The Stata Journal (2017)
17, Number 4, pp. 901–915

Automatic portmanteau tests with applications to market risk management

Guangwei Zhu
Southwestern University of Finance and Economics
Chengdu, China
zhugw@swufe.edu.cn

Zaichao Du
Southwestern University of Finance and Economics
Chengdu, China
duzc@swufe.edu.cn

Juan Carlos Escanciano
Department of Economics
Indiana University
Bloomington, IN
jescanci@indiana.edu

Abstract. In this article, we review some recent advances in testing for serial correlation, provide code for implementation, and illustrate this code's application to market risk forecast evaluation. We focus on the classic and widely used portmanteau tests and their data-driven versions. These tests are simple to implement for two reasons: First, the researcher does not need to specify the order of the tested autocorrelations, because the test automatically chooses this number. Second, its asymptotic null distribution is chi-squared with one degree of freedom, so there is no need to use a bootstrap procedure to estimate the critical values. We illustrate the wide applicability of this methodology with applications to forecast evaluation for market risk measures such as value-at-risk and expected shortfall.

Keywords: st0504, dbptest, rtau, autocorrelation, consistency, power, Akaike's information criterion, Schwarz's Bayesian information criterion, market risk

1 Introduction

Testing for serial correlation has held a central role in time-series analysis since its inception (see the early contributions by Yule [1926] and Quenouille [1947]). Despite the many proposals and variations since the seminal contribution of Box and Pierce (1970), the so-called portmanteau tests are still the most widely used. In its simplest form, the employed statistic is just the sample size times the sum of the first p-squared sample autocorrelations, which is compared with critical values from a chi-squared distribution with p degrees of freedom (with a correction if the test is applied to residuals). The basic Box–Pierce statistic has been slightly modified to improve its finite sample performance; see Davies, Triggs, and Newbold (1977); Ljung and Box (1978); Davies and Newbold (1979); or Li and McLeod (1981). The properties of the classical Box–Pierce tests have been extensively studied in the literature; see, for example, the monograph by Li (2004) for a review of this literature. Much of the theoretical literature on Box–Pierce tests was developed under the independence assumption and hence is generally invalid when applied to dependent data (the asymptotic size of the test is different from the nominal level); see Newbold (1980) or, more recently, Francq, Roy, and Zakoïan (2005) for valid tests. This limitation of classical Box–Pierce tests is by now well understood. In this article, we focus on a different limitation: the selection of the employed number of autocorrelations is arbitrary. We review the contribution of Escanciano and Lobato (2009), who proposed a data-driven portmanteau statistic where the number of correlations is not fixed but selected automatically from the data. In this article, we give a synthesis of this methodology, introduce new general assumptions for its validity, review new applications in risk management, and provide code for its implementation.

2 Automatic portmanteau tests: A synthesis

Given a strictly stationary process $\{Y_t\}_{t\in Z}$ with $E(Y_t^2) < \infty$ and $\mu = E(Y_t)$, define the autocovariance of order j as

$$\gamma_j = \mathrm{Cov}(Y_t, Y_{t-j}) = E\left\{(Y_t - \mu)(Y_{t-j} - \mu)\right\}, \qquad \text{for all } j \geq 0$$

and the jth order autocorrelation as $\rho_j = \gamma_j/\gamma_0$. We aim to test the null hypothesis

$$H_0 : \rho_j = 0, \qquad \text{for all } j \geq 1$$

against the fixed alternative hypotheses

$$H_1^K : \rho_j \neq 0, \qquad \text{for some } 1 \leq j \leq K$$

and some $K \geq 1$.

Suppose we observe data $\{Y_t\}_{t=1}^n$. γ_j can then be consistently estimated by the sample autocovariance

$$\widehat{\gamma}_j = \frac{1}{(n-j)} \sum_{t=1+j}^{n} \left(Y_t - \overline{Y}\right)\left(Y_{t-j} - \overline{Y}\right), \quad j = 0, \ldots, n-1$$

where $\overline{Y}$ is the sample mean, and we can also introduce $\widehat{\rho}_j = \widehat{\gamma}_j/\widehat{\gamma}_0$ to denote the jth order sample autocorrelation.

The Box–Pierce Q_p statistic (Box and Pierce 1970) is just

$$Q_p = n\sum_{j=1}^{p} \widehat{\rho}_j^2$$

which is commonly implemented via the Ljung and Box (LB, 1978) modification

$$\mathrm{LB}_p = n(n+2)\sum_{j=1}^{p}(n-j)^{-1}\widehat{\rho}_j^2$$

When $\{Y_t\}_{t=1}^n$ are independent and identically distributed (i.i.d.), both Q_p and LB_p converge to a chi-squared distribution with p degrees of freedom, or χ_p^2. When $\{Y_t\}_{t=1}^n$

are serially dependent, for example, when Y_t is a residual from a fitted model, the asymptotic distribution of Q_p or LB_p is generally different from χ^2_p and depends on the data-generating process in a complicated way; see Francq, Roy, and Zakoïan (2005) and Delgado and Velasco (2011).

In this section, we synthesize the AQ test methodology that was suggested in Escanciano and Lobato (2009) and extend the methodology to other situations. The main ingredients of the methodology are 1) the following asymptotic results for individual autocorrelations: for $j = 1, \ldots, d$, where d is a fixed upper bound,

$$\sqrt{n}\,(\widehat{\rho}_j - \rho_j) \xrightarrow{D} N(0, \tau_j) \tag{1}$$

for a positive asymptotic variance $\tau_j > 0$, with[1]

$$\widehat{\tau}_j \xrightarrow{P} \tau_j \tag{2}$$

and 2) a data-driven construction of p, given below. For i.i.d. observations, $\tau_j = 1$, and trivially, we can take $\widehat{\tau}_j = 1$, but in other more general settings with weak dependence or estimation effects, we will have an unknown $\tau_j \neq 1$ that needs to be estimated. Our definitions of portmanteau tests allow for general cases. Define

$$Q^*_p = n \sum_{j=1}^{p} \widetilde{\rho}^2_j$$

where $\widetilde{\rho}_j = \widehat{\rho}_j / \sqrt{\widehat{\tau}_j}$ is called a "generalized autocorrelation" here. Then, the AQ test is given by

$$\text{AQ} = Q^*_{\widetilde{p}} \tag{3}$$

where

$$\widetilde{p} = \min\{p : 1 \le p \le d; L_p \ge L_h, h = 1, 2 \ldots, d\}$$

with

$$L_p = Q^*_p - \pi(p, n, q)$$

$\pi(p, n, q)$ is a penalty term that takes the form

$$\pi(p, n, q) = \begin{cases} p \log n, & \text{if} \quad \max_{1 \le j \le d} \sqrt{n}\,|\widetilde{\rho}_j| \le \sqrt{q \log n} \\ 2p, & \text{if} \quad \max_{1 \le j \le d} \sqrt{n}\,|\widetilde{\rho}_j| > \sqrt{q \log n} \end{cases} \tag{4}$$

and $q = 2.4$. The penalty term in (4) has been proposed by Inglot and Ledwina (2006b) for testing the goodness of fit for a distribution. The value of $q = 2.4$ is motivated from extensive simulation evidence in Inglot and Ledwina (2006a) and Escanciano and Lobato (2009). The value of $q = 0$ corresponds to the Akaike information criterion (AIC); see Akaike (1974). The value of $q = \infty$ corresponds to the Bayesian information criterion (BIC); see Schwarz (1978). In the context of testing for serial correlation, AIC

1. In this article, we use $\xrightarrow{D}$ and $\xrightarrow{P}$ to denote convergence in distribution and in probability, respectively.

is good at detecting nonzero correlations at long lags but leads to size distortions. In contrast, BIC controls the size accurately and is good for detecting nonzero correlations at short lags. As shown empirically in figures 1 and 2 in Escanciano and Lobato (2009), the choice of $q = 2.4$ provides a "switching effect" in which one combines the advantages of AIC and BIC. Thus, we recommend $q = 2.4$ in applications. The upper bound d does not affect the asymptotic null distribution of the test, although it may have an impact on power if it is chosen too small. The finite sample performance of the automatic tests is not sensitive to the choice of d for moderate and large values of this parameter, as shown in table 5 of Escanciano and Lobato (2009) and table 6 of Escanciano, Lobato, and Zhu (2013). Extensive simulation experience suggests that the choice of d that is equal to the closest integer around $\sqrt{n}$ performs well in practice.

Theorem 1 *Under the null hypothesis, (1) and (2),* $\text{AQ} \xrightarrow{D} \chi^2_1$.

This theorem justifies the rejection region

$$\text{AQ} > \chi^2_{1,1-\alpha}$$

where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$ quantile of the χ^2_1. The following theorem shows the consistency of the test.

Theorem 2 *Assume* $\widehat{\rho}_j \xrightarrow{P} \rho_j$ *for* $j = 1, \ldots, d$, *and let (2) hold. Then, the test based on* AQ *is consistent against* H_1^K, *for* $K \leq d$.

Note that joint convergence of the vector of autocorrelations is unnecessary, in contrast to much of the literature. Thus, the methodology of this article does not require estimation of large dimensional asymptotic variances.

The proofs of both theorems follow from straightforward modification of those in Escanciano and Lobato (2009) and are hence omitted.

Remark 1. The methodology can be applied to any setting where (1) and (2) can be established. This includes raw data or residuals from any model. There is an extensive literature proving conditions such as (1) and (2) under different assumptions; see examples below.

Remark 2. The reason for the χ^2_1 limiting distribution of the AQ test is that under the null hypothesis, $\lim_{n\to\infty} P(\widetilde{p} = 1) = 1$. Heuristically, under the null hypothesis, Q_p^* is small, and $\pi(p, n, q)$ increases in p, so the optimal choice selected is the lowest dimensionality $p = 1$ with high probability.

3 Applications to risk management

We illustrate the general applicability of the methodology with new applications in risk management. There is a very extensive literature on the quantification of market risk for derivative pricing, for portfolio choice and for risk management purposes. This literature has long been particularly interested in evaluating market risk forecasts, or the so-called backtests; see Jorion (2007) and Christoffersen (2009) for comprehensive reviews. A leading market risk measure has been the value at risk (VAR), and more recently, expected shortfall (ES). VAR summarizes the worst loss over a target horizon that will not be exceeded at a given level of confidence called "coverage level". ES is the expected value of losses beyond a given level of confidence.[2] We review popular backtests for VAR and ES and derive automatic versions using the general methodology above.

Let R_t denote the revenue of a bank at time t, and let Ω_{t-1} denote the risk manager's information at time $t-1$, which contains lagged values of R_t and possibly lagged values of other variables, say, X_t. That is, $\Omega_{t-1} = \{X_{t-1}, X_{t-2}, \ldots; R_{t-1}, R_{t-2}, \ldots\}$. Let $G(\cdot, \Omega_{t-1})$ denote the conditional cumulative distribution function of R_t given Ω_{t-1}, that is, $G(\cdot, \Omega_{t-1}) = \Pr(R_t \leq \cdot|\Omega_{t-1})$. Assume $G(\cdot, \Omega_{t-1})$ is continuous. Let $\alpha \in [0, 1]$ denote the coverage level. The α-level VAR is defined as the quantity $\text{VAR}_t(\alpha)$ such that

$$\Pr\{R_t \leq -\text{VAR}_t(\alpha)|\Omega_{t-1}\} = \alpha \tag{5}$$

That is, the $-\text{VAR}_t(\alpha)$ is the αth percentile of the conditional distribution G,

$$\text{VAR}_t(\alpha) = -G^{-1}(\alpha, \Omega_{t-1}) = -\inf\{y : G(y, \Omega_{t-1}) \geq \alpha\}$$

Define the α-violation or hit at time t as

$$h_t(\alpha) = 1\{R_t \leq -\text{VAR}_t(\alpha)\}$$

where $1(\cdot)$ denotes the indicator function. That is, the violation takes the value 1 if the loss at time t is larger than or equal to $\text{VAR}_t(\alpha)$, and it takes the value 0 otherwise. Equation (5) implies that violations are Bernoulli variables with mean α and, moreover, that centered violations are a martingale difference sequence (MDS) for each $\alpha \in [0, 1]$; that is,

$$E\{h_t(\alpha) - \alpha|\Omega_{t-1}\} = 0 \text{ for each } \alpha \in [0, 1]$$

This restriction has been the basis for the extensive literature on backtesting VAR. Two of its main implications, the zero mean property of the hit sequence $\{h_t(\alpha) - \alpha\}_{t=1}^{\infty}$ and its uncorrelation, led to the unconditional and conditional backtests of Kupiec (1995) and Christoffersen (1998), respectively, which are the most widely used backtests. More recently, Berkowitz, Christoffersen, and Pelletier (2011) have proposed the Box–Pierce-type test for VAR,

$$C_{\text{VAR}}(p) = n \sum_{j=1}^{p} \widehat{\rho}_j^2$$

with $\widehat{\rho}_j = \widehat{\gamma}_j/\widehat{\gamma}_0$ and $\widehat{\gamma}_j = 1/(n-j)\sum_{t=1+j}^{n}\{\widehat{h}_t(\alpha) - \alpha\}\{\widehat{h}_{t-j}(\alpha) - \alpha\}$, and where $\{\widehat{h}_t(\alpha) = R_t \leq -\widehat{\text{VAR}}_t(\alpha)\}_{t=1}^{n}$, for an estimator of the VAR, $\widehat{\text{VAR}}_t(\alpha)$. An automatic version of the test statistic in Berkowitz, Christoffersen, and Pelletier (2011) can be

2. Other names for ES are conditional VAR, average VAR, tail VAR, or expected tail loss.

computed following the algorithm above with $\tau_j = 1$. This test is valid only when there are no estimation effects. If T is the in-sample size for estimation and n is the out-of-sample size used for forecast evaluation, the precise condition for no estimation effects in backtesting VAR and ES is that both $T \to \infty$ and $n \to \infty$ at a rate such that $n/T \to 0$ (that is, the in-sample size is much larger than the out-of-sample size). More generally, Escanciano and Olmo (2010) provided primitive conditions for the convergences (1) and (2) to hold in a general setting where there are estimating effects from estimating VAR. When estimation effects are present, τ_j no longer equals 1, but Escanciano and Olmo (2010) provide suitable estimators, $\widehat{\tau}_j$, satisfying (2). Let AC_{VAR} denote the AQ version of $C_{\mathrm{VAR}}(p)$.

More recently, there has been a move in the banking sector toward ES as a suitable measure of market risk able to capture "tail risk" (the risk coming from very big losses). ES is defined as the conditional expected loss given that the loss is larger than $\mathrm{VAR}_t(\alpha)$, that is,

$$\mathrm{ES}_t(\alpha) = E\left\{-R_t|\Omega_{t-1}, -R_t > \mathrm{VAR}_t(\alpha)\right\}$$

Definition of a conditional probability and a change of variables yield a useful representation of $\mathrm{ES}_t(\alpha)$ in terms of $\mathrm{VAR}_t(\alpha)$,

$$\mathrm{ES}_t(\alpha) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VAR}_t(u)du \tag{6}$$

Unlike $\mathrm{VAR}_t(\alpha)$, which contains information only on one quantile level α, $\mathrm{ES}_t(\alpha)$ contains information from the whole left tail by integrating all VARs from 0 to α. As we did with (6), we define the cumulative violation process,

$$H_t(\alpha) = \frac{1}{\alpha}\int_0^\alpha h_t(u)du$$

Because $h_t(u)$ has mean u, then by Fubini's theorem, $H_t(\alpha)$ has mean $1/\alpha\int_0^\alpha udu = \alpha/2$. Moreover, again by Fubini's theorem, the MDS property of the class $\{h_t(\alpha) - \alpha : \alpha \in [0,1]\}_{t=1}^\infty$ is preserved by integration, which means that $\{H_t(\alpha) - \alpha/2\}_{t=1}^\infty$ is also an MDS. For computational purposes, it is convenient to define $u_t = G(R_t, \Omega_{t-1})$. Because $h_t(u) = 1\{R_t \leq -\mathrm{VAR}_t(u)\} = 1(u_t \leq u)$, we obtain

$$\begin{aligned} H_t(\alpha) &= \frac{1}{\alpha}\int_0^\alpha 1(u_t \leq u)du \\ &= \frac{1}{\alpha}(\alpha - u_t)1(u_t \leq \alpha) \end{aligned}$$

Like violations, cumulative violations are distribution free because $\{u_t\}_{t=1}^\infty$ comprises a sample of i.i.d. $U[0,1]$ variables (see Rosenblatt [1952]). Cumulative violations have been recently introduced in Du and Escanciano (2017). The variables $\{u_t\}_{t=1}^\infty$ necessary to construct $\{H_t(\alpha)\}_{t=1}^\infty$ are generally unknown because the distribution of the data G

is unknown. In practice, researchers and risk managers specify a parametric conditional distribution $G(\cdot, \Omega_{t-1}, \theta_0)$, where θ_0 is some unknown parameter in $\Theta \subset \mathbb{R}^p$, and proceed to estimate θ_0 before producing VAR and ES forecasts. Popular choices for distributions $G(\cdot, \Omega_{t-1}, \theta_0)$ are those derived from location-scale models with Student's t distributions, but other choices can be certainly entertained in our setting. With the parametric model at hand, we can define the "generalized errors"

$$u_t(\theta_0) = G(R_t, \Omega_{t-1}, \theta_0)$$

and the associated cumulative violations

$$H_t(\alpha, \theta_0) = \frac{1}{\alpha} \left\{\alpha - u_t\left(\theta_0\right)\right\} 1(u_t(\theta_0) \leq \alpha)$$

As with VARs, the arguments above provide a theoretical justification for backtesting ES by checking whether $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^{\infty}$ have zero mean (unconditional ES backtest) and whether $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^{\infty}$ are uncorrelated (conditional ES backtest).

Let $\widehat{\theta}$ be an estimator of θ_0 and construct residuals

$$\widehat{u}_t = G(R_t, \Omega_{t-1}, \widehat{\theta})$$

and estimated cumulative violations

$$\widehat{H}_t(\alpha) = \frac{1}{\alpha}(\alpha - \widehat{u}_t)1(\widehat{u}_t \leq \alpha)$$

Then, we obtain

$$\widehat{\gamma}_j = \frac{1}{n-j} \sum_{t=1+j}^{n} \left\{\widehat{H}_t(\alpha) - \alpha/2\right\} \left\{\widehat{H}_{t-j}(\alpha) - \alpha/2\right\} \text{ and } \widehat{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0}$$

Du and Escanciano (2017) construct the Box–Pierce test statistic

$$C_{\text{ES}}(p) = n \sum_{j=1}^{p} \widehat{\rho}_j^2$$

and derive its asymptotic null distribution. In particular, they establish conditions for (1) and (2) to hold and provide expressions for the corresponding $\widehat{\tau}_j$. Let AC_{ES} denote the AQ version of $C_{\text{ES}}(p)$.

Compared with the existing backtests, these automatic backtests select p from the data and require only estimation of marginal asymptotic variances of marginal correlations to obtain known limiting distributions.

4 Implementation

We introduce the `dbptest` command to implement the AQ test (3). Notice that $\tau_j = 1$ for i.i.d. observations and for backtesting VAR and ES without estimation effects.

We also provide the `rtau` command to estimate τ_j for more general cases, including MDS as in Escanciano and Lobato (2009), as well as backtests for VAR and ES with estimation effects as in Escanciano and Olmo (2010) and Du and Escanciano (2017), respectively.

4.1 Syntax

Automatic Q test

`dbptest` *varname* [*if*] [*in*] [, `mu(#)` `q(#)` `tauvector(`*matname*`)` `nlags(#)`]

Estimating τ_j

`rtau` *varname* [*if*] [*in*] , `nlags(#)` `seriestype(`*type*`)` [`cl(#)` `nobs(#)`]

4.2 Options

Automatic Q test

`mu(#)` specifies the mean of the tested variable. The default is the variable's sample mean.

`q(#)` is a fixed positive number to control the switching effect between the AIC and BIC. The default is `q(2.4)`.

`tauvector(`*matname*`)` specifies a column vector containing variances of the autocorrelations. The default is a vector of 1s.

`nlags(#)` specifies the maximum number of lags of autocorrelations. The default is the closest integer around $\sqrt{n}$, where n is the number of observations. If it is larger than the dimension of `tauvector()`, it will be replaced by the dimension of `tauvector()`.

Estimating τ_j

`nlags(#)` specifies the number of lags of autocorrelations. `nlags()` is required.

`seriestype(`*type*`)` specifies one of the following types: `mds`, `var`, or `es`. `seriestype()` is required.

`seriestype(mds)` specifies *varname* to be an MDS as in Escanciano and Lobato (2009).

`seriestype(var)` corresponds to backtesting VAR. *varname* assumes a first-order autoregressive mean model and a conditional variance model with squared residuals with lags of order 1 and variance components with lags of order 1 [AR(1)–GARCH(1,1)] model with Student's t innovations when deriving the estimation effects.

`seriestype(es)` corresponds to backtesting ES, and *varname* assumes an AR(1)–GARCH(1,1) model with Student's t innovations when deriving the estimation effects.

`cl(#)` specifies the coverage level of VAR and ES. The default is `cl(0.05)`.

`nobs(#)` specifies the in-sample size when backtesting VAR and ES.

4.3 Remarks

One needs to `tsset` the data before using `dbptest` and `rtau`.

Automatic Q test

`dbptest` implements a data-driven Box–Pierce test for serial correlations. The test automatically chooses the order of autocorrelations. The command reports not only the usual outputs of the Box–Pierce test as `wntestq`, that is, the Q statistics and the corresponding p-value, but also the automatic number of lags chosen.

Estimating τ_j

`rtau` estimates the asymptotic variances of autocorrelations when necessary. This includes

1. MDS data; and
2. backtesting ES and VAR with estimation effects.

`cl(#)` and `nobs(#)` are required only when `seriestype(var)` or `seriestype(es)` is specified.

4.4 Stored results

Automatic Q test

`dbptest` stores the following in `r()`:

Scalars

`r(stat)`	Q statistic	`r(lag)`	the number of lags
`r(p)`	probability value		

Estimating τ_j

`rtau` stores the following in `e()`:

Matrix

`e(tau)`	variances of autocorrelations

4.5 Example

To illustrate the use of the two commands, we consider the DAX Index return data from January 1, 1997, to June 30, 2009 as in Du and Escanciano (2017). The in-sample period is from January 1, 1997, to June 30, 2007. The out-of-sample period is from July 1, 2007, to June 30, 2009, which is the financial crisis period.

We use the in-sample data to fit an AR(1)–GARCH(1,1) model with Student's t innovations. After getting the estimates for u_t, $h_t(\alpha)$, and $H_t(\alpha)$ using the out-of-sample data, we implement the conditional backtests for VAR and ES using the new `dbptest` command.

Without estimation effects

Here we carry out the AQ test (3) without considering the estimation effects, that is, $\tau_j = 1$.

```
. set matsize 509
. import delimited "dax.csv", varnames(1)
(3 vars, 3,168 obs)
. scalar nin = 2658
. scalar nout= 509
. scalar ninout = nin + nout
. keep lret date
. drop in 1
(1 observation deleted)
. generate sin = (_n <= nin)
. generate sout= (_n > nin & _n<=ninout)
. keep if _n<=ninout
(0 observations deleted)
. generate date_num=_n
. tsset date_num
        time variable:  date_num, 1 to 3167
                delta:  1 unit
. arch lret if sin==1, noconstant arch(1) garch(1) ar(1) distribution(t)
  (output omitted)
. matrix define awab=e(b)
. matrix define covm=e(V)
. scalar ahat   = awab[1,1]
. scalar alphat = awab[1,2]
. scalar bethat = awab[1,3]
. scalar omghat = awab[1,4]
. scalar vhat = round(e(tdf))
. predict resin, residuals
. generate resout = resin if sin==0
(2,658 missing values generated)
. replace resin=. if sin==0
(509 real changes made, 509 to missing)
```

```
. predict convar, variance
. generate fith = convar if sin==0
(2,658 missing values generated)
. replace convar=. if sin==0
(509 real changes made, 509 to missing)
. generate fitsig = sqrt(fith)
(2,658 missing values generated)
. generate epsj = resout/fitsig
(2,658 missing values generated)
. generate utj = t(vhat, epsj*sqrt(vhat/(vhat-2)))
(2,658 missing values generated)
. scalar jalp = 0.1
. generate h = (utj <= jalp) if sout==1
(2,658 missing values generated)
. generate utalp = utj - jalp if h == 1
(3,108 missing values generated)
. replace utalp=0 if utalp==. & sout==1
(450 real changes made)
. generate H =-utalp/jalp
(2,658 missing values generated)
. dbptest H, mu(0.05)
Automatic Portmanteau test for serial correlation
----------------------------------------------------------
 Variable: H
----------------------------------------------------------
 Portmanteau (Q) statistic                   =    2.8417
 Prob > chi2(1)                              =    0.0918
 The number of lag(s) (from 1 to 23)         =         1
----------------------------------------------------------
```

The displayed results are for cumulative violations at a 10% coverage level, that is, $H_t(0.1)$. Under the correct model specification, we have $E\{H_t(\alpha)\} = \alpha/2$, so we set `mu()` to be 0.05. We get an AQ statistic of 2.8417 and a p-value of 0.0918. Hence, the ES model is rejected at a 10% significance level. It also reports the number of lags chosen, which is 1 in this case.

Likewise, we carry out the conditional backtest for VAR using $h_t(\alpha)$. Following the rule of thumb that the coverage level for ES is twice (or approximately twice) that of VAR, we examine the autocorrelations of $h_t(0.05)$.

```
. capture drop h
. scalar jalp=0.05
. generate h = (utj <= jalp) if sout==1
(2,658 missing values generated)
. dbptest h, mu(0.05)
Automatic Portmanteau test for serial correlation

 Variable: h

 Portmanteau (Q) statistic                  =    0.7972
 Prob > chi2(1)                             =    0.3719
 The number of lag(s) (from 1 to 23)        =         1
```

We now get an AQ statistic of 0.7972 and a p-value of 0.3719, so we fail to reject the VAR model.

With estimation effects

To account for the estimation effects, we use the `rtau` command to estimate τ_j before we run the `dbptest` command.

```
. rtau lret,  nlags(15) seriestype(es) cl(0.1) nobs(2658)

Asymptotic Variances of Autocorrelations

    Order      Tau for ES
      1         1.0027636
      2         1.0192228
      3         1.0192343
      4         1.004399
      5         1.0030891
      6         1.0021455
      7         1.0137747
      8         1.0016341
      9         1.0094143
      10        1.0012676
      11        1.0011319
      12        1.0080588
      13        1.0077699
      14        1.0033674
      15        1.0017961

. matrix Tau_ES = e(tau)
. dbptest H, mu(0.05) tauvector(Tau_ES)
Automatic Portmanteau test for serial correlation

 Variable: H

 Portmanteau (Q) statistic                  =    2.8338
 Prob > chi2(1)                             =    0.0923
 The number of lag(s) (from 1 to 15)        =         1
```

```
. rtau lret,  nlags(15) seriestype(var) cl(0.05) nobs(2658)
─────────────────────────────────────────────
Asymptotic Variances of Autocorrelations
─────────────────────────────────────────────
    Order      Tau for VaR
      1         1.01014
      2         1.0029985
      3         1.0023986
      4         1.0023737
      5         1.0027832
      6         1.0021056
      7         1.001556
      8         1.0014201
      9         1.0011457
      10        1.0009844
      11        1.001431
      12        1.0013224
      13        1.0013889
      14        1.0009824
      15        1.0011676
─────────────────────────────────────────────
. matrix Tau_VaR = e(tau)
. dbptest h, mu(0.05) tauvector(Tau_VaR)
Automatic Portmanteau test for serial correlation
─────────────────────────────────────────────────────────
 Variable: h
─────────────────────────────────────────────────────────
 Portmanteau (Q) statistic                 =    0.7892
 Prob > chi2(1)                            =    0.3743
 The number of lag(s) (from 1 to 15)       =         1
─────────────────────────────────────────────────────────
```

Notice that the in-sample size here is 2,658. The AQ test statistics for ES and VAR here are slightly lower than those without estimation effects. The test conclusions remain the same, although the *p*-values are slightly higher than before.

5 References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Berkowitz, J., P. Christoffersen, and D. Pelletier. 2011. Evaluating value-at-risk models with desk-level data. *Management Science* 57: 2213–2227.

Box, G. E. P., and D. A. Pierce. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65: 1509–1526.

Christoffersen, P. 2009. Value–at–risk models. In *Handbook of Financial Time Series*, ed. T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, 753–766. Berlin: Springer.

Christoffersen, P. F. 1998. Evaluating interval forecasts. *International Economic Review* 39: 841–862.

Davies, N., and P. Newbold. 1979. Some power studies of a portmanteau test of time series model specification. *Biometrika* 66: 153–155.

Davies, N., C. M. Triggs, and P. Newbold. 1977. Significance levels of the Box-Pierce portmanteau statistic in finite samples. *Biometrika* 64: 517–522.

Delgado, M. A., and C. Velasco. 2011. An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *Journal of the American Statistical Association* 106: 946–958.

Du, Z., and J. C. Escanciano. 2017. Backtesting expected shortfall: Accounting for tail risk. *Management Science* 63: 940–958.

Escanciano, J. C., and I. N. Lobato. 2009. An automatic Portmanteau test for serial correlation. *Journal of Econometrics* 151: 140–149.

Escanciano, J. C., I. N. Lobato, and L. Zhu. 2013. Automatic specification testing for vector autoregressions and multivariate nonlinear time series models. *Journal of Business & Economic Statistics* 31: 426–437.

Escanciano, J. C., and J. Olmo. 2010. Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics* 28: 36–51.

Francq, C., R. Roy, and J.-M. Zakoïan. 2005. Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100: 532–544.

Inglot, T., and T. Ledwina. 2006a. Data driven score tests of fit for semiparametric homoscedastic linear regression model. Technical Report 665, Institute of Mathematics Polish Academy of Sciences.

———. 2006b. Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and its Applications* 417: 124–133.

Jorion, P. 2007. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. New York: McGraw–Hill.

Kupiec, P. H. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3: 73–84.

Li, W. K. 2004. *Diagnostic Checks in Time Series*. Boca Raton, FL: Chapman & Hall/CRC.

Li, W. K., and A. I. McLeod. 1981. Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society, Series B* 43: 231–239.

Ljung, G. M., and G. E. P. Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65: 297–303.

Newbold, P. 1980. The equivalence of two tests of time series model adequacy. *Biometrika* 67: 463–465.

Quenouille, M. H. 1947. A large-sample test for the goodness of fit of autoregressive schemes. *Journal of the Royal Statistical Society* 110: 123–129.

Rosenblatt, M. 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23: 470–472.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.

Yule, G. U. 1926. Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society* 89: 1–63.

About the authors

Guangwei Zhu is a lecturer in the Institute of Chinese Financial Studies at the Southwestern University of Finance and Economics, Chengdu, China. His research is funded by the Excellent Doctoral Dissertation Foundation of Southwestern University of Finance and Economics.

Zaichao Du is a professor in the Research Institute of Economics and Management at the Southwestern University of Finance and Economics, Chengdu, China. His research is funded by the National Natural Science Foundation of China, 71401140.

Juan Carlos Escanciano is a professor in the Department of Economics at Indiana University, Bloomington, IN. His research is funded by the Spanish Plan Nacional de I+D+I, reference number ECO2014-55858-P.