# The Value-at-Risk Backtest Package*

Thor Pajhede

July 1, 2016

**Abstract**

Evaluating Value-at-Risk forecasts is a common feature of papers in financial econometrics. However, there is a tendency to use older backtests with low power, despite several newer tests having much improved power as well as power against more general forms of dependence. This paper describes the Value-at-Risk Backtest package for Matlab which contain most existing backtests. In addition, the Monte Carlo method of Dufour [2006] is implemented to allow exact small sample inference.

**Keywords:** Value-at-Risk, Backtesting, Risk Management, Matlab

## Contents



*The newest version of the Value-at-Risk Backtest package is available from the authors website at econ.ku.dk/pajhede/backtest.

# 1 Introduction

The package is written in the Matlab language, though some functions are implemented in C++ to increase computational speed[1]. Each backtest described in this paper is implemented in a single function, the function has a header describing its usage, inputs, outputs and a small example. To use the backtest, simply add the backtest toolbox to the Matlab directory and call the functions as described in their header, a number of examples are presented in Section3.

The following backtests are included in the package:

- Proportion of failures test of Kupiec [1995], function fPFtest fjointtest
- Markov backtest of Christoffersen [1998], function fMarkovtest
- Generalized Markov backtests of Pajhede [2014], function fGeneralizedMarkovtest and function fDuration-Markovtest
- Duration backtest of Christoffersen [2004], function fDurCtest
- Discrete duration backtest of Haas [2006], function fDurDtest
- GMM-J test of Candelon et al. [2011], function fGMMtest
- Dynamic quantile backtest of Engle and Manganelli [1999], function fDynamicQuantileTest

For descriptions of these see 2 and for instructions on using the function see the function header or section 3.

There exist a number of caveats to be aware of when conducting Value-at-Risk (VaR) backtests.

- The test due to **?** and Pajhede [2014] can be implemented in various ways, however because numbers between zero and one are raised to potentially very large powers this can cause floating point errors. We use expressions robust to this.
- If a backtest function is given data that does not fit its data requirement a warning is returned to the user and a NaN value is returned as test value.
- Due to the discrete nature of the data the asymptotic distributions of the test statistics is often not a good approximation for the finite sample distribution, see Christoffersen [2004]. For this reason the package implements the Monte Carlo technique of Dufour [2006] for each backtest. This allows the users to easily obtain valid p-values regardless of the observations available.

# 2 Hit-sequence Based Backtesting

Let $R_t$ denote the realization of a return of an asset or a portfolio of assets at time $t$. The *ex ante* VaR for time $t$ and *coverage rate* $p$, denoted as $\mathrm{VaR}_{t|t-1}(p)$, conditional on all information, $\mathcal{F}_{t-1}$, available at time $t-1$ (for example past returns and macroeconomic indicators) is defined as the $p$'th conditional quantile of the distribution of $R_t$:

$$P(R_t < \mathrm{VaR}_{t|t-1}(p)|\mathcal{F}_{t-1}) = p, \qquad t = 1, ..., T.$$

Since its introduction in the 90s Value-at-Risk (VaR), has become widely used when reporting aggregate market risk. Typically the coverage rate used is 1% or 5%.

Backtesting is the procedure of comparing *realized* losses to the *forecasted* VaR. To implement backtesting of a VaR forecast define the *hit-sequence*, $\{I_t\}_{t=1}^T$, as follows:

$$I_t := 1\left(R_t < \mathrm{VaR}_{t|t-1}(p)\right), \qquad t = 1, ...T \tag{2.1}$$

Where $1(\cdot)$ is the indicator function. Thus, the hit-sequence is by construction a binary time series indicating whether a loss at time $t$ greater than the VaR, termed a *violation* or a *hit*, was realized.

A VaR forecast is valid, in the sense of actually having forecasted the desired quantile, only if the associated hit-sequence satisfies the following criteria due to Christoffersen [1998]:

[1] Matlab functions are included in these cases to facilitate documentation.

- **The unconditional coverage criteria:** The unconditional probability of a violation must be exactly equal to the coverage rate $p$:

$$H_{UC}: \ P(I_t = 1) = p$$

- **The independence criteria**: The conditional probability of a violation must be constant:

$$H_{Ind}: \ P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1)$$

Combining these criteria we obtain the conditional coverage criteria:

- **The conditional coverage criteria**: The probability of a violation must be constant and equal to the coverage rate:

$$H_{CC}: \ P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1) = p$$

It follows, see Christoffersen [1998], that the hit-sequence of a valid VaR forecast, is in fact a sequence of i.i.d Bernoulli distributed variables:

$$I_t \underset{i.i.d}{\sim} Bernoulli(p), \qquad t = 1, ..., T. \tag{2.2}$$

The waiting time between hits, termed the *duration,* is given by:

$$D_i = t_i - t_{i-1},$$

where $t_i$ denotes the time of violation number $i$. It follows that the durations is a sequence of i.i.d geometrically distributed variables:

$$D_i \underset{i.i.d}{\sim} Geometric(p) \tag{2.3}$$

The backtests described in this paper are based on either the distribution of the hit-sequence or the distribution of the duration, as given by equations (2.1) and (2.3).

## 2.1 The Generalized Markov Framework

Christoffersen [1998] used a first order Markov model to derive backtests, while Pajhede [2014] extends the model to $k$'th order dependence which contain the tests of Christoffersen [1998] as a special case. The hit-sequence, (2.2), is modeled as a $k$'th order Markov chain:

$$I_t|\mathcal{F}_{t-1,k} \underset{i.i.d}{\sim} Bernoulli(p_t(\theta)), \quad \mathcal{F}_{t-1,k} = I_{t-1}, ..., I_{t-k}, \qquad t = 1, ..., T. \tag{2.4}$$

The likelihood for this model conditioned on $k$ observations prior to $t = 1$ fixed, is given by $\mathcal{L}_T(\theta) = \prod_{t=1}^{T} p_t(\theta)^{I_t}(1 - p_t(\theta))^{1-I_t}$. The transition probabilities of (2.4), $p_t(\theta)$, can be quite general, but this also yields a large parameter vector. Instead it is suggested that the probability of a hit at time $t$, $p_t(\theta)$ be a function of a hit occurring in the $k$ latest observations, reducing the number of parameters to two:

$$p_t(\theta) = J_{t-1}p_E + (1 - J_{t-1})p_S, \quad J_{t-1} := 1\left(\sum_{i=1}^{k} I_{t-1} > 0\right). \tag{2.5}$$

The likelihood is then given by, $\mathcal{L}_T(\theta) = (1 - p_S)^{T_{00}} p_S^{T_{01}}(1 - p_E)^{T_{10}} p_E^{T_{11}}$, where $T_{ij} = \sum_{t=1}^{T} I_t J_{t-1}$. Defining the unrestricted estimator, the estimated restricted under $H_{Ind}$ estimator restricted under $H_{CC}$ as

$$\hat{\theta} = (\hat{p}_S, \hat{p}_E)', \quad \tilde{\theta} = H\hat{\phi} \quad \text{and} \quad \theta_0 = Hp.$$

Where $H = (1, 1)'$, $\hat{p}_S = T_{01}/(T_{01} + T_{00})$, $\hat{p}_E = T_{11}/(T_{11} + T_{10})$ and $\hat{\phi} = (T_{01} + T_{11})/(T_{01} + T_{11} + T_{00} + T_{10})$. It follows that the test of independence with asymptotics for $T \to \infty$ and restriction in parenthesis, are given by

$$\begin{aligned} Q_{G-Ind}(\theta = H\phi) = -2\{&log(1 - \hat{\phi})(T_{00} + T_{10}) + log(\hat{\phi})(T_{01} + T_{11}) \\ &- log(1 - \hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1 - \hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \\ &\xrightarrow{d} \chi^2(1) \end{aligned} \tag{2.6}$$

the test of conditional coverage with asymptotics for $T \to \infty$, by

$$\begin{aligned}Q_{G-CC}(\theta = Hp) = &-2\{log(1-p)(T_{00}+T_{10}) + log(p)(T_{01}+T_{11}) \\ &- log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1-\hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \\ &\xrightarrow{d} \chi^2(2)\end{aligned} \tag{2.7}$$

and the test of unconditional coverage with asymptotics for $T \to \infty$, by

$$Q_{G-UC}(H\phi = Hp) = Q_{G-CC}(\theta = Hp) - Q_{G-Ind}(\theta = H\phi) \xrightarrow{d} \chi^2(1)$$

Tests derived from the first specification of equation (2.5) are referred to as *generalized Markov tests*. The test of unconditional coverage, $Q_{G-UC}(H\phi = Hp)$, was first derived in Kupiec [1995]. We require $T_{11} + T_{10} > 0$ corresponding to at least 1 hit in the hit-sequence which must occur before $T$, to calculate the test.

An alternative specification of the transition probability is that the probability of a hit at time $t$, $p_t(\theta)$, is a function of the number of observations since the last hit (the *duration*) in the preceding $k$ lags, after which the probability is a constant. This reduces the parameters of the model to $k+1$, or equivalently,

$$p_t(\theta) = J(1)_{t-1}p_{E1} + ... + J(k)_{t-1}p_{Ek} + (1 - \sum_{i=1}^{k} J(i)_{t-1})p_S, \tag{2.8}$$

where $J(1)_{t-1} := 1\,(I_{t-1} = 1)\ \ , ...,\ J(k)_{t-1} := 1\,(I_{t-1} = 0, ..., I_{t-k} = 1)$. The likelihood is then given by, $\mathcal{L}_T(\theta) = (1-p_S)^{T_{00}} p_S^{T_{01}} \prod_{i=1}^{k}(1-p_{Ei})^{T_{10}(i)} p_{Ei}^{T_{11}(i)}$, where $T_{10}(i) = \sum_{t=1}^{T}(1-I_t)J(i)_{t-1}$. Defining the estimated unrestricted estimator, the estimated restricted under $H_{Ind}$ estimator restricted under $H_{CC}$ as

$$\hat{\theta} = (\hat{p}_S, \hat{p}_{E1}, ..., \hat{p}_{Ek})', \quad \tilde{\theta} = H\hat{\phi} \quad \text{and} \quad \theta_0 = Hp.$$

Where $H = (1, ..., 1)'$ is a $k \times 1$ vector, with maximum likelihood estimates (MLE) $\hat{\phi}$ and $\hat{p}_S$ unchanged while $\hat{p}_{Ei} = T_{11}(i)/(T_{11}(i) + T_{10}(i))$, for $i = 1, ..., k$. It follows that the test of independence with asymptotics for $T \to \infty$, are given by

$$\begin{aligned}Q_{D-Ind}(\theta = H\phi) &= -2\Big(log(1-\hat{\phi})(T_{00}+T_{10}) \times log(\hat{\phi})(T_{01}+T_{11}) - log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} \\ &- \sum_{i=1}^{k} log(1-\hat{p}_{Ei})T_{10}(i) - \sum_{i=1}^{k} log(\hat{p}_{Ei})T_{11}(i)\Big), \\ &\xrightarrow{d} \chi^2(k)\end{aligned} \tag{2.9}$$

the test of conditional coverage with asymptotics for $T \to \infty$, by

$$Q_{D-CC}(\theta = Hp) = -2(log(1-p)(T_{00}+T_{10}) \times log(p)(T_{01}+T_{11}) - log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01}$$

$$- \sum_{i=1}^{k} log(1-\hat{p}_{Ei})T_{i0}(i) - \sum_{i=1}^{k} log(\hat{p}_{Ei})T_{i1}), \tag{2.10}$$

$$\xrightarrow{d} \chi^2(k+1) \tag{2.11}$$

and the test of unconditional coverage with asymptotics for $T \to \infty$, by

$$Q_{D-UC}(H\phi = Hp) = Q_{D-CC}(\theta = Hp) - Q_{D-Ind}(\theta = H\phi) \xrightarrow{d} \chi^2(1)$$

Tests derived from the second specification of equation (2.8) are referred to as Markov duration tests. See Pajhede [2014] for details. For $k = 1$ the tests, of either specification, reduce to the tests of Christoffersen [1998].

## 2.2 Duration Based Tests

Since the durations follows a geometric distribution, see equation (2.3), this implied a constant hazard rate $P(D_i = d|D_i \geq d) = p$. The duration based backtests are then constructed by modeling the distribution of $D_i$ by some other distribution with a non-constant hazard rate but which nests the geometric distribution under some restrictions which can be tested using likelihood ratio tests.

A small complication is that the durations are subject to right and left censoring, specifically if $I_1 = 1$, then $D_1$ is the time between that 1 and the next 1. If on the other hand $I_1 = 0$, then the $D_1$ is the time until the first 1 and is a left censored observation. If $I_T = 0$, then the last duration, $D_{N(T)}$, is the time between the last hit (which we designate $N(T)$) and the remaining length of the hit sequence and is considered a right censored observation. This leads to the log-likelihood:

$$ln(\mathcal{L}_{\mathcal{T}}(\theta)) = C_1 ln(S(D_1)) + (1-C_1) ln(f(D_1)) + \sum_{i=2}^{N(T)-1} ln(f(D_i)) + C_{N(T)} ln(S(D_{N(T)})) + (1-C_{N(T)}) ln(f(D_{N(T)}))$$

Where $\theta$ are the parameters of the survival and probability mass functions and $S(D_i) = 1 - F(D_i)$ is the survival function. The censoring series $C_i$, indicates if there is censoring ($C_i = 1$) or no censoring ($C_i = 0$) for hit $i$. If the hit sequence starts with a 0 indicating no hit, then $C_1 = 1$ because the duration will be censored. However, if the hit sequence starts with a violation, then $C_1 = 0$. The procedure is similar for the last observation. Other $C_j$ values are always 0.

Christoffersen [2004] introduced the first duration test. The durations are modeled using the (continuous) Weibull distribution which contain the exponential distribution, the continuous analogue of the geometric distribution, as a special case. The Weibull distribution is able to model increasing and decreasing hazard rates and has PDF and CDF as follows:

$$f(D; a, b) = ba^{-b} D^{b-1} e^{-(\frac{D}{a})^b} \qquad , \qquad F(D; a, b) = 1 - e^{-(\frac{D}{a})^b} \tag{2.12}$$

Independence is tested using a likelihood-ratio statistic of the restriction $H_{Ind}: \ b = 1$ which reduces the Weibull distribution to the exponential distribution:

$$Q_{C-Weibull-Ind} = -2\left[log(\mathcal{L}_T(a, 1)) - log(\mathcal{L}_T(a, b))\right]$$

This test will not have standard $\chi^2(1)$ asymptotics due to the use of continuous distributions. Instead simulation methods are required for finding the asymptotic distribution. The tests require two or more durations, at least one of which is not censored, to be calculated.

Haas [2006] suggested a discrete version of the duration test with the discrete Weibull distribution of Nakagawa and Osaki [1975] and nests the geometric distribution. A re-parametrized version of the discrete Weibull is recommended by the author based on ease of estimation. Its PMF and CDF is:

$$g(D; a, b) = e^{-a^b (D-1)^b} - e^{-a^b d^b} \qquad , \qquad G(D; a, b) = 1 - e^{-a^b D^b} \tag{2.13}$$

The independence criteria is tested using a likelihood-ratio statistic by the restriction $H_{Ind}: \ b = 1$ and conditional coverage by the restrictions $H_{cc}: \ b = 1, \, a = -log(1-p)$. The test statistics for testing IND and CC with asymptotics for $T \to \infty$ are given as

$$Q_{D-Weibull-Ind} = -2\left[log(\mathcal{L}_T(a, 1)) - log(\mathcal{L}_T(a, b))\right] \xrightarrow{d} \chi^2(1)$$

and

$$Q_{D-Weibull-CC} = -2\left[log(\mathcal{L}_T(-log(1-p), 1)) - log(\mathcal{L}_T(a, b))\right] \xrightarrow{d} \chi^2(2).$$

## 2.3 Dynamic Quantile Test

Rather than specify a full distributional model for the hit-sequence,Engle and Manganelli [1999] model it as a regression on the demeaned hit-sequence using the lagged hit-sequence as regressors:

$$I_t - p = \delta + \sum_{k=1}^{K} \beta_k I_{t-k} + \epsilon_t \tag{2.14}$$

Where the parameter vector is given by $\theta = (\delta, \beta_1, ..., \beta_k)'$ which is estimated by ordinary least squares as $\hat{\theta}$ and with $Z$ the design matrix of covariates. The Wald test statistic for the hypothesis of conditional coverage $H_0: \ \delta = \beta_1, ..., \beta_k = 0$ then has closed form expression, with asymptotic distribution for $T \to \infty$, given by:

$$DQ_{CC} = \frac{\hat{\theta}' Z' Z \hat{\theta}}{p(1-p)} \xrightarrow{d} \chi^2(k+1) \tag{2.15}$$

Dumitrescu et al. [2012] suggest replacing the linear model with a probit or logit type model, but find only slight increases in power at the cost of a much more complicated implementation.

## 2.4 GMM Tests

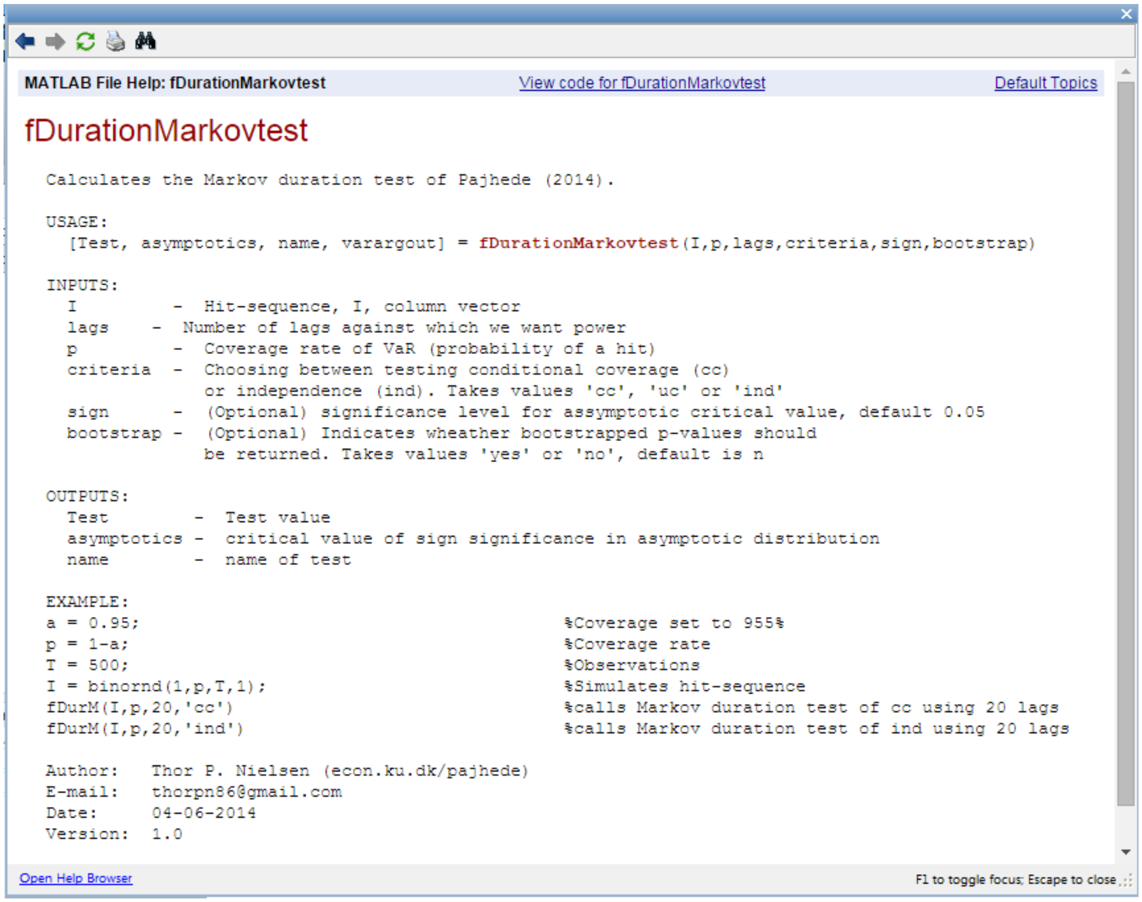
Candelon et al. [2011] suggested a GMM-J test of the durations, comparing the $k$ first theoretical moments of the geometric distribution to those estimated from the observed durations. For a geometric distribution with success probability $p$, a series of orthonormal polynomials are given by the following recursive relation, $\forall d \in N$:

$$M_{j+1}(d;p) = \frac{(1-p)(2j+1) + p(j-d+1)}{(j+1)\sqrt{1-p}} M_j(d;p) - \frac{j}{j+1} M_{j-1}(d;p) \tag{2.16}$$

For any $j \in N$, with $M_{-1}(d;p) = 0$ and $M_0(d;p) = 1$. These polynomials have expectation $E(M_{j+1}) = 0$, are asymptotically independent with unit variance and are known to converge in distribution when squared as $\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_j(d_i, p)\right]^2 \xrightarrow{d} \chi^2(1)$ for $N \to \infty$. From this it follows that the test statistic, with asymptotics for $N \to \infty$, using $k$ such moments can be expressed as

$$J(k) = \left( \frac{1}{\sqrt{n}} \sum_{i}^{N} M(d_i;p) \right)' \left( \frac{1}{\sqrt{n}} \sum_{i}^{N} M(d_i;p) \right) \xrightarrow{d} \chi^2(k) \tag{2.17}$$

Where $M$ is a vector whose elements are the $k$ orthonormal polynomials $M_j(d_i, \alpha)$. When setting $k = 1$, this will test the unconditional coverage criteria while setting $k > 1$ tests the conditional coverage criteria. A test of independence can be found by replacing $p$ with the ML estimator in the geometric distribution, $\hat{p} = \overline{I}$. The GMM tests require at least 1 violation to be computed and the authors find the highest empirical power in simulations for $k = 5$.



MATLAB File Help: fDurationMarkovtest    View code for fDurationMarkovtest    Default Topics

fDurationMarkovtest

```
Calculates the Markov duration test of Pajhede (2014).

USAGE:
  [Test, asymptotics, name, varargout] = fDurationMarkovtest(I,p,lags,criteria,sign,bootstrap)

INPUTS:
  I         -  Hit-sequence, I, column vector
  lags    -  Number of lags against which we want power
  p         -  Coverage rate of VaR (probability of a hit)
  criteria  -  Choosing between testing conditional coverage (cc)
               or independence (ind). Takes values 'cc', 'uc' or 'ind'
  sign      -  (Optional) significance level for assymptotic critical value, default 0.05
  bootstrap -  (Optional) Indicates wheather bootstrapped p-values should
               be returned. Takes values 'yes' or 'no', default is n

OUTPUTS:
  Test        -  Test value
  asymptotics -  critical value of sign significance in asymptotic distribution
  name        -  name of test

EXAMPLE:
a = 0.95;                                        %Coverage set to 955%
p = 1-a;                                         %Coverage rate
T = 500;                                         %Observations
I = binornd(1,p,T,1);                            %Simulates hit-sequence
fDurM(I,p,20,'cc')                               %calls Markov duration test of cc using 20 lags
fDurM(I,p,20,'ind')                              %calls Markov duration test of ind using 20 lags

Author:   Thor P. Nielsen (econ.ku.dk/pajhede)
E-mail:   thorpn86@gmail.com
Date:     04-06-2014
Version:  1.0
```

Open Help Browser    F1 to toggle focus; Escape to close

Figure 3.1: Header for the function fDurationMarkovtest, which implements the backtests of Pajhede [2014].

# 3 Examples and Usage

Once the package is added to the workspace, the headers can be used to describe how each backtest is used. The header in each function is arranged as follows as illustrated in Figure 3.1 which shows the header for the generalized duration backtests of Pajhede [2014].

## 3.1 Example: The Backtest of Pajhede [2014]

The package includes a small minimum working example in the file MinimumExample.m, running this code simulates the hit-sequence of a well-specified VaR forecast and calls the backtests of Pajhede [2014] both with and without bootstrapped p-values. The code is similar to the small working examples included in each test function and is given below:

frame=single

Listing 1: Full Matlab code of the small example included in the package as MinimumExample.m

```
%Add toolbox to workspace (insert the path to where you have the toolbox)
addpath(genpath('insert path to toolbox here'));

%Simulate a hit-sequence of a 5% VaR forecast
p = 0.05;                                    %Coverage rate
T = 500;                                     %Number of observations
I = binornd(1,p,T,1);                        %Simulates hit-sequence

%Calls "generalized Markov" test of CC function using 10 lags
Test = fGeneralizedMarkovtest(I,p,10,'cc');

%Uses bootstrap of Dufour to get p-value at and sets size to 5%
[Test, asymptotics, name, pval] = fGeneralizedMarkovtest(I,p,10,'cc',0.05,'yes');

%Presents results
disp(['Test Name:   ' name]);
disp(['Test statistic = ' num2str(Test)]);
disp(['Asymptotic 95% Critical value = ' num2str(asymptotics)]);
disp(['Asymptotic p-value  = ' num2str(1-chi2cdf(Test,2))]);
disp(['Bootstrapped p-value  = ' num2str(pval)]);
```

# References


Bertrand Candelon, Gilbert Colletaz, Christophe Hurlin, and Sessi Tokpavi. Backtesting Value-at-Risk: A GMM Duration-Based Test. *Journal of Financial Econometrics*, 9(2):314–343, Spring 2011.

Peter Christoffersen. Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2(1):84–108, 2004.

Peter F Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–62, 1998.

Jean-Marie Dufour. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133(2):443–477, 2006.

Elena-Ivona Dumitrescu, Christophe Hurlin, and Vinson Pham. Backtesting value-at-risk: From dynamic quantile to dynamic binary tests. Technical report, HAL, 2012.

Robert F. Engle and Simone Manganelli. Caviar: Conditional value at risk by quantile regression. Nber working papers, National Bureau of Economic Research, Inc, September 1999.

Markus Haas. Improved duration-based backtesting of value-at-risk. *Journal of Risk*, 8(2):17–38, 2006.

Paul H. Kupiec. Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.), 1995.

Nakagawa and Osaki. The discrete weibull distribution. *IEEE Transactions on Reliability*, 24, 1975.

Thor Pajhede. Backtesting value-at-risk: The generalized markov test. Economics working papers, University of Copenhagen, Department of Economics, 1 2014.