \mathbb{R}^q$. Then, the second equation in (5) is equivalently written as

$$\underline{u}_t = \underline{c}_t + A_t \underline{u}_{t-1} := \begin{pmatrix} \epsilon_t \\ 0_{q-1} \end{pmatrix} + \begin{pmatrix} \mathbf{b}_{1:q-1}\epsilon_t & b_q \epsilon_t \\ I_{q-1} & 0_{q-1} \end{pmatrix} \underline{u}_{t-1}, \qquad (6)$$

where $\mathbf{b}_{1:q-1} = (b_1, \dots, b_{q-1})$ and I_k is the $k \times k$ identity matrix. Notice that $(\underline{c}_t, A_t)$ is an iid sequence of matrices. Let $\|A\| = \sum |a_{ij}|$ for any matrix

[2]Standard bilinear models only allow terms of the form $\epsilon_{t-i} u_{t-j}$ with $i, j > 0$.

$A = (a_{ij})$. Since $E(\log^+ |\epsilon_t|) \leq E|\epsilon_t| < \infty$ we have $E(\log^+ \|A_t\|) < \infty$, and thus we can define the top-Lyapunov exponent $\gamma(\mathbf{A})$ of the sequence $\mathbf{A} = (A_t)$:

$$\gamma(\mathbf{A}) := \inf_{t\in\mathbb{N}^*} \frac{1}{t} E(\log \|A_t A_{t-1} \dots A_1\|) = \lim_{t\to\infty} a.s. \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\|. \tag{7}$$

If $\gamma(\mathbf{A}) < 0$, the unique strictly stationary solution to (6), in view of Bougerol and Picard (1992a, Theorem 1.1), is

$$\underline{u}_t = \underline{c}_t + \sum_{k=1}^{\infty} A_t A_{t-1} \dots A_{t-k+1} \underline{c}_{t-k}. \tag{8}$$

It is straightforward that the strict stationarity of $(\underline{u}_t)$ is equivalent to the strict stationarity of (u_t). It is also seen that the strictly stationary solution is nonanticipative (*i.e.* with u_t function of the ϵ_{t-i}, $i \geq 0$) and ergodic, as a function of the iid process (ϵ_t). By Lemma 2 given in the appendix, and Theorem 2.5 in Bougerol and Picard (1992a), the sufficient condition $\gamma(\mathbf{A}) < 0$ is also necessary for the existence of a nonanticipative strictly stationary solution.

When $q = 1$ we have $\gamma(\mathbf{A}) = E \log |b\epsilon_t|$, and the strict stationarity condition takes the simpler form:

$$|b| < e^{-E \log |\epsilon_t|}. \tag{9}$$

Note that, applying the same method of proof as in Quinn (1982, Theorems 1 and 2), one could show, without using the general results of Bougerol and Picard, that (9) is necessary and sufficient for the strict stationarity in the case $q = 1$ and $E(\log |\epsilon_t|)^2 < \infty$. In particular when ϵ_t is Gaussian, the necessary and sufficient condition is $|b|\sigma_\epsilon < 1.88736$.

When $q > 1$, the strict stationarity region can not be given explicitly. In Figure 2, the strict stationarity region has been evaluated using (7) and simulations of the sequence (A_t) in the case $q = 2$ and $\epsilon_t \sim \mathcal{N}(0,1)$. The strict stationarity curve passes at the points $(b_1, b_2) = \left(\pm e^{-E \log|\epsilon_t|}, 0\right)$, as can be seen from (9), and at the points $(b_1, b_2) = \left(0, \pm e^{-E \log|\epsilon_t|}\right)$, as can be shown by algebraic computations. It is interesting to note that the stationarity region is not symmetric with respect to the diagonal $b_1 = b_2$.

3.2 Second-order stationarity

Results concerning the existence of second-order stationary solutions of bilinear models are well-known, and they can be straightforwardly extended to our model. Let (u_t) be a solution to the 2nd equation in (5). Then it is easily seen that $E(u_t) = 0$ and $E(u_t u_{t-h}) = E(\epsilon_t)E(1+b_1 u_{t-1}+\cdots+b_q u_{t-q})u_{t-h} = 0$ for any $h > 0$. Moreover, we have

$$\left(1 - \sum_{i=1}^{q} b_i^2 \sigma_\epsilon^2\right) E u_t^2 = \sigma_\epsilon^2 > 0.$$

It follows that

$$\sum_{i=1}^{q} b_i^2 \sigma_\epsilon^2 < 1 \tag{10}$$

is a necessary condition for second-order stationarity. Conversely, suppose that this condition holds true. We will show that $\underline{u}_t$ defined in (8) is the limit in L^2 of the Cauchy sequence $(\underline{u}_{tN})_N$ defined by

$$\underline{u}_{tN} = \underline{c}_t + \sum_{k=1}^{N} A_t A_{t-1} \ldots A_{t-k+1} \underline{c}_{t-k}.$$

Let $\|X\|_2^2 = E\|X\|^2$ where $\|\cdot\|$ denotes the Euclidian matrix norm. We have, with $N' > N$,

$$\begin{aligned}
\|\underline{u}_{tN'} - \underline{u}_{tN}\|_2 &\leq \sum_{k=N+1}^{N'} \left\| A_t A_{t-1} \dots A_{t-k+1} \underline{c}_{t-k} \right\|_2 \\
&= \sum_{k=N+1}^{N'} \sqrt{E \underline{c}_t' \otimes \underline{c}_t' \left(E A_t' \otimes A_t' \right)^k \operatorname{vec} I_{q^2}}, \qquad (11)
\end{aligned}$$

where the inequality follows from the Minkowski inequality, and the equality follows from independence and elementary properties of the Kronecker product $\otimes$ of matrices and the vec operator (see *e.g.* Harville (1997) for details about these matrix operators). Denote by $\rho(A)$ the spectral radius of a square matrix A. Using $\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$, it can be shown that the right-hand side of (11) tends to 0 as $N \to \infty$ if $\rho\left\{E\left(A_t \otimes A_t\right)\right\} < 1$. We have

$$E\left(A_t \otimes A_t\right) = \sigma_\epsilon^2 B \otimes B + J \otimes J,$$

where

$$B = \begin{pmatrix} \mathbf{b}_{1:q-1} & b_q \\ 0_{q-1\times q-1} & 0_{q-1} \end{pmatrix}, \qquad J = \begin{pmatrix} 0'_{q-1} & 0 \\ I_{q-1} & 0_{q-1} \end{pmatrix}.$$

By induction, it can be shown that

$$\det\left(\sigma_\epsilon^2 B \otimes B + J \otimes J - \lambda I_{q^2}\right) = (-\lambda)^{q^2} \mathcal{B}\left(\frac{1}{\lambda}\right),$$

where $\mathcal{B}(z) = 1 - \sum_{i=1}^q b_i^2 \sigma_\epsilon^2 z^i$. It is well-known that the roots of the polynomial $\mathcal{B}(z)$ are outside the unit disk if and only if (10) holds (see *e.g.* Francq and Zakoïan (2004), Proposition 1). Thus (10) entails that the spectral radius of $EA_t \otimes A_t$ is strictly less than 1, which allows to conclude that $(\underline{u}_{tN})_N$ is a Cauchy sequence in L^2. Therefore $\underline{u}_t$ is in L^2.

As in GARCH models, strict stationarity is weaker than second-order stationarity (*i.e.* $\gamma(\mathbf{A}) < 0$ does not imply $Eu_t^2 < \infty$), but it will be seen in the proof of Theorem 4.2 below that $\gamma(\mathbf{A}) < 0$ implies $E|u_t|^s$ for some $s > 0$ (see the remark in Section 4.3).

Note that Giraitis and Surgailis (2002) give sufficient conditions for strict stationarity of a class of infinite-order ARCH-type bilinear models encompassing (5). However, when applied to our model, their conditions turn out to be more restrictive than ours, which are both necessary and sufficient in our framework. Whether necessary and sufficient conditions can be obtained for infinite-order models is an open issue, to our knowledge.

The results of this section are summarized in the next theorem.

Theorem 3.1 *The second equation of (5) admits a strictly stationary solution (u_t) if and only if $\gamma(\mathbf{A}) < 0$, where $\mathbf{A} = (A_t)$ is defined in (6). Under this condition, the strictly stationary solution is unique, nonanticipative and ergodic. This solution admits a second order moment if and only if $\sum_{i=1}^{q} b_i^2 \sigma_\epsilon^2 < 1$. In this case, the solution is a conditionally heteroskedastic white noise.*

As illustrated in Figure 2, the second order stationarity region is generally much more restrictive than the strict stationarity region.

4 Mixing properties

We now turn to the most technical part. For the rest of the paper, the only result of interest in this section is Theorem 4.2 below. This theorem concerns mixing properties of the process (u_t), which will be crucial for applying

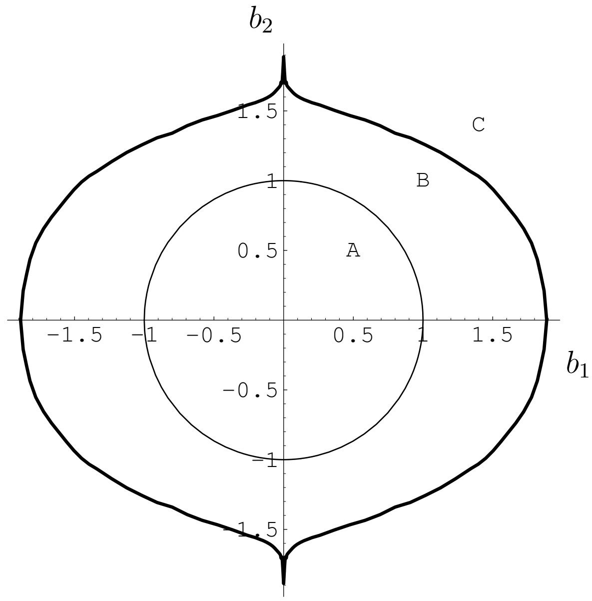

Figure 2: Strict and second-order stationarity regions of the bilinear model

$$u_t = (1 + b_1 u_{t-1} + b_2 u_{t-2})\epsilon_t, \quad \epsilon_t \text{ iid } \mathcal{N}(0,1)$$

A: second order stationarity, $\mathtt{A} \cup \mathtt{B}$: strict stationarity, and C: non stationary.

unit-root tests to Model (5). Markov chain techniques have been widely used in recent years to derive such mixing results. General conditions for ergodicity and mixing of Markov chains are provided in the book by Meyn and Tweedie (1993). References dealing with mixing properties of various classes of processes can be found in Francq and Zakoïan (2005).

4.1 Elements of Markov chain theory

Equation (6) ensures that the process $(\underline{u}_t)$ is a time homogeneous Markov chain with state space $\mathbb{R}^q$. Recall that a Markov chain (X_t) with state space $E \subset \mathbb{R}^d$ is said to be $\mu - irreducible$ for some measure μ on $(E, \mathcal{E})$ (where $\mathcal{E}$

is the Borel σ-field on E), if for all $x \in E$, for all $B \in \mathcal{E}$ such that $\mu(B) > 0$,

$$\text{there exists } t > 0 \text{ such that} \quad P^t(x, B) > 0.$$

Here $P^t(x, B)$ denotes the *$t-$step transition probability* of moving from x to the set B in t steps. The measure μ is a maximal irreducibility measure if any other irreducibility measure is absolutely continuous with respect to μ. Throughout we assume the chain is μ-irreducible with μ maximal, and we denote by $\mathcal{E}^+$ the set of sets B such that $\mu(B) > 0$. If for each bounded and continuous function g on E, the function of x given by $E(g(X_t)|X_{t-1} = x)$ is continuous, the chain is said to be *a Feller chain.* The Markov chain (X_t) is said to be geometrically ergodic if there exists $\rho \in (0, 1)$ such that

$$\rho^{-t}\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \longrightarrow 0, \quad \text{as } t \to \infty,$$

for each $x \in E$, where π denotes the *invariant measure* of the Markov chain, *i.e.* a probability measure π such that $\pi(B) = \int_E \pi(dx)P(x, B)$ for all $B \in \mathcal{E}$, and $\|\cdot\|_{TV}$ is the total variation norm. A consequence of geometric ergodicity is *strong mixing with geometric rate.* A strictly stationary Markov chain (X_t) is said to be strongly mixing with geometric rate if there exists constants $K > 0$ and $\rho \in (0, 1)$ such that

$$\sup_{f,g} |\mathrm{Cov}(f(X_0), g(X_t))| \leq K\rho^t, \quad \text{for all } t > 0,$$

where the sup is taken over all functions f and g such that $|f| \leq 1, |g| \leq 1$.

A set $C \in \mathcal{E}$ is called *ν_m-small* if there exists $m > 0$ and a non-trivial measure ν_m on $\mathcal{E}$ such that: $\quad P^m(x, B) \geq \nu_m(B), \quad$ for all $x \in C, B \in \mathcal{E}$. Let C be a ν_M-small set, where the measure ν_M is such that $\nu_M(C) > 0$. Such a measure exists whenever $C \in \mathcal{E}^+$, see Meyn and Tweedie

(1993, Proposition 5.2.4). Let $E_C = \{m \geq 1 |\ C$ is ν_m-small, with $\nu_m = \delta_m \nu_M$, for a positive constant $\delta_m\}$. Then, if (X_t) is a μ-irreducible Markov chain, and if $C \in \mathcal{E}^+$, the greatest common divisor d of the set E_C does not depend on C and is called period of the Markov chain. If $d = 1$, (X_t) is said to be *aperiodic*.

4.2 A general criterion for geometric ergocity

The following criterion for geometric ergocity is obtained from a straightforward adaptation of Meyn and Tweedie (1993, Theorem 19.1.3). It has the particularity of being based on m-step transitions, instead of 1-step transitions as is usually the case.

Theorem 4.1 *Assume that*

(i) (X_t) is a μ-irreducible Feller chain, for some measure μ on $(E, \mathcal{E})$ whose support has non-empty interior,

(ii) (X_t) is an aperiodic chain,

(iii) there exists a compact set $C \subset E$, an integer $m \geq 1$, and a nonnegative continuous function (test function) $g : E \to [0, +\infty)$ such that

$$\begin{aligned} E[g(X_{t+m})|X_t = x] &\leq (1-\beta)g(x) - \beta, \qquad x \in C^c, \\ E[g(X_{t+m})|X_t = x] &\leq b, \qquad x \in C, \end{aligned}$$

for some strictly positive constants β and b. Then (X_t) is geometrically ergodic.

4.3 Application to our bilinear model

Now we are in a position to state the main result of this section.

Theorem 4.2 *Let f be the density of ϵ_t and assume that $f > 0$. If $\gamma(\mathbf{A}) < 0$, where $\mathbf{A} = (A_t)$ is defined in (6), then the strictly stationary solution process (u_t) is strongly mixing with geometric rate.*

The proof of is given in the appendix, and uses the following lemma.

Lemma 1 *Let X be an almost surely positive random variable. If $EX^r < \infty$ for some $r > 0$ and if $E \log X < 0$, then there exists $s > 0$ such that $EX^s < 1$.*

Under a slightly different form, this result is contained in the proof of Lemma 2.3 by Berkes, Horváth and Kokoszka (2003). For the convenience of the reader, a complete proof of the lemma is given in the appendix.

REMARK: The proof of Theorem 4.2 allows to show that the strictly stationary solution verifies

$$E|u_t|^s < \infty \qquad \text{for some } s > 0.$$

Indeed we have

$$E|u_t|^s \leq E\|\underline{u}_t\|^s \leq E\|\underline{c}_1\|^s \left\{1 + \sum_{k=0}^{\infty} \rho^k \sum_{i=1}^{k_0} \left(E\|A_1\|^s\right)^i\right\} < \infty$$

for s satisfying (25) in the Appendix. We also used the elementary inequality $\left(\sum_i a_i\right)^s \leq \sum_i a_i^s$ for any sequence of positive numbers a_i.

5 Unit-root testing

For processes whose differences may exhibit serial correlation, the Phillips-Perron and augmented DF tests are arguably the most popular unit-root tests. Both of them have been derived under precise assumptions, the validity of which is questionable in the model of this paper. We start by considering the Phillips-Perron test.

5.1 Phillips-Perron tests

In his seminal paper, Phillips (1987) studied the random walk

$$x_t = ax_{t-1} + v_t, \quad a = 1, \quad t = 1, 2, \ldots,$$

where the initial value x_0 may be any random variable whose distribution is fixed. He showed that the standard least squares estimator

$$\hat{a}_n := \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2}$$

consistently estimates $a = 1$, under very general assumptions on the error terms v_t. More precisely, denoting by $\alpha_v(k)$ the strong mixing coefficients of the process (v_t), Phillips found that under the assumptions

i) $Ev_t = 0$ for all t,

ii) $\sum_{k=1}^{\infty} \{\alpha_v(k)\}^{\frac{\nu}{2+\nu}} < \infty$, for some $\nu > 0$,

iii) $\sup_t E\,|v_t|^{2+\nu} < \infty$,

iv) $\vartheta_v^2 := \lim_{n\to\infty} \operatorname{Var}\left\{n^{-1/2}\sum_{t=1}^n v_t\right\}$ exists and $\vartheta_v^2 > 0$,

the standardized least squares estimator satisfies

$$Z_\phi := n\left(\hat{a}_n - 1\right) - \frac{n^2\hat{\sigma}_{\hat{a}_n}^2}{2\hat{s}_v^2}\left(\hat{\vartheta}_v^2 - \hat{s}_v^2\right) \Rightarrow \frac{(1/2)\left\{W^2(1) - 1\right\}}{\int_0^1 W^2(t)dt}, \tag{12}$$

where $\{W(t),\ t \in [0,1]\}$ denotes a standard Brownian motion, $\hat{\vartheta}_v^2$ is a weakly consistent estimator of ϑ_v^2 defined in iv) above, $\hat{\sigma}_{\hat{a}_n}^2 = \hat{s}_v^2 / \sum_{t=2}^n x_{t-1}^2$, and

$$\hat{s}_v^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \hat{a}_n x_{t-1})^2 \tag{13}$$

is a weakly consistent estimator of $s_v^2 := Ev_t^2$. Note that $\hat{a}_n$, $\hat{\sigma}^2_{\hat{a}_n}$ and $\hat{s}_v^2$ are available in any standard regression software. For the estimation of ϑ_v^2, a HAC-type estimator can be used, as proposed by Phillips (1987). Phillips also found the asymptotic distribution of the associated regression t statistics:

$$Z_t := \frac{\hat{s}_v}{n\hat{\vartheta}_v \hat{\sigma}_{\hat{a}_n}} Z_\phi \Rightarrow \frac{(1/2)\left\{W^2(1) - 1\right\}}{\left\{\int_0^1 W^2(t)dt\right\}^{1/2}}. \tag{14}$$

5.2 Validity of the Phillips-Perron test for the bilinear process

We are interested in testing the unit-root assumption

$$H_0 : \phi = 0$$

in Model (5). We keep the notation of the previous section, with x_t replaced by y_t (and thus $v_t = y_t - y_{t-1}$). The next theorem states that (12) and (14) hold under H_0. A drift term and/or a deterministic time trend could be added to our model, leading to the limiting distributions obtained by Phillips and Perron (1988). The stochastic unit-root hypothesis can then be tested by the standard Phillips-Perron tests, in exactly the same way as when the unit root is not stochastic.

Theorem 5.1 *Assume that in Model (5), the zeroes of the polynomial $\psi(z) := 1 - \sum_{i=1}^{p} \psi_i z^i$ are outside the unit disk, that the stationary solution of the second equation in (5) satisfies $E|u_t|^{2+\nu} < \infty$ for some $\nu > 0$, and that the assumptions of Theorem 4.2 are satisfied. Under H_0 the weak convergences (12) and (14) hold.*

The proof is given in the appendix. The estimator $\hat{s}_v^2$ can be replaced by the simpler estimator $n^{-1}\sum_{t=1}^{n}(x_t - x_{t-1})^2$. Phillips (1987, Theorem 4.2) shows that there exists a consistent HAC estimator $\hat{\vartheta}_v^2$ under the addition moment assumption $E|u_t|^{4+\nu} < \infty$. As stated in Theorem 5.1, other estimators than the HAC may be employed. The choice of the estimators of s_v^2 and ϑ_v^2 may however be important for the behavior of the statistics Z_ϕ and Z_t in finite samples and/or under the alternative $\phi \neq 0$.

For $\alpha \in (0,1)$, let $\mathfrak{df}_\phi(\alpha)$ and $\mathfrak{df}_t(\alpha)$ be the α-quantiles of the distributions defined in the right-hand sides of (12) and (14). These quantiles are given in Fuller (1976, p. 371). In particular $\mathfrak{df}_\phi(5\%) = -8.1$ and $\mathfrak{df}_t(5\%) = -1.95$. The alternative we consider is

$$H_1 : (1-z)\psi(z) - \phi z \neq 0 \text{ when } |z| \leq 1.$$

Under H_1 we assume that (y_t) is the nonanticipative stationary solution of (5). The following result shows, as an immediate consequence of Theorem 5.1, that the asymptotic level of the standard Phillips-Perron test remains valid in our framework. The consistency is less trivial, and is shown in the appendix.

Corollary 5.1 *We suppose that the assumptions of Theorem 5.1 are satisfied. Under the unit-root assumption H_0,*

$$\lim_{n\to\infty} P\left\{Z_\phi \leq \mathfrak{df}_\phi(\alpha)\right\} = \alpha \quad \textit{and} \quad \lim_{n\to\infty} P\left\{Z_t \leq \mathfrak{df}_t(\alpha)\right\} = \alpha$$

and under the stationarity assumption H_1,

$$\lim_{n\to\infty} P\left\{Z_\phi \leq \mathfrak{df}_\phi(\alpha)\right\} = 1 \quad \textit{and} \quad \lim_{n\to\infty} P\left\{Z_t \leq \mathfrak{df}_t(\alpha)\right\} = 1.$$

The last limit is obtained with the restrictions $\limsup_{n\to\infty} \hat{\vartheta}_v^2 < \infty$ *a.s and* $\hat{\vartheta}_v^2 > 0$ *a.s for all* n*. The consistency of the* Z_ϕ*-based test is obtain whatever the nonnegative estimator* $\hat{\vartheta}_v^2$*.*

5.3 Augmented DF tests

The approach followed by Dickey and Fuller (1979) is based on the pth-order autoregression defined by the first equation of (5):

$$\Delta y_t = (\phi, \boldsymbol{\psi}')X_t + u_t, \qquad \text{where} \qquad X_t = (y_{t-1}, V_t')',$$

$V_t = (\Delta y_{t-1}, \dots \Delta y_{t-p})'$ and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)'$. The least-squares estimator of $(\phi, \boldsymbol{\psi}')'$ is defined by

$$(\hat{\phi}, \hat{\boldsymbol{\psi}}')' = \left(\sum_{t=1}^n X_t X_t'\right)^{-1} \sum_{t=1}^n \Delta y_t X_t, \qquad \hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \dots, \hat{\psi}_p)'.$$

The following theorem is similar to Theorem 5.1-Corollary 5.1. For the sake of conciseness we only consider the test based on $\hat{\phi}$, and we omit the studentized version.

Theorem 5.2 *Assume Model (5) satisfies the assumptions of Theorem 5.1. Under* H_0

$$\mathrm{DF}_\phi := n\frac{\hat{\phi}}{1 - \hat{\psi}_1 - \dots - \hat{\psi}_p} \Rightarrow \frac{(1/2)\left\{W^2(1) - 1\right\}}{\int_0^1 W^2(t)dt} \tag{15}$$

and, under the additional moment assumption $Eu_t^4 < \infty$*,*

$$\sqrt{n}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right) \Rightarrow \mathcal{N}\left\{0, \Sigma_{\boldsymbol{\psi}} := \left(EV_tV_t'\right)^{-1}\left(Eu_t^2 V_t V_t'\right)\left(EV_tV_t'\right)^{-1}\right\}. \tag{16}$$

We have $\lim_{n\to\infty} P\left\{\mathrm{DF}_\phi \le \mathfrak{df}_\phi(\alpha)\right\} = \alpha$ *under the unit-root assumption* H_0*, and* $\lim_{n\to\infty} P\left\{\mathrm{DF}_\phi \le \mathfrak{df}_\phi(\alpha)\right\} = 1$ *under the stationarity assumption* H_1*.*

As in the case of an independent noise, the asymptotic null-distribution of $\sqrt{n}\left(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}\right)$ is the same whether the variable y_{t-1} is included or not in the regression (of course only in the case $\phi=0$). However, the asymptotic variance $\Sigma_{\boldsymbol{\psi}}$ depends on the noise distribution through the b_i coefficients and the moments of ϵ (see the example below). This is not surprising because the asymptotic variance of the LS estimator in stationary ARMA models is modified when, in the noise assumptions, independence is replaced by uncorrelatedness (see Francq and Zakoïan, 1998). Interestingly, this is not the case for the distribution of $\hat{\phi}$ which turns out to be the same as for an independent noise.

In the simple case $p=1$ with $\phi=\psi_1=0$, $b_1^4<1/3$ and $\epsilon_t\sim\mathcal{N}(0,1)$, straightforward computations show that

$$\Sigma_{\boldsymbol{\psi}}=\frac{(1-b_1^2)(1+3b_1^2+12b_1^4)}{1-3b_1^4}.$$

It is seen that this asymptotic variance can be arbitrarily bigger (for b_1 close to 1/3) than for an iid noise.

6 Small sample properties of the standard unit-root tests

We have seen in the previous section that the standard Phillips-Perron and augmented DF tests are asymptotically valid for testing the stochastic unit-root hypothesis in Model (5). In other words the asymptotic behaviour of the tests is not affected by the presence of bilinear terms. In this section we investigate the finite-sample properties of the tests. The results of Monte-

Carlo experiments are presented in Tables 1-7 below. We start by analyzing the properties of the tests under H_0.

6.1 Size analysis

In this section, the data generating process (DGP) is (5) with $p = 0$, $q = 1$, $\phi = 0$ and $\epsilon_t \sim \mathcal{N}(0, 1)$. Five values of b (ranging from 0 to 0.99) and six sample sizes (ranging from 100 to 3,000) are considered. To save space only partial results are reported in the tables 1-7, but complementary results are available from the authors. For each experiment, the number of replications is $N = 10,000$. To estimate the long-run variance ϑ_v^2 with HAC-type estimators, the following kernels are used :

QS	Andrews Quadratic-Spectral kernel
Parzen	Parzen kernel
Fejer	Fejer, Bartlett kernel
Tuk-Han	Turkey-Hanning kernel
Triang	Triangular kernel

See Newey and West (1987) and Andrews (1991) for definitions. To flatten the spectrum of the residuals and to make the estimation of the asymptotic variance ϑ_v^2 simpler, the AR-filtering can be used (see Lee and Philipps, 1993). In this section, and in the next Section 6.2, the results are obtained without AR smoothing. The influence of AR-smoothing on the size and power of the test is considered in Section 6.3 below. To gauge if, over the N replications, the difference between the relative frequency of rejection, denoted $\hat{\alpha}$, and the nominal level α is significant or not, we compute the statistic

$$z = \frac{\hat{\alpha} - \alpha}{\sqrt{\alpha(1-\alpha)/N}}. \tag{17}$$

Since N is large, this statistic roughly follows a standard gaussian distribution when α is the actual size of the test. In Tables 1-7 below, the value of z are displayed into parenthesis. Table 1 displays finite sample results for the Z_ϕ test. Results for the Z_t test are given in Table 2. The output concerning the DF test based on the studentized statistics DF_t is presented in Table 3 (without constant in the model) and in Table 4 (with a constant).

Analysing Tables 1-4 the following commentaries can be made:

1. For small values of b (corresponding to models with high-order moments for u_t), the observed frequencies of rejection are generally very close to the nominal levels, even for small sample sizes. This is true for both tests, although the Z_ϕ test performs slightly better that the Z_t test.

2. For values of b that are close to unity (that is close to violation of the second-order stationarity condition), the size distortion of both tests is higher than for small b. As expected, the size distortion decreases when n increases.

3. For small b the results for all five kernels are nearly identical, but for large b the Turkey-Hanning and the triangular kernels perform better than the other three kernels.

4. Comparing the results for the Phillips-Perron test in Table 2 with those of the DF test in Tables 3 and 4, we may notice that for b close to unity

the size distortion of the DF test is higher, so that the Phillips-Perron test looks preferable.

Table 5 summarizes results of Monte Carlo experiments for a large sample size of $n = 10,000$. We observe a very small, but still significant, deviation between the distribution of Z_ϕ and Z_t and their asymptotic distribution.

6.2 Power analysis

We turn to the properties of the tests under the stationarity assumption H_1. The DGP is now (5) with $p = 0$, $q = 1$, $\epsilon_t \sim \mathcal{N}(0,1)$ and $\phi = -0.1$ or $\phi = -0.01$. This corresponds to an AR(1) coefficient $a = \phi + 1 = 0.9$ or $a = 0.99$. Table 6 gives the relative frequencies of rejection of H_0 for the Philipps-Perron and for the DF_t statistics. The results of the Z_t-test are nearly identical for the different kernels, so that only the results for the triangular kernel (as it performs slightly better at the previous experiments) are given. From Table 6 we may notice that:

1. For $n = 100$ the power of both tests increases when b increases.
2. The powers of the two tests are very close to each other.
3. As expected, the powers of the tests are decreasing when the autoregressive coefficient is approaching unity, and are getting close to unity rapidly when the sample size is increasing.

6.3 Influence of AR-filtering

This section proposes a Monte Carlo analysis of the properties of the Phillips-Perron Z_t test when the long-run asymptotic variance ϑ_v^2 is estimated using AR-filtering.

We first investigate the possible size distortion of the test. The Monte Carlo experiment is conducted as in Section 6.1, with the only difference that the parameter p in (5) takes the values 0, 1 and 2. For the estimation of the asymptotic long-run variance ϑ_v^2, the AR-filtering is used. This transformation flattens the spectrum at the frequency zero and facilitates the estimation. The computations are based on

$$\frac{\vartheta_v^2}{2\pi} = f_v(0) = \frac{f_w(0)}{\Phi^2(0)}, \qquad \Phi(B)w_t = v_t,$$

where B is the backshift operator, Φ is an AR polynomial of order g, and $f(0)$ denotes the spectrum at frequency zero for the corresponding process. The results from the series of experiments, with different combinations of parameters p and g, are summarized in Table 7. It can be seen that:

1. For $g < p$, the size distortion of the test is highly significant, and the situation does not much improve when the sample size increases.

2. For $g > p$, the distortion of the test is slightly more important than for $g = p$, but diminishes when the sample size n or/and the parameter b increase. For small n, the relative frequency of rejection is often significantly less than the nominal level (especially for the higher values of p). Nevertheless the case $g > p$ looks preferable to the case $g < p$.

3. For $g = p$ the results are similar for those in Table 2 and are more precise than for $g \neq p$.

So, in practical situation where p in (5) is unknown, underparametrization is more risky than overparametrization (especially when the sample size n is larger than 500).

The sensitivity of the power of the Z_t test to AR-filtering was also considered, using a DGP of the form (5) with $q = 1$, $p = 1$, and $\phi = -0.5$ or $\psi_1 = 0.3$ (the inverted roots of $(1 - z)\psi(z) - \phi z$ are $0.4 \pm 0.3742i$, so that the process is stationary). Monte Carlo experiments not reported here reveal that the power of the test is not sensitive to the way of estimating the spectrum at the frequency zero, and is getting close to unity very fast, even for small sample sizes and for b close to one.

7 Conclusion

In this paper we considered a class of AR models with bilinear innovations, in the spirit of Charemza et al (2005) but suitable for $I(1)$ series. This specification can be seen as a stochastic unit-root model. From another viewpoint this model is also of the GARCH type and displays asymmetries. By comparison with the standard GARCH, the news impact curve is shifted both horizontally, and vertically towards zero. We established necessary and sufficient strict and second-order stationarity conditions. We showed that the strict stationary solution is geometrically ergodic. Testing for unit roots in the presence of conditional heteroscedasticity is clearly important in financial applications, in particular to know if the economic shocks are persistent or not. The ergodicity results were used to demonstrate that the standard Phillips-Perron and augmented DF tests are asymptotically valid in this framework, which is not the case for other stochastic unit-roots models recently considered in the literature. Of course other statistical problems are of interest for the model of this paper. Estimation and testing issues are the object of a companion paper which is available from the authors.

The output of our Monte Carlo experiments can be summarized as follows. The presence of bilinear terms is sensible in finite samples, however the size distortion is tiny for moderate and large sample sizes. Another teaching from our experiments is that the Phillips-Perron test performs slightly better than the augmented DF test. From these numerical experiments and the asymptotic study, we draw the conclusion that the range of application of the conventional unit-root tests is broader than the sole detection of deterministic unit-roots.

References

Berkes, I., Horváth, L. and P.S. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bougerol, P. and N. Picard (1992a) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.

Bougerol, P. and N. Picard (1992b) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Charemza, W.W., M. Lifshits and S. Makarova (2005) Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results. *Journal of Economic Dynamics and Control*, **29**, 63–96.

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

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

Francq, C. and J-M. Zakoïan (1998) Estimating Linear Representations of Nonlinear Processes. *Journal of Statistical Planning and Inference*, 68, 145–165.

Francq, C., and J.M. Zakoïan (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli*, 10, 605-637.

Francq, C., and J.M. Zakoïan (2005) Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory* to appear.

Fuller, W. A. (1976) *Introduction to Statistical Time Series.* Wiley, New York.

Giraitis, L., Robinson, P.M., and D. Surgailis (2000) A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability* 10, 1002–1024.

Giraitis, L. and D. Surgailis (2002) ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications* 100, 275–300.

Granger, C.W.J. and A.P. Andersen (1978) *An introduction to bilinear time series models.* Vandenhoek and Ruprecht, Gottingen.

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

Hamilton, J.D. (1994) *Time Series Analysis.* Princeton University Press, New Jersey.

Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective.* Springer, New-York.

Herrndorf, N. (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

Kim, K. and P. Schmidt (1993) Some evidence on the accuracy of Phillips-Perron tests using alternative estimates of nuisance parameters. *Journal of Econometrics* 59, 345–350.

Lee, C.C. and P.C.B. Phillips (1993) *An ARMA-Prefiltered Estimator of the Long Run Variance with an Application to Unit Roots tests.* Cowles Foundation Yale University, mimeographed.

Leybourne, S.J., McCabe, B.P.M. and A.R. Tremayne (1996) Can Economic Time Series Be Differenced to Stationarity? *Journal of Business and Economic Statistics*, 14, 435–446.

Ling, S. (2004) Estimation and testing of stationarity for double-autoregressive models. *Journal of the Royal Statistical Society: Series B*, 66, 63–78.

Ling, S. and W.K. Li (2003) Asymptotic inference for unit root with GARCH (1,1) errors. *Econometric Theory* 19, 541–564.

Liu, J. and P.J. Brockwell (1988) On the general bilinear time series model. *Journal of Applied Probability*, 25, 553–564.

Meyn, S.P. and R.L. Tweedie (1993) *Markov Chains and Stochastic stability*, 3rd Edition, Springer, London.

Nelson, D. B. (1991) Conditional Heteroskedasticity in Asset Returns : a New Approach. *Econometrica*, 59, 347–370.

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

Phillips, P.C.B (1987) Time Series Regression with a Unit Root. *Econometrica*, 55, 277–301.

Phillips P.C.B. and P. Perron (1988) Testing for a Unit Root in Time Series Regression *Biometrika*, 75, 335–346.

Quinn, B.G. (1982) A note on the existence of strictly stationary solutions to bilinear equations. *Journal of Time Series Analysis*, 3, 249–252.

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47, 67–84.

Rodrigues, P.M.M. and A. Rubia (2005) The performance of unit root tests under level-dependent heteroskedasticity. *Economics Letters*, to appear.

Seo, B. (1999) Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, 91, 113–144.

Sentana, E. (1995) Quadratic ARCH models. *Review of Economic Studies*, 62, 639–661.

Subba Rao, T. and M.M. Gabr (1984) *An introduction to bispectral analysis and bilinear time series models.* Lecture Notes in Statistics, Springer, New-York.

Zakoïan, J.M. (1994) Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, 18, 931–955.

APPENDIX

We first establish a lemma, which allows to apply Theorem 2.5 in Bougerol and Picard (1992a). An affine subspace H of $\mathbb{R}^q$ is said to be invariant under (6) if it satisfies

$$\forall x \in H, \quad A_1 x + \underline{c}_1 \in H \quad \text{a.s.} \tag{18}$$

Model (6) is said to be irreducible if $\mathbb{R}^q$ is the unique invariant affine subspace. Note that this notion of irreducibility is different from the one used in Section 4.

Lemma 2 *Model (6) is irreducible.*

Proof. For simplicity, we only give the proof for $q = 2$. The arguments are the same for $q > 2$, but the proof requires tedious notations in the general case. Let H be an affine subspace of $\mathbb{R}^2$ satisfying (18). By stationarity, we have, $\forall x = (x_1, x_2)' \in H$,

$$A_2(A_1 x + \underline{c}_1) + \underline{c}_2 = \begin{pmatrix} \epsilon_2(b_1^2 x_1 \epsilon_1 + b_1 b_2 x_2 \epsilon_1 + b_2 x_1 + b_1 \epsilon_1 + 1) \\ \epsilon_1(b_1 x_1 + b_2 x_2 + 1) \end{pmatrix} \in H \quad \text{a.s.} \tag{19}$$

Taking the expectation of the vector defined in (19), we obtain $0 \in H$. Taking $x = 0$ in (18) and (19), we obtain

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in H, \quad \begin{pmatrix} \epsilon_1 \\ 0 \end{pmatrix} \in H \quad \text{a.s.}, \quad \begin{pmatrix} \epsilon_2(\epsilon_1 b_1 + 1) \\ \epsilon_1 \end{pmatrix} \in H \quad \text{a.s.} \tag{20}$$

Since $\sigma_\epsilon > 0$, ϵ_1 is not almost surely equal to 0. Thus (20) entails that the linear subspace $H = \mathbb{R}^2$.

□

Proof of Lemma 1. The moment generating function of $Y = \log X$ is given by $M(u) = Ee^{uY} = EX^u$. The function M is continuously differentiable over $[0, r]$ and we have, for $u > 0$

$$\frac{M(u) - M(0)}{u} = \int \frac{e^{uy} - 1}{u} \, dP_Y(y). \tag{21}$$

We begin to show that

$$\text{for all } \tau > 0, \quad \text{and for all } u \in]0, \tau], \qquad \left| \frac{e^{uy} - 1}{u} \right| \leq \frac{e^{\tau|y|}}{\tau}. \tag{22}$$

This result is obtained, for instance, by considering the function defined by $g(v) = \frac{e^{vy}-1}{v}$ for $v \neq 0$ and $g(0) = y$. The function g being increasing on $\mathbb{R}$, we have for $y \geq 0$,

$$\frac{e^{uy} - 1}{u} \leq \frac{e^{\tau y} - 1}{\tau} \leq \frac{e^{\tau y}}{\tau},$$

and for $y < 0$

$$\frac{1 - e^{uy}}{u} \leq -y \leq \frac{e^{-\tau y}}{\tau}$$

which establishes (22). Now note that

$$\int \frac{e^{\tau|y|}}{\tau} dP_Y(y) = \frac{Ee^{\tau \log X} + 1}{\tau} \leq \frac{EX^r + 1}{\tau} < \infty$$

when $\tau \in]0, r]$. By the Lebesgue theorem, it follows that the right-derivative of M at 0 is, in view of (21)

$$\int y dP_Y(y) = E(\log X) < 0.$$

Since $M(0) = 1$, there exists $s > 0$ such that $M(s) = EX^s < 1$.

□

Proof of Theorem 4.2. To establish the geometric ergodicity of $(\underline{u}_t)$ defined by (8) we verify the three conditions of Theorem 4.1.

Let for $x = (x_1, \dots, x_q)' \in \mathbb{R}^q$, $\psi(x) = 1 + \sum_{i=1}^q b_i x_i$. We have

$$u_t = \psi(\underline{u}_{t-1})\epsilon_t.$$

Let λ denote the Lebesgue measure on $\mathbb{R}$. For any bounded continuous function h,

$$E(h(\underline{u}_t)|\underline{u}_{t-1} = x) = \int h(\psi(x)\epsilon, x_1, \dots, x_{q-1}) f(\epsilon)\lambda(d\epsilon)$$

is a continuous function of $x = (x_1, \dots, x_q)$, by continuity of ψ and h and by application of the Lebesgue theorem. It follows that $(\underline{u}_t)$ is a Feller chain.

Now we will check that $(\underline{u}_t)$ is λ_q-irreducible, where λ_q is the Lebesgue measure on $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$. To avoid cumbersome notations we will only establish this result when $q = 2$, the extension to higher dimensions being straightforward. For $B \in \mathcal{B}(\mathbb{R}^2)$ and $x = (x_1, x_2) \in \mathbb{R}^2$ we have

$$P^2(x, B) = P\left\{(u_2, u_1) \in B\right\}, \text{ where } u_1 = \epsilon_1\psi(x),\ u_2 = \epsilon_2\psi(u_1, x_1). \quad (23)$$

First consider x such that $\psi(x) \neq 0$. Let $T_x : (\epsilon_1, \epsilon_2) \mapsto (u_1, u_2)$. Let ϵ_1^0 be the point such that $\psi(u_1^0, x_1) = 0$ where $u_1^0 = \psi(x)\epsilon_1^0$. The mapping T_x is one-to-one from $(\mathbb{R} \setminus \{\epsilon_1^0\}) \times \mathbb{R}$ to $(\mathbb{R} \setminus \{u_1^0\}) \times \mathbb{R}$, and admits continuous derivatives. Since (ϵ_1, ϵ_2) admits a density, the change-of-variables theorem shows that (u_1, u_2) also admits a density. In view of (23), it follows that $P^2(x, B) > 0$ whenever $\lambda_2(B) > 0$.

Now consider x such that $\psi(x) = 0$. The previous argument fails because the distribution of $(u_1, u_2) = (0, u_2)$ has no density with respect to λ_2. The problem is easily solved by considering three-steps transition probabilities, and by showing that (u_2, u_3) has a density whenever $\psi(0, x_1) \neq 0$. When $\psi(x) = \psi(0, x_1) = 0$, four-steps transition probabilities allow to conclude that (u_3, u_4) has a density. Hence for all x, if $\lambda_2(B) > 0$ then $P^t(x, B) > 0$ for some $t \in \{2, 3, 4\}$. This completes the proof of (i).

To prove (ii) we will still limit ourselves to the case $q = 2$. Let C be a compact subset of $\mathbb{R}^2$ such that $\lambda_2(C) > 0$ and $\psi(x) \neq 0$ for any $x \in C$. We have just seen that, for any $x \in C$, $P^2(x, B) > 0$ whenever $\lambda_2(B) > 0$. Moreover, by continuity of the function $x \to P^2(x, B)$, the compactness of C entails that $\inf_{x \in C} P^2(x, B) = P^2(x^*, B) > 0$, for some $x^* \in C$. Setting $\nu_2(B) = P^2(x^*, B)$, we define a non-trivial measure on $\mathcal{B}(\mathbb{R}^2)$. It follows that C is a ν_2-small set. Now, consider the five-step transitions. We have

$$\begin{aligned} P^5(x, B) &\geq \int_C P^3(x, dy) P^2(y, B) \\ &\geq P^2(x^*, B) \inf_{x \in C} P^3(x, C) = P^2(x^*, B) P^3(x^{**}, C), \end{aligned}$$

for some $x^{**} \in C$. By arguments similar to those used in the proof of step (i), we show that $P^3(x, C) > 0$ for all $x \in C$, and thus we have $P^3(x^{**}, C) > 0$. Hence C is also ν_5-small, with $\nu_5 = P^3(x^{**}, C)\nu_2$. We can conclude that $m = 2$ and $m = 5$ belong to the set E_C defined in Section 4.1. Hence $d = 1$ and the aperiodicity of $(\underline{u}_t)$ is established.

Finally, we will verify condition (iii). Since $\gamma(\mathbf{A}) < 0$, there exists an integer $k > 0$ such that $E(\log \|A_t A_{t-1} \dots A_{t-k}\|) < 0$ (see the first definition of $\gamma(\mathbf{A})$ given in (7) and use the strict stationarity of the sequence (A_t)). On the other hand, we have

$$\begin{aligned} E(\|A_t A_{t-1} \dots A_{t-k}\|) &\leq E\|A_t\| E\|A_{t-1}\| \dots E\|A_{t-k}\| \\ &\leq (E\|A_t\|)^{k+1} < \infty \qquad (24) \end{aligned}$$

using the facts that the norm is multiplicative and that the matrices A_t are iid. Lemma 1 entails the existence of some $s \in]0,1[$ such that

$$\rho := E(\|A_t A_{t-1} \dots A_{t-k}\|^s) < 1. \tag{25}$$

By a recursive expansion of the first equality in (6) we get

$$\underline{u}_t = \underline{c}_t + A_t \underline{c}_{t-1} + \cdots + A_t \dots A_{t-k+1} \underline{c}_{t-k} + A_t \dots A_{t-k} \underline{u}_{t-k-1}$$

and thus, the norm being multiplicative,

$$\|\underline{u}_t\| \le \sum_{i=0}^{k} \|A_t \dots A_{t-i+1}\| \|\underline{c}_{t-i}\| + \|A_t \dots A_{t-k}\| \|\underline{u}_{t-k-1}\|,$$

the first term in the sum, for $i = 0$, being equal to $\|\underline{c}_t\|$ by convention. Because $s \in [0,1)$, it follows from the elementary inequality $(a+b)^s \le a^s + b^s$, for $a \ge 0$ and $b \ge 0$, that

$$\|\underline{u}_t\|^s \le \sum_{i=0}^{k} \|A_t \dots A_{t-i+1}\|^s \|\underline{c}_{t-i}\|^s + \|A_t \dots A_{t-k}\|^s \|\underline{u}_{t-k-1}\|^s.$$

Taking the expectations in both sides, conditionally on $\underline{u}_{t-k-1} = \underline{x}$, yields

$$\begin{aligned} E(\|\underline{u}_t\|^s \mid \underline{u}_{t-k-1} = \underline{x}) &\le \sum_{i=0}^{k} E\|A_t \dots A_{t-i+1}\|^s E\|\underline{c}_{t-i}\|^s + \rho\|\underline{x}\|^s \\ &\le K + \rho\|\underline{x}\|^s. \end{aligned} \tag{26}$$

The first inequality uses the independence between the A_{t-j} and $\underline{c}_{t-i}$ for $i > j$, and the independence between these matrices and $\underline{u}_{t-k-1}$ for $k \ge i$. The latter independence is a consequence of the fact that the stationary solution is nonanticipative. The second inequality in (26) follows from arguments similar to those used to show (24). Let $\beta > 0$ such that $1 - \beta > \rho$ and let C the subset of $[0, +\infty)^q$ defined by

$$C = \{\underline{x} \mid (1 - \beta - \rho)\|\underline{x}\|^s \le K + \beta\}.$$

Clearly $C \neq \emptyset$ since $K+\beta > 0$. Moreover C is compact because $1-\beta-\rho > 0$. Thus the right-hand side of (26) is bounded by a constant over C, and it is bounded by $(1-\beta)\|\underline{x}\|^s - \beta$ over the complement of C. It follows that condition (iii) in Theorem 4.1 is verified, with $g(\underline{x}) = \|\underline{x}\|^s$, $m = k+1$, and β and C chosen as indicated above.

□

Proof of Theorem 5.1. Note that the existence of $E|u_t|^{2+\nu}$ entails (10). First consider the case $p = 0$. Then $v_t = u_t$, and i)–iv) are straightforwardly satisfied with $\vartheta_v^2 = s_v^2 = \sigma_\epsilon^2/(1 - \sum_{i=1}^q b_i^2 \sigma_\epsilon^2)$. Thus, when the DGP does not contain augmented variables, the result directly follows from Phillips (1987). In the case $p > 0$, it is not obvious to know whether $v_t = \psi^{-1}(B)u_t := \sum_{i=0}^\infty c_i u_{t-i}$ inherits the mixing property of (u_t) or not. Fortunately, conditions i)-iv) are not necessary for (12) and (14). Conditions i)-iv) are those given by Herrndorf (1984) to establish the functional central limit theorem (FCLT) for (v_t). Other conditions ensuring the FCLT rely on the concept of near-epoch dependence (NED), see Davidson (1994). The process (v_t) is geometrically L_2-NED on the process (u_t) because the sequence

$$\begin{aligned} \|v_t - E(v_t|u_{t-m}, \ldots, u_{t+m})\|_2 &= \sum_{i=m+1}^\infty |c_i| \, \|u_{t-i} - E(u_{t-i}|u_{t-m}, \ldots, u_{t+m})\|_2 \\ &\leq 2\|u_t\|_2 \sum_{i=m+1}^\infty |c_i| \end{aligned}$$

tends to zero at an exponential rate as $m \to \infty$. In view of this property, the exponential decrease of the α-mixing coefficients of (u_t), and the fact that iv) holds with

$$\vartheta_v^2 = \frac{\sigma_\epsilon^2}{(1 - \sum_{i=1}^p b_i^2 \sigma_\epsilon^2)\psi^2(1)} > 0,$$

we can conclude from Corollary 29.7 in Davidson (1994), that

$$\left(\frac{1}{\sqrt{n}\vartheta_v} S_{[nt]}\right)_{t\in[0,1]} \Rightarrow (W(t))_{t\in[0,1]}\,, \qquad (27)$$

where $S_k = v_1 + \ldots + v_k$ and $[\cdot]$ denotes the integer part. As shown by Phillips (1987), (12) and (14) are direct consequences of the FCLT (27) and of the continuous mapping theorem, which completes the proof.

□

Proof of Corollary 5.1. Under H_1 we have

$$y_t = y_{t-1} + \phi y_{t-1} + \sum_{i=1}^{p} \psi_i \Delta y_{t-i} + u_t = \psi^{*-1}(B)u_t = \sum_{i\geq 0} \pi_i u_{t-i},$$

where $\psi^*(z) = (1-z)\psi(z) - \phi z$. The process (y_t) is then stationary, ergodic and centered. Thus with probability one, we have

$$\hat{a}_n \to a^* := \frac{Ey_t y_{t-1}}{Ey_t^2} < 1,$$

where the inequality follows from the Cauchy-Schwarz inequality and the fact that the innovations of (y_t) are non degenerated. Let $\hat{v}_t^* = y_t - \hat{a}_n y_{t-1}$ and $v_t^* = y_t - a^* y_{t-1}$. The ergodic theorem also shows that

$$\begin{aligned} \hat{s}_v^2 &= \frac{1}{n-1}\sum_{t=1}^{n} v_t^{*2} = \frac{1}{n-1}\sum_{t=1}^{n} y_t^2 - \frac{2\hat{a}_n}{n-1}\sum_{t=1}^{n} y_t y_{t-1} + \frac{\hat{a}_n^2}{n-1}\sum_{t=1}^{n} y_{t-1}^2 \\ &\to s_{v^*}^2 = Ev_t^{*2} = (1-a^{*2})Ey_t^2. \end{aligned}$$

Therefore we have almost surely

$$\begin{aligned} \limsup_{n\to\infty} Z_\phi/n &= \limsup_{n\to\infty} \left\{ \hat{a}_n - 1 - \frac{1}{\frac{2}{n}\sum_{t=1}^{n} y_t^2}(\hat{\vartheta}_v^2 - \hat{s}_v^2) \right\} \\ &\leq a^* - 1 + \frac{s_{v^*}^2}{2Ey_t^2} = -(1-a^*)\left(1 - \frac{1+a^*}{2}\right) < 0, \end{aligned}$$

which shows the consistency of the Z_ϕ-based test. The consistency of the Z_t-based test comes from

$$\limsup_{n\to\infty} Z_t/\sqrt{n} \quad \leq \quad \frac{\sqrt{Ey_t^2}}{\limsup_{n\to\infty} \hat{\vartheta}_v}(a^*-1)\left(1-\frac{1+a^*}{2}\right) < 0.$$

□

Proof of Theorem 5.2. We have

$$\Lambda \begin{pmatrix} \hat{\phi} \\ \hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \end{pmatrix} = \left(\Lambda^{-1}\sum_{t=1}^n X_t X_t' \Lambda^{-1}\right)^{-1} \Lambda^{-1}\sum_{t=1}^n u_t X_t \qquad (28)$$

where $\Lambda = \mathrm{Diag}(n, \sqrt{n}, \ldots, \sqrt{n})$. We have seen that the functional CLT (27) applies to $v_t := \Delta y_t = \psi^{-1}(B)u_t$. Therefore the analogue of the results (a) and (b) of Theorem 3.1 in Phillips (1987) holds. Using also the ergodic theorem, we deduce

$$\Lambda^{-1}\sum_{t=1}^n X_t X_t' \Lambda^{-1} \Rightarrow \begin{pmatrix} \frac{Eu_t^2}{\psi^2(1)} \int_0^1 W^2(t)dt & 0_p' \\ 0_p & EV_tV_t' \end{pmatrix}.$$

Using Proposition 17.2 in Hamilton (1994) and the functional CLT applied to (u_t),

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^n u_t y_{t-1} &= \frac{1}{n}\sum_{t=1}^n u_t(y_0 + v_1 + \cdots + v_{t-1}) \\ &= \frac{1}{n\psi(1)}\sum_{t=1}^n u_t(u_1 + \cdots + u_{t-1}) + o_P(1) \\ &\Rightarrow \frac{Eu_t^2}{2\psi(1)}\left\{W^2(1) - 1\right\}. \end{aligned}$$

Moreover it is easy to show that $\hat{\psi}(1) = 1 - \hat{\psi}_1 - \cdots - \hat{\psi}_p \to \psi(1)$ almost surely. The convergence (15) follows. The convergence (16) comes from the

CLT applied to square integrable stationary martingale difference $(u_t V_t)$:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t V_t \Rightarrow \mathcal{N}\left(0, E u_t^2 V_t V_t'\right).$$

□

n	kernel	α = 0.01		α = 0.05		α = 0.10	
		b = 0.25	*b = 0.99*	*b = 0.25*	*b = 0.99*	*b = 0.25*	*b = 0.99*
100	QS	0.010 (0.101)	0.026 (16.081)	0.054 (1.652)	0.075 (11.333)	0.105 (1.6667)	0.129 (9.500)
	Parzen	0.011 (0.603)	0.025 (15.277)	0.053 (1.468)	0.072 (10.140)	0.104 (1.433)	0.126 (8.733)
	Fejer	0.010 (0.201)	0.028 (17.789)	0.052 (0.918)	0.076 (11.930)	0.103 (1.067)	0.129 (9.700)
	Tuk-Han	0.012 (1.809)	0.024 (14.473)	0.052 (1.055)	0.065 (6.791)	0.103 (0.933)	0.113 (4.167)
	Triang	0.011 (1.407)	0.025 (14.597)	0.052 (0.872)	0.069 (8.534)	0.103 (0.967)	0.119 (6.300)
500	QS	0.010 (0.201)	0.020 (10.452)	0.050 (0.138)	0.065 (6.974)	0.100 (0.000)	0.113 (4.200)
	Parzen	0.010 (-0.201)	0.020 (9.849)	0.050 (-0.046)	0.064 (6.470)	0.100 (-0.033)	0.110 (3.467)
	Fejer	0.010 (-0.201)	0.021 (11.055)	0.050 (-0.138)	0.066 (7.433)	0.100 (0.067)	0.114 (4.667)
	Tuk-Han	0.010 (0.201)	0.015 (4.824)	0.050 (-0.046)	0.055 (2.157)	0.100 (0.200)	0.098 (-0.537)
	Triang	0.010 (-0.201)	0.016 (6.432)	0.050 (-0.092)	0.058 (3.671)	0.100 (0.233)	0.102 (0.700)
1000	QS	0.012 (2.412)	0.019 (8.945)	0.050 (-0.184)	0.062 (5.644)	0.099 (-0.500)	0.115 (4.900)
	Parzen	0.012 (2.312)	0.018 (8.241)	0.050 (-0.046)	0.060 (4.726)	0.098 (-0.533)	0.113 (4.367)
	Fejer	0.012 (2.312)	0.019 (9.045)	0.050 (-0.046)	0.061 (5.231)	0.098 (-0.533)	0.115 (4.900)
	Tuk-Han	0.013 (2.613)	0.015 (5.025)	0.049 (-0.275)	0.053 (1.193)	0.098 (-0.800)	0.105 (1.500)
	Triang	0.013 (2.714)	0.016 (5.528)	0.049 (-0.413)	0.055 (2.065)	0.098 (-0.733)	0.106 (2.067)
2000	QS	0.010 (-0.402)	0.017 (7.437)	0.048 (-1.285)	0.060 (4.726)	0.101 (0.200)	0.109 (3.133)
	Parzen	0.010 (-0.402)	0.016 (6.131)	0.048 (-1.101)	0.059 (3.992)	0.101 (0.200)	0.109 (2.833)
	Fejer	0.009 (-0.804)	0.016 (6.030)	0.048 (-1.101)	0.059 (4.221)	0.101 (0.367)	0.109 (2.933)
	Tuk-Han	0.009 (-0.804)	0.013 (3.417)	0.049 (-0.872)	0.054 (1.927)	0.100 (0.067)	0.104 (1.233)
	Triang	0.010 (-0.704)	0.014 (3.618)	0.049 (-0.734)	0.055 (2.202)	0.100 (0.100)	0.104 (1.333)
3000	QS	0.012 (1.809)	0.018 (7.538)	0.052 (0.964)	0.062 (5.598)	0.103 (1.000)	0.114 (4.700)
	Parzen	0.012 (1.809)	0.017 (6.633)	0.052 (1.009)	0.061 (5.047)	0.103 (0.967)	0.113 (4.200)
	Fejer	0.012 (1.508)	0.016 (6.231)	0.052 (1.055)	0.061 (5.047)	0.103 (1.000)	0.112 (4.033)
	Tuk-Han	0.012 (2.010)	0.014 (3.819)	0.053 (1.285)	0.055 (2.157)	0.103 (1.000)	0.106 (2.100)
	Triang	0.013 (2.513)	0.013 (3.417)	0.053 (1.468)	0.056 (2.569)	0.103 (1.000)	0.106 (1.967)

Table 1: Rejection relative frequencies of the unit-root hypothesis with the Z_ϕ Phillips-Perron test, where the DGP is a unit-root process with bilinear disturbances (Model (5) with $p = 0$, $\phi = 0$ and $q = 1$). The values of the z statistic are given into parenthesis ($z \sim \mathcal{N}(0,1)$ if the nominal level is correct).

n	kernel	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
		$b = 0.25$	$b = 0.99$	$b = 0.25$	$b = 0.99$	$b = 0.25$	$b = 0.99$
100	QS	0.011 (1.006)	0.028 (17.890)	0.055 (2.157)	0.077 (12.297)	0.109 (2.900)	0.131 (10.467)
	Parzen	0.011 (0.905)	0.027 (17.287)	0.055 (2.065)	0.075 (11.287)	0.109 (3.100)	0.129 (9.767)
	Fejer	0.011 (0.603)	0.029 (19.096)	0.054 (1.744)	0.079 (13.214)	0.108 (2.567)	0.132 (10.500)
	Tuk-Han	0.013 (2.513)	0.026 (15.980)	0.053 (1.376)	0.069 (8.626)	0.109 (2.900)	0.116 (5.400)
	Triang	0.012 (1.508)	0.026 (15.880)	0.053 (1.514)	0.072 (10.048)	0.106 (1.933)	0.122 (7.233)
500	QS	0.010 (-0.201)	0.021 (10.854)	0.050 (-0.138)	0.064 (6.240)	0.102 (0.533)	0.111 (3.733)
	Parzen	0.010 (-0.402)	0.020 (10.452)	0.050 (-0.138)	0.062 (5.506)	0.102 (0.533)	0.110 (3.267)
	Fejer	0.010 (-0.302)	0.021 (11.156)	0.049 (-0.505)	0.064 (6.424)	0.102 (0.500)	0.112 (4.133)
	Tuk-Han	0.010 (0.402)	0.016 (5.930)	0.051 (0.275)	0.054 (2.019)	0.100 (0.000)	0.098 (-0.833)
	Triang	0.010 (0.302)	0.017 (7.337)	0.050 (-0.092)	0.057 (3.349)	0.100 (-0.100)	0.102 (0.500)
1000	QS	0.013 (2.915)	0.020 (9.849)	0.051 (0.229)	0.062 (5.598)	0.097 (-0.933)	0.112 (4.100)
	Parzen	0.013 (2.915)	0.018 (8.241)	0.051 (0.321)	0.061 (4.864)	0.097 (-1.033)	0.111 (3.667)
	Fejer	0.013 (2.814)	0.019 (8.945)	0.051 (0.275)	0.062 (5.322)	0.097 (-1.000)	0.112 (4.133)
	Tuk-Han	0.013 (2.915)	0.015 (4.523)	0.051 (0.505)	0.052 (0.918)	0.098 (-0.733)	0.101 (0.300)
	Triang	0.013 (2.814)	0.015 (5.226)	0.051 (0.367)	0.053 (1.606)	0.098 (-0.800)	0.105 (1.633)
2000	QS	0.010 (-0.302)	0.018 (8.040)	0.048 (-0.734)	0.060 (4.451)	0.101 (0.167)	0.109 (2.900)
	Parzen	0.010 (-0.302)	0.016 (6.432)	0.048 (-0.872)	0.059 (3.946)	0.101 (0.200)	0.107 (2.300)
	Fejer	0.010 (0.000)	0.016 (6.432)	0.048 (-0.826)	0.060 (4.451)	0.101 (0.267)	0.108 (2.500)
	Tuk-Han	0.010 (-0.503)	0.014 (3.819)	0.049 (-0.688)	0.054 (1.789)	0.100 (-0.033)	0.102 (0.833)
	Triang	0.010 (-0.302)	0.014 (3.618)	0.049 (-0.505)	0.055 (2.111)	0.100 (-0.100)	0.103 (0.900)
3000	QS	0.012 (2.312)	0.018 (8.241)	0.052 (0.964)	0.062 (5.368)	0.103 (1.000)	0.111 (3.567)
	Parzen	0.012 (2.412)	0.017 (7.236)	0.052 (1.009)	0.060 (4.588)	0.103 (0.967)	0.109 (3.033)
	Fejer	0.012 (2.111)	0.017 (6.834)	0.053 (1.193)	0.060 (4.726)	0.103 (1.000)	0.108 (2.533)
	Tuk-Han	0.013 (2.714)	0.014 (4.020)	0.053 (1.285)	0.054 (1.927)	0.104 (1.200)	0.103 (0.900)
	Triang	0.012 (2.412)	0.014 (3.920)	0.053 (1.331)	0.054 (1.881)	0.104 (1.300)	0.102 (0.633)

Table 2: As Table 1, but with the Z_t Phillips-Perron test.

n	α = 0.01		α = 0.05		α = 0.10	
	b = 0.25	*b = 0.99*	*b = 0.25*	*b = 0.99*	*b = 0.25*	*b = 0.99*
100	0.010 (0.402)	0.034 (23.618)	0.053 (1.193)	0.086 (16.656)	0.104 (1.433)	0.142 (13.933)
500	0.010 (-0.101)	0.028 (18.292)	0.049 (-0.551)	0.075 (11.700)	0.101 (0.333)	0.126 (8.667)
1000	0.013 (2.513)	0.028 (17.990)	0.051 (0.275)	0.077 (12.388)	0.097 (-0.967)	0.126 (8.733)

Table 3: As Table 1, but with the unit-root DF_t-statistics (without constant in the model).

n	α = 0.01		α = 0.05		α = 0.10	
	b = 0.25	*b = 0.99*	*b = 0.25*	*b = 0.99*	*b = 0.25*	*b = 0.99*
100	0.011 (1.106)	0.042 (32.262)	0.056 (2.799)	0.101 (23.446)	0.107 (2.267)	0.157 (19.000)
500	0.011 (0.603)	0.043 (32.764)	0.048 (-0.964)	0.103 (24.135)	0.098 (-0.767)	0.156 (18.700)
1000	0.011 (1.307)	0.042 (32.262)	0.052 (0.873)	0.104 (24.869)	0.102 (0.667)	0.155 (18.400)

Table 4: As Table 3, but a constant is included in the model

statistics	kernel	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
		$b = 0.25$	$b = 0.99$	$b = 0.25$	$b = 0.99$	$b = 0.25$	$b = 0.99$
Z_φ	QS	0.008 (-1.910)	0.014 (4.121)	0.048 (-0.918)	0.056 (2.524)	0.101 (0.267)	0.106 (2.133)
	Parzen	0.008 (-2.111	0.014 (4.322)	0.048 (-1.009)	0.055 (2.294)	0.101 (0.267)	0.105 (1.667)
	Fejer	0.008 (-2.111)	0.014 (4.121)	0.048 (-0.918)	0.055 (2.111)	0.101 (0.333)	0.104 (1.433)
	Tuk-Han	0.008 (-1.910)	0.013 (2.714)	0.048 (-1.055)	0.052 (1.147)	0.101 (0.100)	0.102 (0.567)
	Triang	0.008 (-1.709)	0.012 (1.608)	0.048 (-0.780)	0.050 (0.092)	0.101 (0.067)	0.099 (-0.300)
Z_t	QS	0.008 (-1.608)	0.014 (4.020)	0.048 (-0.380)	0.055 (1.020)	0.101 (0.367)	0.105 (1.533)
	Parzen	0.008 (-1.608)	0.013 (3.317)	0.048 (-0.380)	0.054 (0.800)	0.101 (0.300)	0.103 (1.133)
	Fejer	0.008 (-1.608)	0.014 (4.422)	0.048 (-0.360)	0.053 (0.600)	0.101 (0.400)	0.102 (0.833)
	Tuk-Han	0.008 (-1.809)	0.012 (2.211)	0.048 (-0.460)	0.052 (0.320)	0.101 (0.333)	0.099 (-0.233)
	Triang	0.008 (-1.709)	0.012 (1.910)	0.048 (-0.370)	0.049 (-0.160)	0.101 (0.200)	0.096 (-1.200)

Table 5: As Tables 1 and 2, but for $n = 10,000$.

n	b	α = 0.01				α = 0.05				α = 0.10			
		a = 0.90		a = 0.99		a = 0.90		a = 0.99		a = 0.90		a = 0.99	
		Z_t	DF_t	Z_t	DF_t	Z_t	DF_t	Z_t	DF_t	Z_t	DF_t	Z_t	DF_t
100	0.00	0.328	0.325	0.018	0.017	0.756	0.766	0.083	0.081	0.915	0.924	0.167	0.160
	0.25	0.327	0.334	0.019	0.018	0.744	0.772	0.082	0.081	0.910	0.924	0.165	0.161
	0.50	0.356	0.365	0.019	0.021	0.771	0.774	0.085	0.086	0.918	0.918	0.164	0.167
	0.75	0.381	0.402	0.021	0.028	0.775	0.775	0.090	0.103	0.919	0.909	0.162	0.182
	0.99	0.417	0.438	0.038	0.050	0.773	0.752	0.107	0.129	0.918	0.889	0.187	0.214
250	0.00	0.984	0.990	0.034	0.032	1.000	1.000	0.149	0.146	1.000	1.000	0.280	0.280
	0.25	0.984	0.985	0.034	0.033	1.000	1.000	0.148	0.147	1.000	1.000	0.275	0.274
	0.50	0.983	0.976	0.033	0.035	0.999	0.998	0.144	0.152	1.000	0.999	0.277	0.285
	0.75	0.980	0.959	0.041	0.049	0.998	0.995	0.156	0.172	0.999	0.998	0.282	0.299
	0.99	0.973	0.927	0.056	0.079	0.996	0.987	0.167	0.205	0.997	0.995	0.286	0.327
500	0.00	1.000	1.000	0.082	0.078	1.000	1.000	0.317	0.315	1.000	1.000	0.525	0.523
	0.25	1.000	1.000	0.079	0.079	1.000	1.000	0.315	0.317	1.000	1.000	0.521	0.523
	0.50	1.000	1.000	0.080	0.085	1.000	1.000	0.307	0.317	1.000	1.000	0.523	0.524
	0.75	1.000	0.999	0.085	0.103	1.000	1.000	0.316	0.332	1.000	1.000	0.524	0.540
	0.99	1.000	0.995	0.114	0.156	1.000	0.999	0.329	0.379	1.000	0.999	0.536	0.559
1000	0.00	1.000	1.000	0.312	0.307	1.000	1.000	0.750	0.754	1.000	1.000	0.916	0.920
	0.25	1.000	1.000	0.305	0.307	1.000	1.000	0.753	0.755	1.000	1.000	0.916	0.922
	0.50	1.000	1.000	0.306	0.310	1.000	1.000	0.756	0.760	1.000	1.000	0.921	0.925
	0.75	1.000	1.000	0.312	0.336	1.000	1.000	0.756	0.756	1.000	1.000	0.924	0.916
	0.99	1.000	0.998	0.346	0.401	1.000	0.999	0.772	0.743	1.000	0.999	0.931	0.888

Table 6: As Tables 1 and 3, but the DGP is a stationary AR(1) process with an autoregressive coefficient a and a bilinear noise with coefficient b

g \ p	$p = 0$ $\Delta y_t = u_t$			$p = 1$ $\Delta y_t = -0.5\,\Delta y_{t-1} + u_t$			$p = 2$ $\Delta y_t = 0.6\,\Delta y_{t-1} + 0.08\,\Delta y_{t-2} + u_t$		
$g = 0$ no *AR*-filtering	*see* Table 2				$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$
				n = 100	0.014 (-16.564)	0.018 (-14.774)	n = 100	0.016 (-15.554)	0.020 (-13.903)
				n = 1000	0.028 (-10.232)	0.028 (-10.003)	n = 1000	0.030 (-9.268)	0.030 (-9.039)
$g = 1$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$
	n = 100	0.055 (1.000)	0.076 (5.120)	n = 100	0.051 (0.367)	0.060 (4.680)	n = 100	0.071 (9.590)	0.081 (14.361)
	n = 1000	0.051 (0.180)	0.054 (0.760)	n = 1000	0.050 (0.184)	0.052 (0.964)	n = 1000	0.062 (5.644)	0.064 (5.561)
$g = 2$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$
	n = 100	0.057 (1.460)	0.070 (3.920)	n = 100	0.045 (-2.294)	0.043 (-3.120)	n = 100	0.043 (-3.028)	0.042 (-3.625)
	n = 1000	0.051 (0.280)	0.053 (0.520)	n = 1000	0.051 (0.413)	0.048 (-1.009)	n = 1000	0.051 (0.367)	0.047 (-1.468)
$g = 5$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$		$b = 0.25$	$b = 0.99$
	n = 100	0.065 (2.920)	0.079 (5.700)	n = 100	0.029 (-9.498)	0.032 (-8.351)	n = 100	0.029 (-9.773)	0.030 (-9.993)
	n = 1000	0.052 (0.420)	0.051 (0.140)	n = 1000	0.048 (-0.918)	0.043 (-3.166)	n = 1000	0.048 (-0.964)	0.043 (-3.395)

Table 7: Rejection relative frequencies of the unit-root hypothesis H_0 for the Phillips-Perron Z_t test with AR-filtering of order g. The nominal significance level is 5%