journal of statistical planning and inference

Journal of Statistical Planning and Inference 68 (1998) 203–220

Geometric versus arithmetic random walk The case of trended variables[1]

E. Guerre[a,*,2], F. Jouneau[b,3]

[a] URA 1321-ISUP, BP 157, Université Paris 6, Tour 45-55 4, Place Jussieu, 75252 Paris Cedex 05, France

[b] CORE, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium

Received 30 January 1996; received in revised form 13 November 1996; accepted 9 December 1996

Abstract

The asymptotic behavior of the empirical means and variances for the geometric and arithmetic random walks are studied, when the underlying random walk is trended. Thus the effect of misspecifications can be described, and two tests are proposed. The first test uses a classical approach in model selection and is based on the comparison of estimated quasi likelihoods. The second one is obtained by estimating some nuisance parameters in a Neyman–Pearson test. Some bounds for the power functions are given, which suggest that the second test may be very powerful and better than the first one. © 1998 Elsevier Science B.V. All rights reserved.

Résumé

Le comportement asymptotique des moyennes et variances empiriques dans le cas de marches aléatoires géométriques et arithmétiques est obtenu, lorsque la marche aléatoire sous-jacente a une tendance. En conséquence, l'effet d'une mauvaise spécification peut être analysé, et deux tests sont proposés. Le premier test est d'inspiration classique en sélection de modèles, et est basé sur la comparaison des vraisemblances estimées. Le second test s'obtient en estimant des paramètres de nuisance dans un test de Neyman–Pearson. On donne des bornes pour la puissance qui suggèrent que le second test peut être très puissant, et est meilleur que le premier. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Box–Cox transformation; Integrated time series; Misspecification; Hypothesis testing; Model selection

AMS classifications: primary 62M10; secondary 62M02

* Corresponding author. E-mail: eguerre@ccr.jussieu.fr.

[1] The second author gratefully acknowledges CORE Human Capital grant.

[2] Also at CREST.

[3] Also at Université Lille III.

0378-3758/98/$19.00
PII S0378-3758(97)00142-0

1. Introduction

The need of nonlinear models in time-series analysis has been early recognized among applied statisticians. One of the first statistical nonlinear model can be found in the seminal book of Box and Jenkins (1970), where it is proposed to apply the popular Box–Cox transformation to linear time series. It is widely accepted that many nonstationary time series are not linear, and Box and Jenkins (1970) have proposed to transform the data before fitting an ARMA model (say) to the increments.

In many cases, the power parameter of the transformation is unknown and must be estimated. The literature dealing with the Box–Cox transformation under the so-called i.i.d. assumption is enormous, and have been surveyed in, for instance, Sakia (1992). This transformation has been applied in many fields, from macroeconomic modelling to biometrics studies.

In the context of time series, the model of interest is

$$h_{\lambda}(X_t) - h_{\lambda}(X_{t-1}) = \mu + \eta_t,$$

$X_t = 0$ for $t \leqslant 0$, where μ is a real number, $\{\eta_t\}_{t\in\mathbb{Z}}$ a centered stationary process, and $\{h_{\lambda}(.)\}_{\lambda\in\mathbb{R}}$ a family of power transformations. For instance, the popular Box–Cox transformation (see Box and Cox, 1964) is given by $h_{\lambda}(x) = (x^{\lambda} - 1)/\lambda$, $\lambda \neq 0$, $h_{\lambda}(x) = \ln x$ for $\lambda = 0$, but various transformations has been used in the literature (see Sakia, 1992). In general, the family of transformations is chosen in order to nest the logarithmic transformation ($\lambda = 0$) and identity ($\lambda = 1$). Assuming that the function h_{λ} is invertible, we get that

$$X_t = h_{\lambda}^{-1}\left(\mu t + \sum_{i=1}^{t} \eta_i\right). \tag{1}$$

Thus, the model is nonlinear because of the function h_{λ}^{-1}. The nonlinear structure of this model is theoretically simple to handle, since it is given by the power parameter λ. If $\{\eta_t\}_{t\in\mathbb{Z}}$ is a linear process, considering $h_{\lambda}(X_t)$ yields a integrated linear one. Hence, the nonlinearity (and thus its conditional heteroscedasticity) of the process is easily taken into account by a simple transformation of the variables. Moreover, the increments $\{\Delta h_{\lambda}(X_t)\}_{t\geqslant 0}$ are a linear stationary process.

This model has been widely used in the statistical literature: see classical books in time series as Box and Jenkins (1970), Brockwell and Davis (1987), the discussion following Chatfield and Prothero (1973a), the answer of the authors (Chatfield and Prothero, 1973b). Ansley et al. (1977) propose an algorithm for computing the maximum likelihood estimator. Forecasting models involving a power transformation has been widely investigated, see among others Granger and Newbold (1976), Hopwood et al. (1984), and Nelson and Granger (1979), for an empirical study. Guerrero (1993) provides some interesting ideas for identifying the unknown degrees of the underlying ARMA process and estimating this model. Burridge and Guerre (1996) have proposed a simple procedure to test if a process is a monotonous transformation of a random walk. Granger and Hallman (1991) consider various kind of nonlinear transformations.

The model (1) can also be interesting in financial and economical applications, since it is strongly linked to a particular class of diffusion processes. Indeed, assume that $X_s = g_\lambda(\mu s + \sigma W(s))$ where $\{W(s)\}_{s\in\mathbb{R}^+}$ is a standard Brownian motion, $g_\lambda = h_\lambda^{-1}$, and that $h_\lambda(x)$ is twice continuously differentiable with respect to x. This model is the continuous counterpart of the discrete-time model (1). The Ito formula yields that $\{X_s\}_{s\in\mathbb{R}^+}$ is a diffusion process such that

$$\mathrm{d}X_s = \left(\mu g_\lambda'(\mu t + \sigma W(s)) + \frac{\sigma^2}{2} g_\lambda''(\mu t + \sigma W(s))\right)\mathrm{d}s + \sigma g_\lambda'(\mu t + \sigma W(s))\,\mathrm{d}W(s)$$

$$= \frac{2\mu h_\lambda'(X_s)^2 - \sigma^2 h_\lambda''(X_s)}{2h_\lambda'(X_s)^3}\,\mathrm{d}s + \frac{\sigma}{h_\lambda'(X_s)}\,\mathrm{d}W(s).$$

The drift and volatility functions of this diffusion process are some (polynomial if $\lambda \neq 0$) functions depending on λ. The drawback of this model is that the same parameter is used both drift and volatility. Nevertheless, it covers the 'price in level' model $\mathrm{d}X_s = \mu\,\mathrm{d}s + \sigma\,\mathrm{d}W(s)$ ($\lambda = 1$) as well as the 'rate' model $\mathrm{d}X_s = \tilde{\mu}X_s\,\mathrm{d}s + \sigma X_s\,\mathrm{d}W(s)$ ($\lambda = 0$). Moreover, Marsh and Rosenfeld (1983) have proposed a similar class of diffusions in Finance.

In Economics, it is not very surprising that authors also tend to disagree on the transformations that should be used for integrated time series. Layson and Seak (1984) provide three examples in the empirical literatures for growth (see Andersen and Jordan, 1968; Carlson, 1978), short-run money demand (Hafer and Hein, 1980) and poverty (Thornton et al., 1978[4]).

In the financial literature, Dyl and Maberley (1986) seem to disagree with the usual rate formulation (emphasized among many others by Black and Scholes (1973) famous article) and proposed to consider price-level change to estimate hedge ratios (see also Black (1976), for a justification of the price-level formulation). It should be noticed that the rate formulation although popular is not very attractive for the study of pure arbitrage position, i.e., when the net acquisition price of the portfolio is zero. Clearly, the choice of the formulation is crucial for price estimations of derivative, hedge analysis, and so forth (see Park, 1991).

Despite all these interesting features, little is known on the statistical inference associated to (1). Guerre (1995a,b) suggest that the asymptotic behavior of estimators should be rather untypical and complicated, due to a degeneracy of the empirical information matrix. Moreover, many authors have focused only on particular values of λ, i.e. $\lambda = 0$ or $\lambda = 1$. Indeed, $\lambda = 0$ gives a geometric random walk and $\lambda = 1$ corresponds to the arithmetic one. This is the case for Dyl and Maberly (1986), who are interested in determining whether it is better to consider price in level or in rate. Layson and Seaks (1984) and Park (1991) both propose to test the null hypotheses $\lambda = 0$ or $\lambda = 1$.

[4] One could also go back to Malthus to trace the conflict between level and rate formulations for integrated time series, see Blaug (1979), for a presentation of Malthus' works.

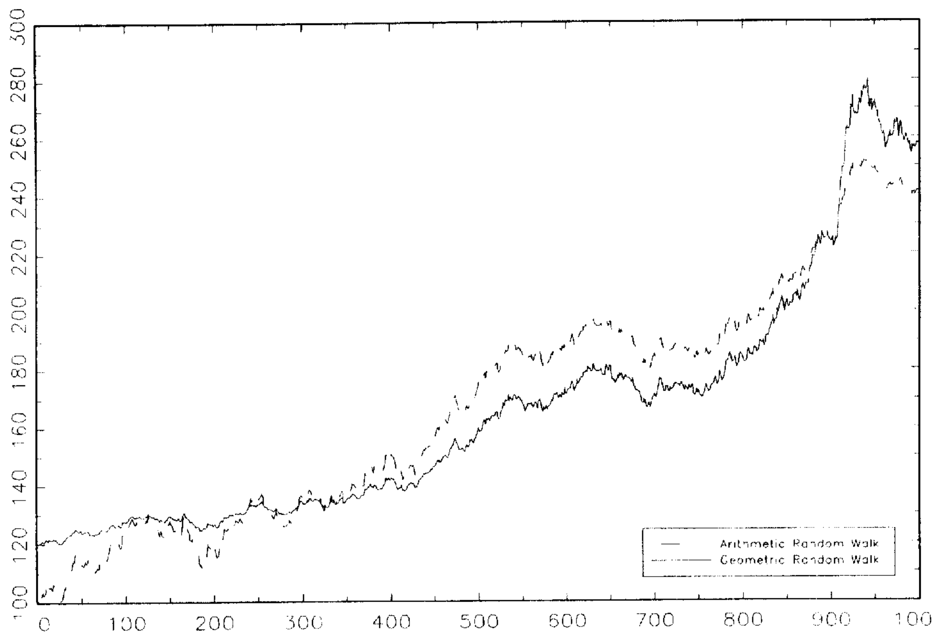

Fig. 1.

Considering the asymptotic behavior of the geometric ($X_t^g = X_0^g \exp(\mu_g t + \sigma_g \sum_{i=1}^t \varepsilon_i)$) and arithmetic ($X_t^a = X_0^a + \mu_a t + \sigma_a \sum_{i=1}^t \varepsilon_i$) random walks suggests that it should be easy to distinguish the two models at the sight of trajectories. This may be not possible as illustrated by Fig. 1, which gives such paths for the same realizations of i.i.d. $\mathcal{N}(0,1)$ ε_i's. This graphic has been obtained as follows. X_t^g, $t = 0, \dots, 1000$ is computed with $\mu_g = 0.002$, $\sigma_g = 0.015$. Then we use $\mu_a = \bar{X}^g$, $\sigma_a^2 = \sum_{t=0}^{1000} (X_t^g - \mu_a)^2/1001$ to compute the X_t^a's, using the same ε_t's as for the X_t^g's. The two trajectories exhibit similar features. It is also possible to obtain the same kind of graphics by proceeding symmetrically, i.e. computing first the arithmetic random walk.

This is why we propose here a formal statistical approach, to test a geometric random walk against an arithmetic one.[5] Our approach is based on quasi-(or pseudo) likelihood, using the Gaussian likelihood without assuming the process is a Gaussian one. Section 2 gives our basic definitions and notations. Because the quasi-likelihoods of our two models depend upon the empirical mean and variance computed under each assumption, and upon a Jacobian term, Section 3 is devoted to the asymptotic study of these statistics. As a consequence, it allows to evaluate the impact of a misspecification.

Section 4 deals with testing issues. The interesting point is that two approaches can be used in order to eliminate the nuisance parameters. The classical approach for model selection is to compare the quasi-likelihood computed with the estimated

[5] Actually, our framework also allows to test $\lambda = 0$ against $\lambda = \lambda_0$, λ_0 known.

parameters. Our second approach is inspired from the well-known Neyman–Pearson lemma, and is an attempt to eliminate parameters in critical region associated to the test of two simple hypotheses. We provide some bounds for the probability of misclassification which suggest that our second approach is probably better than the classical one. Model selection is briefly investigated in Section 5, and proofs are gathered in Section 6.

2. General framework and notation

Despite its simplicity, the Box–Cox model (1) may be difficult to estimate. Moreover, the geometric and arithmetic random walks are the easiest to interpret submodels of (1), and are two competing models which are frequently used in many applications, as explained in the introduction. Thus, we consider the two following models, $\{\varepsilon_t\}_{t\geqslant 1}$ being a strong white noise of variance 1:

$$\mathrm{H_a}=\mathrm{H_a}(\mu_\mathrm{a},\sigma_\mathrm{a}):\ X_t-X_{t-1}=\mu_\mathrm{a}+\sigma_\mathrm{a}\varepsilon_t,\ \ t>0,\qquad X_t=X_\mathrm{a}\ \text{ for } t\leqslant 0,$$

$$\mathrm{H_g}=\mathrm{H_g}(\mu_\mathrm{g},\sigma_\mathrm{g}):\ \ln X_t-\ln X_{t-1}=\mu_\mathrm{g}+\sigma_\mathrm{g}\varepsilon_t,\ \ t>0,\qquad X_t=X_\mathrm{g}\ \text{ for } t\leqslant 0,$$

with $\sigma_\mathrm{a}>0$, $\sigma_\mathrm{g}>0$, $\mu_\mathrm{g}>0$, $\mu_\mathrm{a}>0$, and where X_a and X_g are positive constants. $\mathbb{P}_\mathrm{a}$ and $\mathbb{P}_\mathrm{g}$ are the probability measures associated to $\mathrm{H_a}$ and $\mathrm{H_g}$, respectively. The model $\mathscr{A}^{\mathrm{tr}}=\bigcup_{\sigma_\mathrm{a}>0,\mu_\mathrm{a}>0}\mathrm{H_a}$ corresponds to arithmetic random walks,

$$X_t=\mu_\mathrm{a}t+\sigma_\mathrm{a}\sum_{i=1}^{t}\varepsilon_i+X_0.$$

X_t behaves like $\mu_\mathrm{a}t$ for large t. Similarly, the model $\mathscr{G}^{\mathrm{tr}}=\bigcup_{\sigma_\mathrm{g}>0,\mu_\mathrm{g}>0}\mathrm{H_g}$ contains geometric ones, with

$$X_t=X_0\exp\left(\mu_\mathrm{g}t+\sigma_\mathrm{g}\sum_{i=1}^{t}\varepsilon_i\right),$$

In X_t being equivalent to $\mu_\mathrm{g}t$. In both cases, the variables are upward trended since μ_a and μ_g are strictly positive. The density of ε_1 is denoted ϕ. Note that our framework is easily extended to the case where the power parameter of the Box–Cox transformation is known. Indeed, if this parameter is set equal to $\lambda_0>0$, $Y_t=X_t^{\lambda_0}$ is an arithmetic random walk.

It is further assumed that:

Assumption. (A1) $\mathbb{E}[\varepsilon_1]=0$, $\mathrm{Var}[\varepsilon_1]=1$ and there exists $\delta>0$ such that $\mathbb{E}|\varepsilon_1|^{2+\delta}<+\infty$.
(A2) There exists a positive constant Φ such that $\sup_{e\in\mathbb{R}}|\phi(e)|\leqslant\Phi$.

(A1) implies that, for any values of $(\mu_\mathrm{g},\sigma_\mathrm{g})$,

$$\mathbb{E}[|\log|1-\exp(-\mu_\mathrm{g}-\sigma_\mathrm{g}\varepsilon_1)||^{2+\delta}]<+\infty.$$

The quasi-loglikelihood functions of our two hypotheses are

$$\mathrm{H_a}\colon\ L_\mathrm{a}(\mu_\mathrm{a},\sigma_\mathrm{a}) = -\frac{T}{2}\ln 2\pi\sigma_\mathrm{a}^2 - \frac{1}{2\sigma_\mathrm{a}^2}\sum_{t=1}^{T}(X_t - X_{t-1} - \mu_\mathrm{a})^2,$$

$$\mathrm{H_g}\colon\ L_\mathrm{g}(\mu_\mathrm{g},\sigma_\mathrm{g}) = -\sum_{t=1}^{T}\ln|X_t| - \frac{T}{2}\ln 2\pi\sigma_\mathrm{g}^2 - \frac{1}{2\sigma_\mathrm{g}^2}\sum_{t=1}^{T}(\ln|X_t| - \ln|X_{t-1}| - \mu_\mathrm{g})^2,$$

$-\sum_{t=1}^{T}\ln|X_t|$ being a Jacobian term. These quasi-loglikelihood are computed under the assumption that $\{\varepsilon_t\}_{t\geqslant 1}$ is Gaussian. Note that L_g is computed using the absolute values of the X_t's, because, under $\mathscr{A}^{\mathrm{tr}}$, this process can be negative. Nevertheless, $\{X_t\}_{t\geqslant 1}$ is positive in finite time almost surely because $\mu_\mathrm{g} > 0$. Furthermore, other similar transformations than $\ln(\cdot)$ can be used, without modifying too much our results. In the sequel, the $\mathrm{LR} = \mathrm{LR}(\mathrm{H_g},\mathrm{H_a}) = \mathrm{LR}(\sigma_\mathrm{g},\mu_\mathrm{g},\sigma_\mathrm{a},\mu_\mathrm{a}) \equiv L_\mathrm{g} - L_\mathrm{a}$.

For any process Y_t, we denote its first difference process $\Delta Y_t = Y_t - Y_{t-1}$. We also define its empirical mean computed from T data by $\bar{Y} = \sum_{t=1}^{T} Y_t/T$. The two loglikelihood functions depend upon the sample via the normalized Jacobian term $\overline{\ln|X|}$ and the (quasi) maximum likelihood estimators of the parameters, which are

$$\hat{\mu}_\mathrm{a} = \frac{1}{T}\sum_{t=1}^{T}\Delta X_t = \frac{X_T - X_0}{T},$$

$$\hat{\sigma}_\mathrm{a}^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta X_t - \hat{\mu}_\mathrm{a})^2,$$

$$\hat{\mu}_\mathrm{g} = \frac{1}{T}\sum_{t=1}^{T}\Delta\ln|X_t| = \frac{\ln|X_T| - \ln|X_0|}{T},$$

$$\hat{\sigma}_\mathrm{g}^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta\ln|X_t| - \hat{\mu}_\mathrm{g})^2.$$

Expressions of the maximum of the loglikelihood functions may be derived straightforwardly, and are

$$\hat{L}_\mathrm{a} = L_\mathrm{a}(\hat{\mu}_\mathrm{a},\hat{\sigma}_\mathrm{a}) = -\frac{T}{2}(\ln 2\pi\hat{\sigma}_\mathrm{a}^2 + 1),$$

$$\hat{L}_\mathrm{g} = L_\mathrm{g}(\hat{\mu}_\mathrm{g},\hat{\sigma}_\mathrm{g}) = -\sum_{t=1}^{T}\ln|X_t| - \frac{T}{2}(\ln 2\pi\hat{\sigma}_\mathrm{g}^2 + 1).$$

3. The asymptotic behavior of the estimators

We study the behavior of the estimators and of the Jacobian term under $\mathscr{G}^{\mathrm{tr}}$ and $\mathscr{A}^{\mathrm{tr}}$. The next theorem is the key result of this paper, and can actually be proven under the weaker assumptions that the Donsker theorem holds for the partial sum of the ε_t's, as some appropriate laws of large numbers.

Theorem 1. *Let* $\{W(s)\}_{s\in[0,1]}$ *be a standard Brownian motion. Under* (A1), *we have*:
(i) *Under* $\mathrm{H_g}\in\mathscr{G}^{\mathrm{tr}}$,

$$((\overline{\ln|X|}-\hat{\mu}_{\mathrm{g}}(T+1)/2)/\sqrt{T},\hat{\sigma}_{\mathrm{g}}^2,\sqrt{T}(\hat{\mu}_{\mathrm{g}}-\mu),(\ln\hat{\sigma}_{\mathrm{a}}^2-2\hat{\mu}_{\mathrm{g}}T)/\sqrt{T},$$

$$(\ln\hat{\mu}_{\mathrm{a}}-\hat{\mu}_{\mathrm{g}}T)/\sqrt{T}),$$

converges in distribution to

$$\left(\sigma_{\mathrm{g}}\int_0^1 W(r)\,\mathrm{d}r-\sigma_{\mathrm{g}}W(1)/2,\sigma_{\mathrm{g}}^2,2\sigma_{\mathrm{g}}W(1),0,0\right).$$

(ii) *Under* $\mathrm{H_a}\in\mathscr{A}^{\mathrm{tr}}$,

$$(\overline{\ln|X/T|},T\hat{\sigma}_{\mathrm{g}}^2,T\hat{\mu}_{\mathrm{g}}-\ln T+\ln X_0,\hat{\sigma}_{\mathrm{a}}^2,\hat{\mu}_{\mathrm{a}}),$$

converges almost surely to

$$\left(\ln\mu_{\mathrm{a}}-1,\sum_{t=1}^{+\infty}\ln^2\left|\frac{X_t}{X_{t-1}}\right|,\ln\mu_{\mathrm{a}},\sigma_{\mathrm{a}}^2,\mu_{\mathrm{a}}\right).$$

Proof. See Section 6.

Under $\mathscr{G}^{\mathrm{tr}}$, $\hat{\sigma}_{\mathrm{a}}^2$ and $\hat{\mu}_{\mathrm{a}}$ diverge at an exponential rate. Actually, the proof of this theorem shows that $\ln\hat{\sigma}_{\mathrm{a}}^2$ is close to $2\ln\sup_{1\leqslant t\leqslant T}X_t$, which is $2\ln X_T=2T\hat{\mu}_{\mathrm{g}}+\ln X_0$ due to the trend. More precisely, we have $\ln\hat{\sigma}_{\mathrm{a}}^2=2\ln X_T+\mathrm{o}_{\mathbb{P}_{\mathrm{g}}}(\sqrt{T})$. $\hat{\sigma}_{\mathrm{a}}^2$ diverges because the observations are not transformed via the logarithm function, and then differentiation is not strong enough to reduce the variability of the process. On the other hand, $\hat{\sigma}_{\mathrm{g}}^2$ and $\hat{\mu}_{\mathrm{g}}$ goes to 0 under $\mathscr{A}^{\mathrm{tr}}$, because the logarithmic transformation 'oversmooths' the sample, and thus eliminates the variability. From a statistical viewpoint, this suggests that the geometric random walk which asymptotically maximizes $L_{\mathrm{g}}(\mu_{\mathrm{g}},\sigma_{\mathrm{g}})$ under $\mathscr{A}^{\mathrm{tr}}$ is degenerate (i.e. $\mu_{\mathrm{g}}=0$, $\sigma_{\mathrm{g}}=0$), which may not be surprising at the sight of the trajectories of the two models.

Theorem 1 is easy to extend to the case where the power parameter is known to be $\lambda_0>0$, that is changing $\mathscr{A}^{\mathrm{tr}}$ into:

$\mathscr{A}^{\mathrm{tr}}(\lambda_0)$: $Y_t=X_t^{\lambda_0}$ is an arithmetic random walk, i.e.

$$X_t^{\lambda_0}-X_{t-1}^{\lambda_0}=\mu_{\mathrm{a}}+\sigma_{\mathrm{a}}\varepsilon_t,\quad X_t^{\lambda_0}=X_{\mathrm{a}}^{\lambda_0}\ \text{for}\ t\leqslant 0.$$

In this case, the new mean and variance estimates are

$$\hat{\mu}_{\mathrm{a}}(\lambda_0)=\frac{X_t^{\lambda_0}-X_0^{\lambda_0}}{T},\qquad \hat{\sigma}_{\mathrm{a}}^2(\lambda_0)=\frac{1}{T}\sum_{t=1}^{T}(\Delta X_t^{\lambda_0}-\hat{\mu}_{\mathrm{a}}(\lambda_0))^2.$$

Under $\mathscr{G}^{\mathrm{tr}}$, the behavior of $(\hat{\sigma}_{\mathrm{a}}^2(\lambda_0),\hat{\mu}_{\mathrm{a}}(\lambda_0))$ is close to the one of $(\hat{\sigma}_{\mathrm{a}}^2,\hat{\mu}_{\mathrm{a}})$, because $\ln X_t=(\ln Y_t)/\lambda_0$. For instance, $\ln\hat{\sigma}_{\mathrm{a}}^2(\lambda_0)$ will be close to $2\lambda_0\ln X_T$, as explained above.

Under $\mathscr{A}^{\rm tr}(\lambda_0)$, the asymptotic behavior of $(\hat{\sigma}_{\rm g}^2, \hat{\mu}_{\rm g})$ is easily deduced from the one of the same statistic under $\mathscr{A}^{\rm tr}$, and this empirical mean and variance also vanish asymptotically.

The behavior of these statistics also extends to the case of dependent observations, as long as some appropriate invariance principle holds. The variances appearing in non-degenerate distributions in Theorem 1 are the same up to a factor $\lim_{T\to+\infty} \mathrm{Var}(\sum_{t=1}^{T} \varepsilon_t)/T$ (if strictly larger than 0). This correction can be estimated using a nonparametric estimator of the spectral density at 0.

4. Testing procedures

Theorem 1 allows to study various statistics, on which asymptotic α-level tests of $\mathscr{G}^{\rm tr}$ against $\mathscr{A}^{\rm tr}$ can be based. A popular approach is nonnested model testing relies on the ratio of the two maximum likelihoods. According to Theorem 1, this statistic has to be modified in order to reach asymptotic convergence under $\mathscr{G}^{\rm tr}$, and we consider

$$\widehat{\mathrm{LR}} = \widehat{\mathrm{LR}}(\hat{\sigma}_{\rm g}^2, \hat{\mu}_{\rm g}, \hat{\sigma}_{\rm a}) = \ln \hat{\sigma}_{\rm a} - \hat{\mu}_{\rm g} T - \ln \hat{\sigma}_{\rm g} - \overline{\ln|X|} + \hat{\mu}_{\rm g}(T+1)/2$$
$$= \ln \hat{\sigma}_{\rm a} - \ln \hat{\sigma}_{\rm g} - \overline{\ln|X|} - \hat{\mu}_{\rm g}(T-1)/2.$$

Theorem 1 shows that, under $\mathscr{G}^{\rm tr}$, $\widehat{\mathrm{LR}}/(\hat{\sigma}_{\rm g}\sqrt{T})$ converges in distribution to

$$-\int_0^1 W(r)\,\mathrm{d}r + W(1)/2,$$

which is a $\mathcal{N}(0, \frac{1}{12})$ random variable. Indeed, Theorem 1 shows that

$$\frac{\ln \hat{\sigma}_{\rm a} - \hat{\mu}_{\rm g} T}{\hat{\sigma}_{\rm g}\sqrt{T}}, \quad \frac{\ln \hat{\sigma}_{\rm g}}{\hat{\sigma}_{\rm g}\sqrt{T}},$$

go to 0 in probability. Thus, the asymptotic distribution of interest is the one of

$$-\frac{\overline{\ln|X|} - \hat{\mu}_{\rm g}(T+1)/2}{\hat{\sigma}_{\rm g}\sqrt{T}}.$$

which, by Theorem 1, is the one given above.

It is shown below that $\widehat{\mathrm{LR}}$ diverges to $-\infty$ under $\mathscr{A}^{\rm tr}$. Thus, we shall consider the test which rejects $\mathscr{G}^{\rm tr}$ if $\widehat{\mathrm{LR}}$ is small, i.e. if the data are in the critical region:

$$\widehat{\mathscr{LR}} = \widehat{\mathscr{LR}}(\alpha, \hat{\sigma}_{\rm g}, \hat{\mu}_{\rm g}, \hat{\sigma}_{\rm a}) = \{\widehat{\mathrm{LR}} < \sqrt{T/12}\,\hat{\sigma}_{\rm g} c_\alpha\},$$

$\mathbb{P}(\mathcal{N}(0,1) < c_\alpha) = \alpha$.

Considering the problem of testing the simple hypothesis $\mathrm{H_g}$ against $\mathrm{H_a}$ under Gaussianity suggests to introduce a Neyman–Pearson test, which has well-known optimality property. Neglecting possible negative values in the sample, this test is based upon the statistic $\mathrm{LR}(\sigma_g, \mu_g, \sigma_a, \mu_a)/T = (L_g - L_a)/T$, i.e.

$$\frac{1}{2}\ln\sigma_a^2 + \frac{1}{2\sigma_a^2}(\hat\sigma_a^2 + (\hat\mu_a - \mu_a)^2) - \overline{\ln|X|} - \frac{1}{2}\ln\sigma_g^2 - \frac{1}{2\sigma_g^2}(\hat\sigma_g^2 + (\hat\mu_g - \mu_g)^2).$$

Such tests typically reject $\mathscr{G}^{\mathrm{tr}}$ if LR/T is small. Because the critical region depends upon the parameters in a complicated manner, it is unlikely to find an optimal test. Eliminating the parameters using the estimators leads to a maximum likelihood ratio test, and we propose a test based upon the comparison of the leading terms of LR/T under both $\mathscr{G}^{\mathrm{tr}}$ and $\mathscr{A}^{\mathrm{tr}}$.

Under $\mathscr{G}^{\mathrm{tr}}$, $\ln\hat\sigma_a^2$ is close to $2\ln X_T = 2\mu_g T + 2\sigma_g \sum_{t=1}^{T} \varepsilon_t + \ln X_0$, and the details of the proof of Theorem 1 show that $\hat\sigma_a^2$ dominates $\hat\mu_a^2$. Thus, the leading term in LR/T is $\hat\sigma_a^2/(2\sigma_a^2)$ under $\mathscr{G}^{\mathrm{tr}}$. Under $\mathscr{A}^{\mathrm{tr}}$, LR/T is equivalent to $-\overline{\ln|X|}$ (itself equivalent to $-\ln T$). Thus, the approximation of $\ln\hat\sigma_a^2$ above suggests the rejection region

$$\left\{ \frac{\hat\sigma_a^2/(2\sigma_a^2) - \overline{\ln|X|}}{\exp(2\mu_g T)} < \frac{\exp(2\sigma_g\sqrt{T}c_\alpha)}{2\sigma_a^2} \right\}. \tag{2}$$

This critical region defines an asymptotic α-level test of $\mathrm{H_g}$ versus $\mathrm{H_a}$. At this stage, several problems arise due to the unknown parameters μ_g and σ_a. Because σ_a is a nuisance parameter under $\mathscr{G}^{\mathrm{tr}}$, we replace $2\sigma_a^2$ by 1 for the sake of simplicity.[6] Second, we cannot change μ_g into $\hat\mu_g$, because Theorem 1 gives that $\hat\sigma_a^2/\exp(2\hat\mu_g T)$ converges to 1 under $\mathscr{G}^{\mathrm{tr}}$, and the LHS of the equality would have a degenerate asymptotic distribution. We use instead a suboptimal estimator of the mean based on the first $T_p = [pT]$ observations of the sample ($p \in]0,1[$), namely $\hat\mu_g(p) = (\ln|X_{T_p}|)/T_p$. We easily obtain that, under $\mathscr{G}^{\mathrm{tr}}$, $\sqrt{T}(\hat\mu_g(p) - \mu) = \sigma_g\sqrt{T}\sum_{t=1}^{T_p} \varepsilon_t/T_p$ converges in distribution to $\sigma_g W(p)/p$, jointly with the statistics of the part (i) in Theorem 1. Thus, $(\ln\hat\sigma_a^2 - 2\hat\mu_g(p)T)/2\sqrt{T}$ converges in distribution to $\sigma_g(W(1) - W(p)/p)$, which is a $\mathscr{N}(0, 1/p - 1)$-random variable. This leads to consider the statistic

$$\widehat{\mathrm{NP}} = \mathrm{NP}(\hat\sigma_a, \hat\mu_g(p)) = (\hat\sigma_a^2 - \overline{\ln|X|})/\exp(2\hat\mu_g(p)T),$$

and the critical region

$$\widehat{\mathscr{NP}} = \mathscr{NP}(\alpha, \hat\sigma_g, \hat\mu_g(p), \hat\sigma_a) = \left\{ \widehat{\mathrm{NP}} < \exp\left(2\sqrt{T\left(\frac{1}{p} - 1\right)}\hat\sigma_g c_\alpha\right)\right\}.$$

The next corollary states the consistency properties of these two tests.

Corollary 1. *Under* (A2), *the tests having for critical regions* $\widehat{\mathscr{LR}}$ *and* $\widehat{\mathscr{NP}}$ *are asymptotically of level* α *and consistent for testing* $\mathscr{G}^{\mathrm{tr}}$ *against* $\mathscr{A}^{\mathrm{tr}}$.

[6] If it is known that $\sigma_a \geqslant \sigma_a^- > 0$ under $\mathscr{A}^{\mathrm{tr}}$, it may be more relevant from a statistical viewpoint to change σ_a into σ_a^-. σ_a^- can also be chosen depending upon T.

Proof. The fact that $\widehat{\mathscr{LR}}$ is asymptotically of level α has been proven above. Under $\mathscr{G}^{\mathrm{tr}}$, $(\hat{\sigma}_{\mathrm{g}}, \hat{\mu}_{\mathrm{g}}(p))$ converges almost surely to $(\sigma_{\mathrm{g}}, \mu_{\mathrm{g}})$, and Theorem 1 yields that $\overline{\ln|X|}/\exp(2\hat{\mu}_{\mathrm{g}}(p)T)$ asymptotically vanishes, since $\mu_{\mathrm{g}}>0$. Thus

$$\lim_{T\to+\infty} \mathbb{P}_{\mathrm{g}}(\widehat{\mathscr{NP}}) = \lim_{T\to+\infty} \mathbb{P}_{\mathrm{g}}\left(\frac{\ln\hat{\sigma}_{\mathrm{a}}^2 - 2\hat{\mu}_{\mathrm{g}}(p)T}{2\sigma_{\mathrm{g}}\sqrt{T(1/p-1)}} < c_\alpha\right) = \alpha,$$

by Theorem 1.

We now deal with the consistency part. Recall that $\sqrt{T}\hat{\sigma}_{\mathrm{g}}$ converges to a positive bounded random variable under $\mathscr{A}^{\mathrm{tr}}$. Thus, it is enough to show that $\widehat{\mathrm{LR}}$ diverges to $-\infty$ and that $\widehat{\mathrm{NP}}$ is a negative real number with a probability tending to 1. We have

$$\widehat{\mathrm{LR}} = \ln\hat{\sigma}_{\mathrm{a}} - \ln(\sqrt{T}\hat{\sigma}_{\mathrm{g}}) + \frac{\ln T}{2} - \overline{\ln|X/T|} - \ln T$$
$$- \left(\frac{T-1}{2}\hat{\mu}_{\mathrm{g}} - \frac{\ln T + \ln|X_0|}{2}\right) - \frac{\ln T + \ln|X_0|}{2}, \tag{3}$$

and Theorem 1 gives that $\widehat{\mathrm{LR}}$ is equivalent to $-\ln T$. Similarly, $pT\hat{\mu}_{\mathrm{g}}(p) - \ln T$ converges almost surely to $\ln(p\mu_{\mathrm{a}})$, and

$$\widehat{\mathrm{NP}} = \frac{\hat{\sigma}_{\mathrm{a}}^2 - \overline{\ln|X/T|} - \ln T}{T^2\exp(2pT\hat{\mu}_{\mathrm{g}}(p) - 2\ln T)}, \tag{4}$$

and is equivalent to $(-\ln T)/(Tp\mu_{\mathrm{a}})$, which is negative since $\mu_{\mathrm{a}}>0$. □

4.1. *Study of the power*

The Neyman–Pearson test is a most powerful test for $\mathrm{H_g}$ against $\mathrm{H_a}$ assuming $\{\varepsilon_t\}_{t\geqslant 1}$ is a Gaussian sequence. It asymptotically rejects the geometric random walk if the observations are in the critical region given by (2). Theorem 1 shows that an appropriate normalization under $\mathscr{A}^{\mathrm{tr}}$ for the statistics in (2) leads to rewrite this set as

$$\hat{\sigma}_{\mathrm{a}}^2/(2\sigma_{\mathrm{a}}^2) - \overline{\ln|X/T|} < \exp(2\mu_{\mathrm{g}}T + 2\sigma_{\mathrm{g}}\sqrt{T}c_\alpha) + 2\sigma_{\mathrm{a}}^2\ln T.$$

Considering the limit of $\hat{\mu}_{\mathrm{g}}(p)$ in (4) shows that our $\widehat{\mathscr{NP}}$ test asymptotically rejects $\mathscr{G}^{\mathrm{tr}}$ if

$$\hat{\sigma}_{\mathrm{a}}^2 - \overline{\ln|X/T|} < T^2 p\mu_{\mathrm{a}}\exp(2\hat{\sigma}_{\mathrm{g}}\sqrt{(1/p-1)T}c_\alpha) + \ln T,$$

$\hat{\sigma}_{\mathrm{g}}\sqrt{T}$ being a bounded (uniformly in T) random variable. Thus, the $\widehat{\mathscr{NP}}$ test suffers from a dramatic loss of power compared to the test defined by (2), due to the strong difference of magnitude between the RHS of the two inequalities above. The reason is that under $\mathscr{A}^{\mathrm{tr}}$ the mean estimator $\hat{\mu}_{\mathrm{g}}(p)$ goes to 0. In order to cope with this, we introduce a new mean estimator, which is bounded away from 0. For instance, we can assume that it is known that $\mu_g > \underline{\mu} > 0$ under $\mathscr{G}^{tr}$, which leads to consider the new mean estimators

$$\tilde{\mu}_g = \hat{\mu}_g \mathbb{1}(\hat{\mu}_g > \underline{\mu}) + \underline{\mu}\mathbb{1}(\hat{\mu}_g \leqslant \underline{\mu}), \qquad \tilde{\mu}_g(p) = \hat{\mu}_g(p)\mathbb{1}(\hat{\mu}_g(p) > \underline{\mu}) + \underline{\mu}\mathbb{1}(\hat{\mu}_g(p) \leqslant \underline{\mu}).$$

We have $\tilde{\mu}_g(p) > \underline{\mu}$, and $\underline{\mu}$ can also be a sequence which goes to 0 with T. This new mean estimator $\tilde{\mu}_g(p)$ defines a new critical region, say $\widetilde{\mathscr{N}\mathscr{P}}$, for which we have the inclusion

$$\{\hat{\sigma}_a^2 - \overline{\ln |X/T|} < \exp(2\underline{\mu}T + 2\hat{\sigma}_g\sqrt{(1/p-1)T}c_\alpha) + \ln T\} \subset \widetilde{\mathscr{N}\mathscr{P}}. \tag{5}$$

This shows that the $\widetilde{\mathscr{N}\mathscr{P}}$ test is more powerful than $\widehat{\mathscr{N}\mathscr{P}}$ (asymptotically at least). The gain suggests that the loss of power is the consequence of a failure to take into account the additional information that $\mu_g > \underline{\mu}$ under $\mathscr{G}^{tr}$.

The proof of Corollary 1 also shows that the power of the $\widehat{\mathscr{L}\mathscr{R}}$ test can be weak. Indeed, (3) shows that under $\mathscr{A}^{tr}$, $\widehat{\mathrm{LR}}$ is equivalent to $-\ln T$, and then the $\widehat{\mathscr{L}\mathscr{R}}$ test asymptotically compares $-\ln T$ to $\sqrt{1/12}\sqrt{\sum_{t=1}^{T} \ln^2 |X_t/X_{t-1}|}c_\alpha$. For usual values of α, c_α is negative, and it may arise that the former random variable has the same order of magnitude, in which case the test may accept the hypothesis of a geometric random walk. Indeed, the distribution of the previous random variable depends upon the distribution of the ε_t's and we have not been able to obtain adequate bounds for its tails in full generality. The item $\ln^2 |X_t/X_{t-1}|$ can be large since it does not have an exponential moment (because of the strong divergence of the integral at $X_{t-1} = 0$). A simple way to avoid the large values of $|\sqrt{T/12}\hat{\sigma}_g c_\alpha|$ is to restrict the parameter set under $\mathscr{G}^{tr}$, that is to assume $\sigma_g < \bar{\sigma} (\bar{\sigma} > 0)$. In order to bound $\ln \hat{\sigma}_g$ from below, we also assume $\sigma_g > \underline{\sigma} > 0$. This leads to consider the new variance estimator

$$\tilde{\sigma}_g = \hat{\sigma}_g \mathbb{1}(\underline{\sigma} < \hat{\sigma}_g < \bar{\sigma}) + \bar{\sigma}\mathbb{1}(\hat{\sigma}_g \geqslant \bar{\sigma}) + \underline{\sigma}\mathbb{1}(\hat{\sigma}_g \leqslant \underline{\sigma}).$$

As before, $(\underline{\sigma}, \bar{\sigma})$ can depend on T. This gives for $\alpha < 50\%$, $\sqrt{T}\tilde{\sigma}_g c_\alpha \geqslant \sqrt{T}\bar{\sigma}c_\alpha$. This is not sufficient for our purpose since we asymptotically compare $-\ln T$ to $\sqrt{T/12}\bar{\sigma}c_\alpha$, and it is necessary to modify our statistic in order to achieve a higher rate of divergence. This is easily done changing $\hat{\mu}_g$ into $\tilde{\mu}_g$ in $\widehat{\mathrm{LR}}$, since (3) shows that the redefined $\widehat{\mathrm{LR}}$ diverges to $-\infty$ at a rate which is $-\underline{\mu}T/2$.

Let us define $\widetilde{\mathrm{LR}}$, $\widetilde{\mathscr{L}\mathscr{R}}$, $\widetilde{\mathrm{NP}}$, $\widetilde{\mathscr{N}\mathscr{P}}$ by changing $\hat{\sigma}_g$, $\hat{\mu}_g$ and $\hat{\mu}_g(p)$ into $\tilde{\sigma}_g$, $\tilde{\mu}_g$ and $\tilde{\mu}_g(p)$ in the definition of $\widehat{\mathrm{LR}}, \widehat{\mathscr{L}\mathscr{R}}, \widehat{\mathrm{NP}}$, and $\widehat{\mathscr{N}\mathscr{P}}$. Because $(\tilde{\sigma}_g, \tilde{\mu}_g, \tilde{\mu}_g(p))$ and $(\hat{\sigma}_g, \hat{\mu}_g, \hat{\mu}_g(p))$ have the same asymptotic behavior under $\mathscr{G}^{tr}$ if $\mu_g > \underline{\mu}$ and $\underline{\sigma} < \sigma_g < \bar{\sigma}$, $\widetilde{\mathscr{L}\mathscr{R}}$ and $\widetilde{\mathscr{N}\mathscr{P}}$ are also asymptotic α-level tests of $\mathscr{G}^{tr}$ against $\mathscr{A}^{tr}$. The following theorem studies the power of these new tests, $\widetilde{\mathscr{L}\mathscr{R}}^c$ and $\widetilde{NP}^c$ being the acceptance region to the geometric-random walk.

Theorem 2. *If $\mu_g > \underline{\mu}$, $\underline{\sigma} < \sigma_g < \bar{\sigma}$ under $\mathscr{G}^{tr}$, and* (A1) *holds, then $\widetilde{\mathscr{L}\mathscr{R}}$ and $\widetilde{\mathscr{N}\mathscr{P}}$ are asymptotic α-level tests of $\mathscr{G}^{tr}$ against $\mathscr{A}^{tr}$. Moreover, if* (A2) *holds, and $\alpha \leqslant 50\%$, there exists a positive constant C depending upon the distribution of the ε_t's such* *that, for $H_a \in \mathscr{A}^{tr}$, for any $q \in [0,1]$:*

$$\mathbb{P}_a(\widetilde{\mathscr{L}\mathscr{R}}^c) \leqslant \mathbb{P}_a(\hat{\sigma}_a^2 \geqslant \underline{\sigma}^2 \exp(q\underline{\mu}(T-1) + 2q\sqrt{T/12}\bar{\sigma}c_\alpha))$$

$$+ \frac{C\sqrt{T}}{\sigma_a} \exp\left(-(1-q)\underline{\mu}\frac{T-1}{2} - (1-q)\sqrt{T/12}\bar{\sigma}c_\alpha\right),$$

$$\mathbb{P}_a(\widetilde{\mathscr{N}\mathscr{P}}^c) \leqslant \mathbb{P}_a(\hat{\sigma}_a^2 \geqslant q \exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_\alpha))$$

$$+ \frac{C\sqrt{T}}{\sigma_a} \exp(-(1-q)\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_\alpha)).$$

Proof. See Section 6.

The number q has been introduced to optimally balance the two terms in the bounds (if possible). As mentioned above, $\bar{\sigma}$, $\underline{\sigma}$ and $\underline{\mu}$ can be some sequences depending on T.

Because $c_\alpha < 0$, the bounds in Theorem 2 suggest that the tests are not powerful if $\underline{\mu}$ is close to 0, that is if the null hypothesis contains a geometric random walk without trend. More precisely, $|c_\alpha|\bar{\sigma}\sqrt{T}/(\underline{\mu}T)$ must be small to have good tests. The case of random walks without trend is left for further studies.

The bounds on the probabilities of errors depends upon the probability that $\hat{\sigma}_a^2$ exceeds a large real number. Under A1, it goes to 0 as an exponential function of T. Assuming that $\{\varepsilon_t\}_{t\geqslant 1}$ is Gaussian gives that $\hat{\sigma}_a^2$ is $\sigma_a^2\chi^2(T-1)/T$, and thus

$$\mathbb{P}_a(\hat{\sigma}_a^2 \geqslant x) \leqslant \mathbb{E}_a[\exp(\hat{\sigma}_a^2)]\exp(-x),$$

$\mathbb{E}_a \exp(\hat{\sigma}_a^2)$ being bounded independently of $T>2$. In this case, the two terms in the bound of $\mathbb{P}_a(\widetilde{\mathscr{N}\mathscr{P}}^c)$ have a similar order of magnitude and goes to 0 very fast due to the $\exp(-\exp(.))$ function, reflecting the strong statistical difference between the arithmetic and geometric random walks.

Thus, the bound for $\mathbb{P}_a(\widetilde{\mathscr{N}\mathscr{P}}^c)$ is likely to go to 0 faster than the one for $\mathbb{P}_a(\widetilde{\mathscr{L}\mathscr{R}}^c)$, suggesting that the $\widetilde{\mathscr{N}\mathscr{P}}$ test is better. The reason comes from the form of the regions of acceptance, for which the following inclusions hold:

$$\widetilde{\mathscr{L}\mathscr{R}}^c = \left\{\ln\hat{\sigma}_a - \ln\tilde{\sigma}_g - \overline{\ln|X|} \geqslant \tilde{\mu}_g(p)\frac{T-1}{2} + \sqrt{T/12}\tilde{\sigma}_g c_\alpha\right\}$$

$$\subset \left\{\ln\hat{\sigma}_a - \overline{\ln|X|} \geqslant \underline{\mu}\frac{T-1}{2} + \sqrt{T/12}\bar{\sigma}c_\alpha + \ln\underline{\sigma}\right\} \tag{6}$$

$$\widetilde{\mathscr{N}\mathscr{P}}^c = \{\hat{\sigma}_a^2 - \overline{\ln|X|} \geqslant \exp(2\tilde{\mu}_g(p)T + 2\sqrt{T(1/p-1)}\tilde{\sigma}_g c_\alpha)\}$$

$$\subset \{\hat{\sigma}_a^2 - \overline{\ln|X|} \geqslant \exp(2\underline{\mu}T + 2\sqrt{T(1/p-1)}\bar{\sigma}c_\alpha)\}, \tag{7}$$

which are obtained using inequalities similar to (3) and (5), by definition of $\tilde{\sigma}_g$ and $\tilde{\mu}_g(p)$ and since $c_\alpha < 0$. The proof of the theorem works by estimating the probability that $\hat{\sigma}_a^2$ and $-\overline{\ln|X|}$ exceed quantities close to the RHS's of the inequalities above. Since the RHS of the $\widetilde{NP}$ test is the highest, the probability of exceedance of $-\overline{\ln|X|}$ is smaller, and we obtain a bound for $\mathbb{P}_a(\widetilde{\mathcal{N}\mathcal{P}}^c)$ which decreases at an extremely fast rate compared with the bound of $\mathbb{P}_a(\widetilde{\mathcal{L}\mathcal{R}}^c)$. Note that a similar bound can be obtained for the Neyman–Pearson-type test (2) of $H_g(\underline{\mu}, \bar{\sigma})$ against $H_a(\mu_a, 1/\sqrt{2})$: This suggests that the $\widetilde{\mathcal{N}\mathcal{P}}$ test should enjoy some asymptotic optimality property, at least in terms of rate of convergence of the power function. Of course, accepting $\mathcal{A}^{tr}$ if there are negative observations in the sample improves the $\widetilde{\mathcal{N}\mathcal{P}}$ test.

5. Model choice procedures

A model choice procedure is a decision rule which decides if the data are generated by a geometric or an arithmetic random walk. It is consistent if the probabilities of error, under $\mathcal{A}^{tr}$ and $\mathcal{G}^{tr}$ goes to 0 with T. This problem may be viewed as an estimation problem, with a new discrete parameter lying in $\{\mathcal{A}^{tr}, \mathcal{G}^{tr}\}$.

A popular procedure is the 'goodness of fit' approach, which selects the model with the smallest variance. This does not work here, because, under $\mathcal{A}^{tr}$, $\hat{\sigma}_a$ converges to σ_a and $\hat{\sigma}_g$ goes to 0. Since the model choice problem is also an estimation problem the decision rule can be based on the likelihood functions: one can prefer the model with the highest likelihood, that is $\mathcal{A}^{tr}$ if $\hat{L}_a > \hat{L}_g$ and $\mathcal{G}^{tr}$ if $\hat{L}_g > \hat{L}_a$. Since

$$\hat{L}_g - \hat{L}_a = T(\ln\hat{\sigma}_a - \ln\hat{\sigma}_g - \overline{\ln|X|}),$$

this approach corrects the comparison of variance criterium (based on $\ln\hat{\sigma}_a - \ln\hat{\sigma}_g$) by taking into account the Jacobian term $\overline{\ln|X|}$.

Corollary 2. *Under* (A1), *choosing the model which has the highest likelikhood is a consistent model choice procedure.*

Proof. Under $\mathcal{G}^{tr}$, $(\hat{L}_g - \hat{L}_a)/(T\sqrt{T})$ is equivalent to $\mu_g T > 0$, by virtue of Theorem 1. Under $\mathcal{A}^{tr}$:

$$(\hat{L}_g - \hat{L}_a)/T = \ln\hat{\sigma}_a - \ln(\sqrt{T}\hat{\sigma}_g) + \tfrac{1}{2}\ln T - \overline{\ln|X/T|} - \ln T,$$

and is equivalent to $(-\ln T)/2$. □

Thus, correcting $\ln\hat{\sigma}_a - \ln\hat{\sigma}_g$ with the Jacobian terms gives a consistent procedure. The same modifications than the one leading to Theorem 2 allow us to bound the probability of error under $\mathcal{A}^{tr}$. The study of the probability of error under $\mathcal{G}^{tr}$ is more delicate. Nevertheless, model choice procedures can be based on the $\widetilde{\mathcal{L}\mathcal{R}}$ and $\widetilde{\mathcal{N}\mathcal{P}}$ tests by taking α going to 0 with T. The results of Theorem 2 can be useful if $(\bar{\sigma}c_\alpha)/\underline{\mu}\sqrt{T})$ goes to 0.

6. Proofs section

6.1. Proof of Theorem 1

Let $[.]$ be the integer part and define $\{E_T(r)\}_{r\in[0,1]}$ as the linear interpolation of the normalized partial sums up to T of the ε_t's i.e.

$$E_T(r) \equiv \frac{\sum_{t=1}^{[Tr]} \varepsilon_t + (Tr - [Tr])\varepsilon_{[Tr]+1}}{\sqrt{T}}.$$

The Donsker theorem (see Billingsley, 1968, p. 68) says that $\{E_T(r)\}_{r\in[0,1]}$ converges weakly to $\{W(r)\}_{r\in[0,1]}$ in $C[0,1]$ equipped with the supremum norm $\|.\|_\infty$. Moreover, by the Skohorod device, the variables can be reconstructed for each T on a rich enough probability space such that the convergence holds almost surely for this norm.

6.1.1. Geometric random walk with trend

The estimator $\hat{\sigma}_{\mathrm{g}}^2$ converges almost surely to σ_{g}^2 by (A1) and the law of large numbers. Because X_T goes to $+\infty$, we have

$$\ln \hat{\mu}_{\mathrm{a}} = \ln X_T + \ln T + \ln(1 - X_0/X_T) = T\hat{\mu}_{\mathrm{g}} + o_{\mathbb{P}_{\mathrm{g}}}(\sqrt{T}).$$

Moreover, since $\sum_{t=1}^{T} t = T(T+1)/2$ and

$$\ln X_t = \mu_{\mathrm{g}} t + \sigma_{\mathrm{g}}\sqrt{T}E_T(t/T) + \ln X_0,$$

we have

$$\hat{\mu}_{\mathrm{g}} = \frac{1}{T}(\ln X_T - \ln X_0) = \mu_{\mathrm{g}} + \frac{\sigma_{\mathrm{g}}}{\sqrt{T}} E_T(1),$$

$$\frac{\overline{\ln X} - \hat{\mu}_{\mathrm{g}}(T+1)/2}{\sqrt{T}} = \frac{\sigma_{\mathrm{g}}}{T}\sum_{t=1}^{T} E_T(t/T) - \sigma_{\mathrm{g}}\frac{T+1}{2T} E_T(1) + \frac{\ln X_0}{\sqrt{T}}.$$

The convergence of these quantities follows from the Donsker theorem, and from the convergence of the Riemann sum of the Brownian motion to the corresponding integral. Thus, we study $(\ln \hat{\sigma}_{\mathrm{a}}^2 - 2\hat{\mu}_{\mathrm{g}} T)/\sqrt{T}$. Let $\hat{\sigma}_{\mathrm{a}0}^2 \equiv \hat{\sigma}_{\mathrm{a}}^2 + \hat{\mu}_{\mathrm{g}}^2$. We begin with

Step 1. *An approximation of* $\ln \hat{\sigma}_{\mathrm{a}0}^2$. Because $X_t = X_{t-1}\exp(\mu_{\mathrm{g}} + \sigma_{\mathrm{g}}\varepsilon_t)$, we have $\hat{\sigma}_{\mathrm{a}0}^2 = \sum_{t=1}^{T} X_t^2(1 - \exp(-\mu_{\mathrm{g}} - \sigma_{\mathrm{g}}\varepsilon_t))^2/T$, and thus

$$\begin{aligned} &\frac{1}{\sqrt{T}} \ln\left(\frac{\min_{1\leqslant t\leqslant T}(1 - \exp(-\mu_{\mathrm{g}} - \sigma_{\mathrm{g}}\varepsilon_t))^2}{T}\right) \leqslant \frac{\ln \hat{\sigma}_{\mathrm{a}0}^2 - 2\ln \max_{1\leqslant t\leqslant T} X_t}{\sqrt{T}}, \\ &\frac{\ln \hat{\sigma}_{\mathrm{a}0}^2 - 2\ln \max_{1\leqslant t\leqslant T} X_t}{\sqrt{T}} \leqslant \frac{1}{\sqrt{T}} \ln\left(\max_{1\leqslant t\leqslant T}(1 - \exp(-\mu_{\mathrm{g}} - \sigma_{\mathrm{g}}\varepsilon_t))^2\right). \end{aligned} \tag{8}$$

Moreover,

$$\left|\ln\left(\min_{1\leqslant t\leqslant T}(1 - \exp(-\mu_{\mathrm{g}} - \sigma_{\mathrm{g}}\varepsilon_t))^2\right)\right| \leqslant 2\max_{1\leqslant t\leqslant T} |\ln|1 - \exp(-\mu_{\mathrm{g}} - \sigma_{\mathrm{g}}\varepsilon_t)||,$$

$$\left|\ln \max_{1\leqslant t\leqslant T}(1-\exp(-\mu_g-\sigma_g\varepsilon_t))^2\right| \leqslant 2\max_{1\leqslant t\leqslant T}|\ln|1-\exp(-\mu_g-\sigma_g\varepsilon_t)||.$$

(A1) and the Markov inequality give, for any small positive η:

$$\mathbb{P}\left(\max_{1\leqslant t\leqslant T}|\ln|1-\exp(-\mu_g-\sigma_g\varepsilon_t)||\geqslant \eta\sqrt{T}\right)$$

$$\leqslant T\mathbb{P}(|\ln|1-\exp(-\mu_g-\sigma_g\varepsilon_t)||\geqslant\eta\sqrt{T})\leqslant \frac{\mathbb{E}|\ln|1-\exp(-\mu_g-\sigma_g\varepsilon_1)||^{2+\delta}}{\eta^{2+\delta}T^{\delta/2}},$$

which implies $\max_{1\leqslant t\leqslant T}|\ln|1-\exp(-\mu_g-\sigma_g\varepsilon_t)||=o_{\mathbb{P}}(\sqrt{T})$, and (8) yields

$$\left(\ln\hat{\sigma}_{a0}^2 - 2\sup_{1\leqslant t\leqslant T}\ln X_t\right)\Big/\sqrt{T}=o_{\mathbb{P}_g}(1). \tag{9}$$

Step 2. $(\sup_{1\leqslant t\leqslant T}\ln X_t-\hat{\mu}_g T)/\sqrt{T}$ *goes to* 0. Because $\sup_{s\in[0.1]}W(s)$ is a bounded variable, we have, using the 'reconstructed' version of the Donsker theorem:

$$\begin{aligned}
&\frac{\sup_{1\leqslant t\leqslant T}\ln X_t-\hat{\mu}_g T}{\sqrt{T}}\\
&=\frac{1}{\sqrt{T}}\left(\sup_{1\leqslant t\leqslant T}(\mu_g T+\sigma_g\sqrt{T}W(t/T))-\mu_g T-\sigma_g\sqrt{T}W(1)\right)+o_{\mathbb{P}_g}(1)\\
&=\frac{1}{\sqrt{T}}(\mu_g T+\sigma_g\sqrt{T}W(1)-\mu_g T-\sigma_g\sqrt{T}W(1))+o_{\mathbb{P}_g}(1)\\
&=o_{\mathbb{P}_g}(1).
\end{aligned}$$

Step 3. $\ln(1-\hat{\mu}_a^2/\hat{\sigma}_{a0}^2)/\sqrt{T}$ *goes to* 0. Note that $\hat{\sigma}_{a0}^2-\hat{\mu}_a^2=\hat{\sigma}_a^2\geqslant 0$ and thus $\hat{\mu}_a^2/\hat{\sigma}_{a0}^2$ lies in [0,1]. Moreover,

$$\frac{\hat{\mu}_a^2}{\hat{\sigma}_{a0}^2}\geqslant\frac{(X_T-X_0)^2}{T(X_T-X_{T-1})^2}=\frac{(1-X_0/X_T)^2}{T(1-\exp(-\mu_g-\sigma_g\varepsilon_T))^2}=O_{\mathbb{P}_g}\left(\frac{1}{T}\right),$$

because X_T goes to $+\infty$ and $\mathbb{P}(\mu_g+\sigma_g\varepsilon_T=0)=0$. Thus

$$0\geqslant\ln(1-\hat{\mu}_a^2/\hat{\sigma}_{a0}^2)\geqslant\ln(1-O_{\mathbb{P}_g}(1/T)),$$

and $\ln(1-\hat{\mu}_a^2/\hat{\sigma}_{a0}^2)/\sqrt{T}=O_{\mathbb{P}_g}(1/T\sqrt{T})$.

Eq. (9) along with steps 2 and 3 end the proof of part (i) of Theorem 1.

6.1.2. Arithmetic random walk with trend

The convergence of $\hat{\sigma}_a^2$, $\hat{\mu}_a$ and $T\hat{\mu}_g-\ln T+\ln(X_0)=\ln(X_T/T)=\ln(\mu_a+\sigma_a\sum_{t=1}^T\varepsilon_t/T)$ is a direct consequence of the strong law of large numbers. We now deal with $\overline{\ln|X/T|}$. The strong law of large numbers also yields that, for each event ω, for each small η there exists a constant $N=N(\omega,\eta)$ such for $t\geqslant N$, we have

$$(\mu_a-\eta)t\leqslant X_t=\mu_a t+\sigma_a\sum_{i=1}^t\varepsilon_i\leqslant(\mu_a+\eta)t,$$

and thus

$$\lim_{T\to+\infty}\inf \frac{1}{T}\sum_{t=N}^{T}\log\left((\mu_a-\eta)\frac{t}{T}\right)\leqslant \lim_{T\to+\infty}\inf \overline{\ln|X/T|}$$
$$\leqslant \lim_{T\to+\infty}\sup \overline{\ln|X/T|}\leqslant \lim_{T\to-\infty}\sup \frac{1}{T}\sum_{t=N}^{T}\log\left((\mu_a+\eta)\frac{t}{T}\right).$$

The lower and upper bounds are convergent Riemann sums, the limit of the former being

$$\int_0^1 \ln((\mu_a+\eta)x)\,dx = x(\ln((\mu_a+\eta)x)-1)]_0^1 = \ln(\mu_a+\eta)-1.$$

Thus, $\lim_{T\to+\infty}\overline{\ln|X/T|}=\ln\mu_a-1$, almost surely.

The limit of $T\hat{\sigma}_g^2$ is the one of

$$T\hat{\sigma}_g^2+T\hat{\mu}_g^2=\sum_{t=1}^{T}\ln^2\left|\frac{X_t}{X_{t-1}}\right|,$$

since $T\hat{\mu}_g^2$ goes to 0. Thus, we have to show that the series in the statement of the theorem converges almost surely. First, note that $\mathbb{P}(|X_t/X_{t-1}|\in\{0,+\infty\})=0$, and thus each item of the series is finite. Second, the strong law of large numbers gives that X_t/X_{t-1} converges to 1, and that, for t large enough:

$$\ln^2\left|\frac{X_t}{X_{t-1}}\right|\leqslant\left(\frac{X_t-X_{t-1}}{X_{t-1}}\right)^2\leqslant\frac{\sigma_g^2\varepsilon_t^2}{(\mu_a-\eta)^2(t-1)^2}.$$

The upper bound has a L_1-norm of order $1/t^2$, and the associated serie is normally L_1-convergent, which implies that the series in the theorem converges almost surely. □

6.2. *Proof of Theorem 2*

We begin with the preliminary lemma:

Lemma 1. *Under* (A1) *and* (A2), *there exists a positive constant C depending on the distribution of the ε_t's such that, for $x\in\mathbb{R}$,*

$$\mathbb{P}_a(-\overline{\ln|X|}\geqslant x)\leqslant C\frac{\sqrt{T}\exp(-x)}{\sigma_a}.$$

Proof. We have

$$\mathbb{P}_a(-\overline{\ln|X|}\geqslant x)=\mathbb{P}_a(\overline{\ln|X|}\leqslant -x)\leqslant\mathbb{P}_a\left(\inf_{1\leqslant t\leqslant T}\ln|X_t|\leqslant -x\right)$$
$$\leqslant\sum_{t=1}^{T}\mathbb{P}_a(|X_t|\leqslant\exp(-x))$$

$$= \sum_{t=1}^{T} \mathbb{P}\left(-\exp(-x) \leqslant \mu_{\mathrm{a}} t + \sigma_{\mathrm{a}} \sum_{i=1}^{t} \varepsilon_i \leqslant \exp(-x) \right)$$
$$= \sum_{t=1}^{T} \mathbb{P}\left(\frac{-\exp(-x) - \mu_{\mathrm{a}} t}{\sigma_{\mathrm{a}}\sqrt{t}} \leqslant \frac{1}{\sqrt{t}} \sum_{i=1}^{t} \varepsilon_i \leqslant \frac{\exp(-x) - \mu_{\mathrm{a}} t}{\sigma_{\mathrm{a}}\sqrt{t}} \right).$$

Under (A1) and (A2), the density of $\sum_{i=1}^{t} \varepsilon_i/\sqrt{t}$ is bounded uniformly in t and the state variable, see Petrov (1975), by a constant which depends on the distribution of the ε_t's. Thus

$$\mathbb{P}_{\mathrm{a}}(-\overline{\ln|X|} \geqslant x) \leqslant \frac{2C\exp(-x)}{\sigma_{\mathrm{a}}} \sum_{t=1}^{T} \frac{1}{\sqrt{t}}$$
$$\leqslant \frac{2C\exp(-x)}{\sigma_{\mathrm{a}}} \int_0^T \frac{\mathrm{d}x}{\sqrt{x}} = C\frac{\sqrt{T}\exp(-x)}{\sigma_{\mathrm{a}}}. \qquad \square$$

Proof of Theorem 2. Using the inclusions (6) and (7), we have

$$\mathbb{P}_{\mathrm{a}}(\widetilde{\mathscr{LR}}^{\mathrm{c}}) \leqslant \mathbb{P}_{\mathrm{a}}\left(\ln\hat{\sigma}_{\mathrm{a}} - \overline{\ln|X|} \geqslant (q+1-q)\left(\underline{\mu}\frac{T-1}{2} + \sqrt{T/12}\bar{\sigma}c_{\alpha} \right) + \ln\underline{\sigma} \right)$$
$$\leqslant \mathbb{P}_{\mathrm{a}}(\hat{\sigma}_{\mathrm{a}}^2 \geqslant \underline{\sigma}^2 \exp(q\underline{\mu}(T-1) + 2q\sqrt{T/12}\bar{\sigma}c_{\alpha}))$$
$$+ \mathbb{P}_{\mathrm{a}}\left(-\overline{\ln|X|} \geqslant (1-q)\left(\underline{\mu}\frac{T-1}{2} + \sqrt{T/12}\bar{\sigma}c_{\alpha} \right) \right)$$
$$\leqslant \mathbb{P}_{\mathrm{a}}(\hat{\sigma}_{\mathrm{a}}^2 \geqslant \underline{\sigma}^2 \exp(q\underline{\mu}(T-1) + 2q\sqrt{T/12}\bar{\sigma}c_{\alpha}))$$
$$+ \frac{C\sqrt{T}}{\sigma_{\mathrm{a}}} \exp\left(-(1-q)\underline{\mu}\frac{T-1}{2} - (1-q)\sqrt{T/12}\bar{\sigma}c_{\alpha} \right),$$
$$\mathbb{P}_{\mathrm{a}}(\widetilde{\mathscr{NP}}^{\mathrm{c}}) \leqslant \mathbb{P}_{\mathrm{a}}(\hat{\sigma}_{\mathrm{a}}^2 - \overline{\ln|X|} \geqslant (q+1-q)\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_{\alpha}))$$
$$\leqslant \mathbb{P}_{\mathrm{a}}(\hat{\sigma}_{\mathrm{a}}^2 \geqslant q\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_{\alpha}))$$
$$+ \mathbb{P}_{\mathrm{a}}(-\overline{\ln|X|} \geqslant (1-q)\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_{\alpha}))$$
$$\leqslant \mathbb{P}_{\mathrm{a}}(\hat{\sigma}_{\mathrm{a}}^2 \geqslant q\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_{\alpha}))$$
$$+ \frac{C\sqrt{T}}{\sigma_{\mathrm{a}}} \exp(-(1-q)\exp(2\underline{\mu}T + 2\sqrt{(1/p-1)T}\bar{\sigma}c_{\alpha})),$$

by virtue of Lemma 1. $\square$

References

Andersen, L.C., Jordan, J.L., 1968. Monetary and fiscal actions: A test of their relative importance in economic stabilization. Rev. St. Louis Fed. Reserve Bank 50, 11–24.

Ansley, C.F., Spivey, W.A., Wrobleski, W.J., 1977. A class of transformation for Box–Jenkins seasonal models. Appl. Statist. 26, 173–178.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.
Black, F., 1976. The pricing of commodity contracts. J. Financial Econom. 3, 167–179.
Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Political Economy 81, 673–659.
Blaug, M., 1979. Economic Theory in Retrospect. 3rd ed. Cambridge University Press, Cambridge.
Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Statist. Soc. B-26, 211–243.
Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer, Berlin.
Burridge, P., Guerre, E., 1996. The limit distribution of level crossings of a random walk, and a simple unit root test. Econometric Theory 12, 705–723.
Carlson, K.M., 1978. Does the St. Louis now believe in fiscal policy? Rev. St. Louis Fed. Reserve Bank 60, 13–19.
Chatfield, C., Prothero, D.L., 1973a. Box–Jenkins seasonal forecasting: problems in a case study. J. Roy. Statist. Soc. A-136, 295–315.
Chatfield, C., Prothero, D.L., 1973b. A reply by Dr Chatfield and Dr Prothero. J. Roy. Statist. Soc. A-136, 295–315.
Dyl, E.A., Maberly, E.D., 1986. The weekly pattern in stock index futures: A further note. J. Finance 41, 1149–1152.
Granger, C.W.J., Hallman, J., 1991. Nonlinear transformations of integrated time-series. J.T.S.A. 12, 207–224.
Granger, C.W.J., Newbold, P., 1976. Forecasting transformed series. J. Roy. Statist. Soc. B-38, 189–203.
Guerre, E., 1995a. On the score function of the Box–Cox transformation for integrated time series. INSEE D.P. no. 9506.
Guerre, E., 1995b. The general asymptotic behavior of estimators of the Box–Cox model for integrated time series. INSEE D.P. 9548.
Guerreo, V.M., 1993. Time-series analysis supported by power transformations. J. Forecasting 12, 37–48.
Hafer, R.W., Hein, S.E., 1980. The dynamics and estimation of the short-run money demand. Rev. St. Louis Federal Reserve Bank 62, 26–35.
Hopwood, W.S., Mc Keown, J.C., Newbold, P., 1984. Time series forecasting models involving power transformations. J. Forecasting 3, 57–61.
Layson, S.K., Seaks, T.G., 1984. Estimation and testing for functional form in first difference models. Rev. Econom. Statist. 66, 338–343.
Marsh, T.A., Rosenfeld, E.R., 1983. Stochastic processes for interest rates and equilibrium bond prices. J. Finance 38, 635–647.
Nelson, Jr, H.L., Granger, C.W.J., 1979. Experience with using the Box–Cox transformation when forecasting economic time series. J. Econometrics 10, 57–69.
Park, T., 1991. Double length regressions for testing the Box–Cox difference transformation. Rev. Econom. Statist. 73, 181–185.
Petrov, V.V., 1975. Sums of Independent Random Variables. Springer, Berlin.
Sakia, R.M., 1992. The Box–Cox transformation technique: a review. Statistician 41, 169–178.
Thornton, J.R., Agnello, R.J., Link, C.R., 1978. Poverty and economic growth: Trickle down Peter out. Econom. Inquiry 16, 385–394.