Testing for normality of weakly dependent data

Zacharias Psaradakis
Department of Economics, Mathematics and Statistics
Birkbeck, University of London
Marián Vávra
Advisor to the Governor and Research Department
National Bank of Slovakia
E-mail: marian.vavra@nbs.sk

NBS Research Seminar, Bratislava, Slovakia, 2014

Introduction

The validity of normality of the marginal law (of economic variables) may be very useful in practice:

- **Econometrics**:
  - econometric modelling/identification;
  - forecasting (testing for Gaussian prediction bands)
  - bootstrap techniques;
  - statistical filters.
- **Macro-Finance**:
  - option pricing;
  - risk management;
  - yield curve modelling (testing for Gaussian affine term structure models).

Some basics about normality tests

There are 3 classes of (**i.i.d. based**) tests in the literature:

- Empirical distribution (characteristic) function tests (e.g. Kolmogorov-Smirnov test, Anderson-Darling test);
- Order statistic based tests (e.g. Shapiro-Wilks test);
- Moment based tests (e.g. Jarque-Bera test).

Bai and Ng (2005) modified a JB test for weakly dependent observations (the only test for w.d.). But the test suffers from some shortcomings

- The finite-sample properties of the test are very poor;
- The test requires the first eight moments to be finite.

The main task of the paper

The main task is to propose a new version of some existing test for normality of weakly dependent data under minimal (moment) conditions.

Which test to modify?

The Anderson-Darling (AD) test is one of the most power test of normality in the literature. The AD test takes the form

$$\mathcal{A}_n = \int_{\mathbb{R}} \frac{(F_n(z) - \Phi(z))^2}{\Phi(z)(1 - \Phi(z))} \, \mathrm{d}\Phi(z), \quad z \in \mathbb{R}, \tag{1}$$

where Φ denotes a standard normal distribution and F_n is the empirical distribution function associated with $\{X_t : t = 1, \ldots, n\}$

$$F_n(z) = \frac{1}{n} \sum_{t=1}^{n} I\left(\frac{X_t - \mu}{\sigma} \leq z\right), \quad z \in \mathbb{R}, \tag{2}$$

where $I(\cdot)$ is a standard indicator function and $\mu = \mathbb{E}(X_t)$ and $\sigma = \sqrt{\mathrm{var}(X_t)}$. It can be shown that as $n \to \infty$ and for given μ and σ, then

$$n\mathcal{A}_n \xrightarrow{d} \int_0^1 \frac{U^2(\omega)}{\omega(1 - \omega)} \, \mathrm{d}\omega, \tag{3}$$

where U is the Brownian bridge.

Complications with the AD test statistic

We face 2 complications when using the AD test in practice:

- Complication 1: once the parameter(s) μ and/or σ in (2) are **unknown** and must be estimated from data. The asymptotic distribution is no longer parameter free ((3) does not hold);
- Complication 2: economic time series (at least some of them) can be characterized as **weakly dependent** process and the asymptotic distribution is no longer parameter free;
- An appropriate bootstrap method has to be used in order to (i) replicate the dependence in data; (ii) impose the normality assumption under H_0.
- We implement an AR-sieve bootstrap to calculate the critical values of the AD test.

Assumptions about the stochastic process

Assumption 1 The underlying stochastic process $\{X_t\}$ is a real-valued stationary and weakly dependent process allowing for a Wold representation given by

$$X_t = \mu + \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{4}$$

where $\mu \in \mathbb{R}$, the roots of the lag polynomial $\psi(q) = 1 - \sum_{j=1}^{\infty} \psi_j q^j$ lie outside the unit disk and $\sum_{j=1}^{\infty} j|\psi_j| < \infty$, the error sequence $\{\epsilon_t\}$ is assumed to be stationary and ergodic such that $\mathbb{E}(\epsilon_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\epsilon_t^2|\mathcal{F}_{t-1}) = s^2 < \infty$, where $\mathcal{F}_t = \{\epsilon_t, \epsilon_{t-1}, \dots\}$ is the σ-field, $\mathbb{E}(\epsilon_t^4) < \infty$ and the density function $f(\epsilon_t)$ is absolutely continuous.

Note

Under an additional mild assumption the process in (4) can be written into an AR(∞) model.

AR-sieve bootstrap

Algorithm 1

(i) Select an appropriate lag order p of an AR model using the AIC.

(ii) Estimate the unknown AR(p) model parameters by the OLS.

(iii) Construct a sequence of the estimated residuals $\{\hat{\epsilon}_t : t = p+1, \ldots, n\}$ by the recursion

$$\hat{\epsilon}_t = X_t - \hat{c} - \sum_{i=1}^{p} \hat{\phi}_i X_{t-i}.$$

(iv) Under the null hypothesis of marginal normality, the hypothesized distribution equals to a standard normal distribution Φ. Therefore, consistently with the null, draw independent random errors $\epsilon_t^* \sim N(0, \hat{s}^2)$, for $t = 1, \ldots, n+100$, where $\hat{s}^2 = (n-2p-1)^{-1} \sum_{t=p+1}^{n} \hat{\epsilon}_t^2$.

AR-sieve bootstrap

Algorithm 1

(v) Generate bootstrap replicates $\{X_t^* : t = 1, \ldots, n + 100\}$ by the recursion

$$X_t^* = \hat{c} + \sum_{i=1}^{p} \hat{\phi}_i X_{t-i}^* + \epsilon_t^*,$$

where the process is initiated by a vector of sample averages: $(X_{-p+1}^*, \ldots, X_0^*) = (\bar{X}, \ldots, \bar{X})$. The first 100 data points are then discarded in order to eliminate start-up effects and the remaining n data points are used.

(vi) Construct a bootstrap analogy of the BAD test $\mathcal{A}_n^*$ calculated from a bootstrap sample $\{X_t^* : t = 1, \ldots, n\}$.

AR-sieve bootstrap

Algorithm 1

(vii) Repeat steps (iv)–(vi) independently B times to get a sample of the BAD statistics $\{\mathcal{A}^*_{n,i} : i = 1, \dots, B\}$. Then, the sampling distributions of the BAD test statistic is approximated by the empirical distribution functions associated with $\{\mathcal{A}^*_{n,i} : i = 1, \dots, B\}$: $H^*_n(u) = B^{-1} \sum_{i=1}^{B} I(\mathcal{A}^*_{n,i} \leq u)$. Finally, a bootstrap test of the nominal level α rejects the null hypothesis of normality if

$$\hat{\mathcal{A}}_n > \inf\{u : H^*_n(u) \geq (1 - \alpha)\},$$

where $\hat{\mathcal{A}}_n$ is the BAD test statistic obtained from the observed sample $\{X_t : t = 1, \dots, n\}$.

Multivariate extension

- Since the estimation of the multivariate EDF-based tests is computationally expensive, some dimensionality reduction technique is desirable for multiple time series applications;
- A natural solution utilizes the fact that if a $(k \times 1)$ random vector $\boldsymbol{x}_t$ is distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under H_0, then $y_t = (\boldsymbol{x}_t - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu})$ is distributed as $\chi^2(k)$;
- Requires the VAR-sieve bootstrap = not very convenient in our case;
- We introduce a simple, linear, yet flexible, dimensionality reduction method which is in spirit similar to the quadratic-form one. The method is motivated by the well-known Cramér-Wold theorem.

Multivariate extension...

Theorem (Cramér-Wold)

For a $(k \times 1)$ *random vectors* $\boldsymbol{x}_t = (X_{1t}, \ldots, X_{kt})'$ *and* $\boldsymbol{x} = (X_1, \ldots, X_k)'$*, a necessary and sufficient condition for* $\boldsymbol{x}_t \xrightarrow{d} \boldsymbol{x}$ *with a joint distribution* $F(\mathbf{x})$ *as* $t \to \infty$ *is that* $\boldsymbol{\lambda}'\boldsymbol{x}_t \xrightarrow{d} \boldsymbol{\lambda}'\boldsymbol{x}$ *with a marginal distribution function* $F(\boldsymbol{\lambda}'\boldsymbol{x})$ *for each* $\boldsymbol{\lambda} \in \boldsymbol{R}^k$*.*

An ultimate question is, however, how to determine the aggregation vector $\boldsymbol{\lambda}$.

Multivariate extension...

We propose a two-step procedure (orthogonalization and aggregation):

- Step 1: The orthogonalization can be done by the eigenvalue decomposition which decomposes a $(k \times k)$ symmetric and positive-definite variance-covariance matrix as follows: $\mathrm{cov}(\boldsymbol{x}_t) = \boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{P}'$, where $\boldsymbol{P}$ is a square matrix;
- Step 2: Orthogonalized components $\boldsymbol{z}_t = \boldsymbol{P}^{-1}\boldsymbol{x}_t = (Z_{1t}, \ldots, Z_{kt})'$ are then aggregated using the skewness-based weighting function defined as $\boldsymbol{w} = [w_i]$, where $w_i = 1$ if $skew(Z_{it}) \geq 0$ and $w_i = -1$ if $skew(Z_{it}) < 0$, for $i \in \{1, \ldots, k\}$, and $skew(\cdot)$ can by any measure of skewness;
- Finally, one can then apply the testing procedure described in Algorithm 1 to the transformed scalar process $X_t = \boldsymbol{\lambda}'\boldsymbol{x}_t$, where $\boldsymbol{\lambda} = \boldsymbol{w}'\boldsymbol{P}^{-1}$.

Monte Carlo setup

The finite-sample properties of the BN and BAD tests are assessed using the following DGPs:

M1: $X_t = 0.5X_{t-1} + \varepsilon_t$

M2: $X_t = 0.8X_{t-1} + \varepsilon_t$

M3: $X_t = 0.8X_{t-1} - 0.4X_{t-2} - 0.5\varepsilon_{t-1} + \varepsilon_t$

M4: $X_t = 0.5X_{t-1} - 0.3X_{t-1}\epsilon_{t-1} + \epsilon_t$

M5: $X_t = 1.5S_t - 0.5(1 - S_t) + 0.5X_{t-1} + \epsilon_t$

M6:

$$\boldsymbol{x}_t = \begin{pmatrix} 0.4 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \boldsymbol{x}_{t-1} + \begin{pmatrix} a_t \\ \varepsilon_t \end{pmatrix}$$

where $a \sim N(0,1)$ and $\varepsilon \in \{N(0,1), S1, S2, S3, A1, A2, A3\}$.

Monte Carlo setup...

Table : Parameters of Generalized Lambda Distribution

	λ_1	λ_2	λ_3	λ_4	skewness	kurtosis
S1	0.000000	-1.000000	-0.080000	-0.080000	0.0	6.0
S2	0.000000	-0.397912	-0.160000	-0.160000	0.0	11.6
S3	0.000000	-1.000000	-0.240000	-0.240000	0.0	126.0
A1	0.000000	-1.000000	-0.007500	-0.030000	1.5	7.5
A2	0.000000	-1.000000	-0.100900	-0.180200	2.0	21.1
A3	0.000000	-1.000000	-0.001000	-0.130000	3.2	23.8

Monte Carlo results

		$n = 100$		$n = 200$		$n = 500$	
DGP	distr.	BN	BAD	BN	BAD	BN	BAD
M1	N	0.03	0.05	0.05	0.05	0.09	0.04
	S1	0.01	0.23	0.04	0.40	0.22	0.78
	S2	0.04	0.44	0.09	0.69	0.34	0.97
	S3	0.04	0.63	0.11	0.88	0.33	1.00
	A1	0.21	0.81	0.83	0.97	1.00	1.00
	A2	0.10	0.67	0.35	0.89	0.79	1.00
	A3	0.45	1.00	0.97	1.00	1.00	1.00
M2	N	0.01	0.06	0.02	0.06	0.04	0.05
	S1	0.00	0.11	0.00	0.14	0.02	0.17
	S2	0.01	0.19	0.02	0.24	0.06	0.37
	S3	0.01	0.28	0.03	0.39	0.09	0.65
	A1	0.00	0.25	0.03	0.43	0.46	0.80
	A2	0.01	0.32	0.06	0.44	0.36	0.77
	A3	0.00	0.59	0.03	0.88	0.73	1.00

Why does it work?

- The formal proof is given in Bühlmann (1997), some modifications are, however, necessary,
- **An intuitive explanation:** Bickel and Bühlmann (1997) explain that the closure of the Wold representation is fairly large. It means that for any non-linear stochastic process there exist another process in the closer of linear processes having identical sample paths with probability exceeding $1/e \approx 0.37$;
- **A rule of thumb for applications:** Kreiss et al. (2011) prove that if the distribution of a relevant statistic is determined solely by the first two moments, then the AR-sieve bootstrap is expected to work.

Example: symmetric or asymmetric fan-charts?

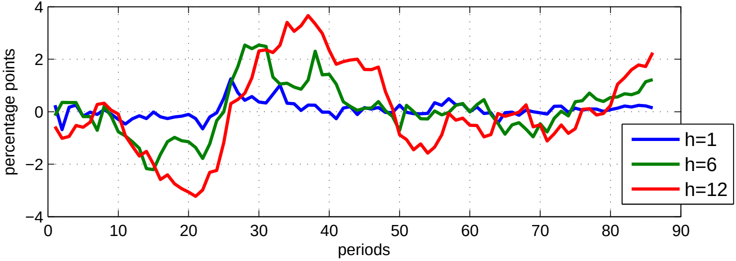
- Prediction bands (fan-charts) have become a standard tool of central banks in assessing risk about the future development of the economy (see Clements and Hendry (2008, Chap. 2,3));
- But we hold the view that **only correct prediction bands** may be useful in practice;
- Researches face an ultimate dilemma whether to construct symmetric or asymmetric fan-charts (**symmetric** = Bank of Canada, Riksbank, Norges Bank; **asymmetric** = Bank of England, NBS, National Bank of Poland, Bank of Italy);
- Other applications of the BAD test: calculating the probability of deflation; density forecast evaluation (DSGE vs. (B)VAR models).

Example: NBS inflation forecast errors

- The forecast error is defined as $X_t(h) = \pi_{t+h} - \pi_t(h)$ for a forecast horizon $h \in \{1, \ldots, 12\}$, where $\pi_t(h)$ stands for the h-step ahead forecast of the (year-on-year) CPI inflation rate and π_{t+h} denotes the actual inflation rate.
- We focus on testing for marginal and joint normality of the NBS inflation forecast errors:
  - marginal hypothesis:
    $H_0 : F(X_t(h)) = N(0, \sigma_h^2)$ against
    $H_1 : F(X_t(h)) \neq N(0, \sigma_h^2)$;
  - joint hypothesis:
    $H_0 : F(X_t(1), \ldots, X_t(12)) = N(\mathbf{0}, \mathbf{\Sigma})$ against
    $H_1 : F(X_t(1), \ldots, X_t(12)) \neq N(\mathbf{0}, \mathbf{\Sigma})$.

Example: symmetric or asymmetric fan-charts?

Figure : NBS Inflation Forecast Errors



Example: symmetric or asymmetric fan-charts?

Table : P-values of the BAD Test for Marginal and Joint Normality of the Inflation Forecast Errors

hypothesis	horizon	$B = 1000$	$B = 5000$	$B = 10000$
marginal	$h = 1$	0.04	0.04	0.04
	$h = 2$	0.36	0.35	0.35
	$h = 3$	0.24	0.24	0.24
	$h = 4$	0.40	0.41	0.40
	$h = 5$	0.49	0.50	0.50
	$h = 6$	0.63	0.61	0.63
	$h = 7$	0.91	0.89	0.87
	$h = 8$	0.84	0.86	0.86
	$h = 9$	0.87	0.86	0.87
	$h = 10$	0.73	0.74	0.73
	$h = 11$	0.57	0.59	0.57
	$h = 12$	0.56	0.59	0.58
joint	$h = 1, \ldots, 12$	0.65	0.62	0.62

References

Bickel, P. and P. Bühlmann (1997). Closure of linear processes. *Journal of Theoretical Probability 10*, 445–479.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli 3*, 123–148.

Clements, M. and D. Hendry (2008). *A Companion to Economic Forecasting*. Wiley.

Kreiss, J., E. Paparoditis, and D. Politis (2011). On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics 39*, 2103–2130.

Thanks

Thank you for attention.