Poster session

THE DISTRIBUTION OF CONDITIONAL ON REALISED VOLATILITY FINANCIAL RETURNS

***Max V. Moldovan*[1], *Andrew C. Worthington*[2], *Nicholas A. Nechval*[3], *Helen Higgs*[4]**

[1] ESH Group Ltd.,
6-450 Ezermalas Street, Riga, Latvia, LV-1016
Ph: (+371) 9 82 69 66. Fax: (+371) 7 518 519. E-mail: max@esh.lv

[2] School of Economics and Finance, Queensland University of Technology,
2 George Street, Brisbane, Australia, Qld 4001
Ph: (+61) 07 3864 2658. Fax: (+61) 07 3864 1500. E-mail: a.worthington@qut.edu.au

[3] Mathematical Statistics Department, University of Latvia,
19 Raina Blvd, Riga, Latvia, LV-1050
Ph: (+371) 7034701. Fax: (+371) 7034702. E-mail: nechval@junik.lv

[4] School of Economics and Finance, Queensland University of Technology,
2 George Street, Brisbane, Australia, Qld 4001
Ph: (+61) 07 3864 4252. Fax: (+61) 07 3864 1500. E-mail: h.higgs@qut.edu.au

Abstract

The following study explores the stochastic properties of financial time series. Special attention is devoted to the unusual two-peaked shape of the empirically observed distribution of the conditional on realised volatility financial returns. The performed analysis leads to the conclusion that the conditional on realised volatility returns are distributed with the specific previously undocumented distribution. The probability density that represents this distribution is derived, characterised and applied for improving the forecasts of Value at Risk.

Key words: Realised volatility, High-frequency data, Value at Risk, Two-component effect, Monte Carlo simulation

1. Introduction

A distribution of financial returns plays a central role in financial econometrics and its applications. It is important to distinguish between unconditional and conditional return distributions. The unconditional distribution reflects stochastic characteristics of returns observed in the markets. Since unconditional returns can be directly observed, they have been well studied. The situation is more complicated in the case of conditional returns. The term conditional generally refers to conditioning returns on underlying volatility. Since financial volatility is not directly observable, volatility estimates are used for conditioning. Realised volatility is an unbiased, consistent and highly efficient estimator of actual financial volatility. In this study realised volatility estimates are applied to conditioning financial return series. Despite the theoretical expectations, it has been found that conditional on realised volatility returns are not standard normally distributed but follow the specific previously undocumented two-peaked distribution. The probability density that represents this distribution is derived and applied to forecasting the future return probability density quintiles, or so-called Value at Risk.

The paper is structured as follows. Section 2 presents the traditional return decomposition and introduces the realised volatility estimator. In Section 3 the unusual previously undocumented distribution of conditional on realised volatility return series is documented and probability density that represents this distribution suggested. Section 4 demonstrates how the new probability density can be used in the financial model with application to Value at Risk forecasting. Section 5 summarises the results of the study.

2. Return Decomposition and Realised Volatility

We assume that daily return generating process can be expressed as follows.

$$r_t = \sigma_t z_t \tag{1}$$

where $r_t = \ln(P_t) - \ln(P_{t-1})$, P_t is a last price observed over day t, and $z_t \sim N(0,1)$ and σ_t are a stochastic component and standard deviation of return over day t respectively.

This view on the return generating process relies on the Efficient Market Hypothesis exposed by Fama [1] and widely accepted among academics and practitioners (see Campbell et al. [2]). Following this, the conditional on volatility returns $z_t = r_t / \sigma_t$ are expected to be standard normally distributed: $z_t \sim N(0,1)$. For such conditioning it is required to estimate volatility over each day t. The recent developments in financial econometrics allow estimating the variability of financial returns using more frequently sampled intra-period data in place of traditional multi-period datasets. Andersen and Bollerslev [3] and Andersen et al. [4] related intra-period accumulation of squared returns to the mathematical concept of quadratic variation and proposed the realised volatility estimator. Realised volatility is an unbiased, consistent and efficient volatility estimator that does not requires information outside an estimation interval. Realised volatility in day t based on k intraday return observations is expressed as follows.

$$v_{t,k} = \sum_{n=1}^{k} x_{t,n}^2 \tag{2}$$

where $v_{t,k}$ is a realised volatility estimate computed by summing squared intraday returns $x_{t,n}^2$ that sampled k times over equally spaced time intervals within day t.

The especially useful property of the realised volatility estimator is consistency: $v_{t,k} \to \sigma_t^2$ as $k \to \infty$. However, the value of k should not be the highest available in order to avoid the domination of the microstructure bias, though high enough to minimise the measurement error. For selection of the optimal sampling frequency Andersen et al. [5] suggested the volatility signature plot approach, which is illustrated in Appendix 2 on the example of data described in Appendix 1. It has been found that the optimal sampling frequency for series considered in the study is $k = 8$. Hence, the realised volatility series $v_{t,8}$ has been computed for return conditioning.

3. The Two-Component Effect and J Distribution

The conditional return series has been computed as follows.

$$z_t = r_t / \sigma_{t,8} \tag{3}$$

where the log return r_t over day t is standardised by the realised standard deviation $\sigma_{t,8} = \sqrt{v_{t,8}}$.

According the theoretical view on the return generating process given by (1), conditional returns should be approximately standard normally distributed $z_t \sim N(0,1)$. Indeed, Andersen et al. [6] performed the standardisation (3) and found that distributions of the resulted series are "remarkably close to a standard normal". Authors interpreted this result as consistent with Mixture-of-Distributions hypothesis suggested by Clark [7], completely ignoring platykurtosis, which clearly violets normality. In this study it has been found that standardised return series computed by (3) has a highly unusual two-peaked distribution.

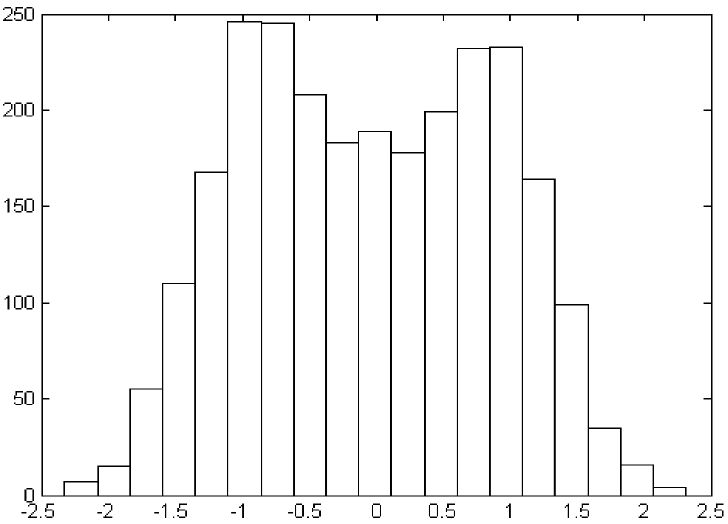

Figure 1. *Standardised by* $\sigma_{t,8}$ *AUD/USD futures returns*

This empirical effect has been entitled the two-component effect. After the careful examination it has been found that the two-component effect arrises due to the special case of heteroscedasticity in intraday returns. Specifically, the simple standard deviation of returns over the first intraday period is more than three times larger than standard deviations of the rest of intraday returns. The table below demonstrates this.

Table 1. Simple standard deviations of intraday returns

n =	1	2	3	4	5	6	7	8
s.d. of returns	0,0051	0,0015	0,0015	0,0015	0,0014	0,0012	0,0011	0,0012

In mathematical terms this situation can be expressed as follows.

$$z_t = \frac{(x_{t,1} + j_t) + x_{t,2} + \ldots + x_{t,k}}{\sqrt{(x_{t,1} + j_t)^2 + x_{t,2}^2 + \ldots + x_{t,k}^2}} \tag{4}$$

where $x_{t,n} \sim N(0, \sigma_t^2 / k)$, $j_t \sim N(0, \sigma_{j,t}^2)$, $\sigma_{j,t}^2 > 0$ and k is the number of intraday returns within day t.

It has been found that with $k = 8$, $\sigma_t^2 = 1$ and $\sigma_{j,t}^2 = 3$ the random variable given by (4) has a distribution that is visually identical to the distribution of empirically observed standardised returns. Recognising that if $k \to \infty$ then $x_{t,n} \to 0$ and relating the homogenous variance σ_t^2 to the variance of a jump as $\sigma_{j,t}^2 = \alpha \cdot \sigma_t^2$, the function (4) can be presented as follows.

$$z_t = \frac{y_t + j_t}{\sqrt{\sigma_t^2 + j_t^2}} \tag{5}$$

where $y_t \sim N(0, \sigma_t^2)$, $j_t \sim N(0, \alpha \cdot \sigma_t^2)$, $\alpha \geq 0$ reflects the portion of σ_t^2 in $\sigma_{j,t}^2$ and $Cov(y_t, j_t) = 0$.

Since the denominator of (5) is an unbiased and consistent estimator of the standard deviation of the numerator and the variance of y_t is directly related to the variance of j_t through the parameter α, it can be demonstrated that the distribution of z_t does not depend on the variance of y_t. Therefore (5) can be generalised as follows.

$$z_t = \frac{y_t + j_t}{\sqrt{1 + j_t^2}} \tag{6}$$

where $y_t \sim N(0,1)$, $j_t \sim N(\gamma, \alpha)$, $\alpha \geq 0$ and $Cov(y_t, j_t) = 0$.

The new function is a combination of two i.n.d. random variables one of which is standard normally distributed and the other is normally distributed with arbitrary parameters. Since the distribution of z_t given by (6) depends only on two parameters γ and α, this probability density will be referred to as $J(\gamma, \alpha)$. Figure 2 demonstrates some shapes that $J(\gamma, \alpha)$ takes with variation of parameters γ and α.

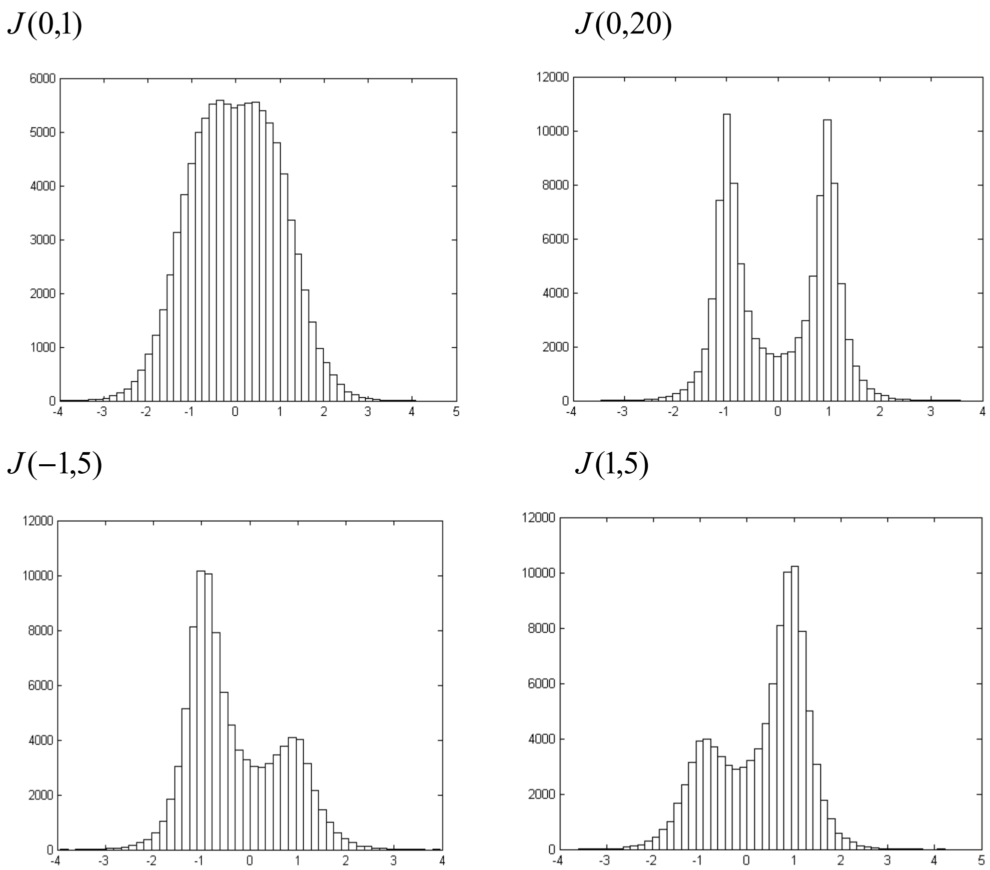

Figure 2. *Histograms of $J(\gamma, \alpha)$ distributed random variables*

It should be noted that if $\gamma = 0$ then z_t given by (6) always have expected unconditional mean and variance parameters 0 and 1 respectively. More detailed characterisation of $J(\gamma, \alpha)$ together with visual examples and technical comments is given in Moldovan [8]. It also should be noted that since at the present moment $J(\gamma, \alpha)$ has not a probability density function, this distribution is simulated but not theoretical. However, since the random number generator is already (see [9]), J can already be applied practically. In the next section it is demonstrated how $J(\gamma, \alpha)$ can be used in the financial model for forecasting purposes.

4. Application of $J(\gamma,\alpha)$: Forecasting Probability Quintiles of Future Price Distribution

In this section is shown how $J(\gamma,\alpha)$ can be applied for improving the forecasting performance of the financial model. Specifically, it is demonstrated how taking into account the previously unnoticed distributional specifics of conditional on realised volatility returns can substantially improve forecasts of Value at Risk (VaR), which after the acceptance by the Basle Committee (Basel Committee on Banking supervision [10]) has become the most widely used risk measurement tool in the banking sector.

VaR of a financial asset or portfolio of assets is defined as the maximum loss that can be expected with a certain level of confidence over a particular interval of time. Thus, VaR is equivalent to a quintile of a future gain/losses distribution on a considered asset or portfolio of assets that corresponds to a desired confidence level. One of the most efficient ways of VaR forecasting is based on the Monte Carlo simulation technique (see Jorion [11]). This technique can be implemented by forecasting the future gain/losses distribution and selecting a required quintile. The following financial model is used for simulation.

$$P_t = P_{t-1} \cdot \exp(r_t) \tag{7a}$$

$$r_t = \sigma_{t,k} z_t \qquad z_t \sim i.i.d.N(0,1) \tag{7b}$$

$$q_t = \ln(\sigma_{t,k}) \tag{7c}$$

$$q_t = c_1 + c_2 q_{t-1} + c_3 \eta_{t-1} + \eta_t \tag{7d}$$

$$\eta_t = \xi_t \sqrt{h_t} \qquad \xi_t \sim i.i.d.N(0,1) \tag{7e}$$

$$h_t = w + a\eta_{t-1}^2 + bh_{t-1} \tag{7f}$$

where P_t and r_t are the price and log return at time t respectively, $\sigma_{t,k}$ is the realised standard deviation estimated from returns sampled k times per time interval, c_1, c_2 and c_3 are the unconditional mean, AR and MA are coefficients in (7d) respectively, η_t is the serially uncorrelated innovation, h_t is the conditional second moment of η_t. w, a and b are the unconditional mean, ARCH and GARCH coefficients in (7f) respectively.

For evaluation of the forecasting performance of the 1000 days backtesting procedure the following confidence intervals have been used (Kupiec [12]).

$$\text{95\% confidence level:} \quad 37 < X < 65 \tag{8a}$$

$$\text{99\% confidence level:} \quad 4 < X < 17 \tag{8b}$$

where X is a umber of prices given by (7a) that are less than the actually observed prices.

In words, over 1000 days we expect 50 and 10 prices, which called exceptions, to lie below the actually observed prices for 95% and 99% confidence levels respectively. However, since VaR is a stochastic measure some deviations from the expected values should be anticipated. Values that are within the intervals given by (8a) and (8b) can be attributed to a chance. Values outside these intervals will indicate the model misspecification. The results of the backtesting procedure applied for 1 to 10 days ahead VaR forecasts are shown below with * indicating the numbers of exceptions that lie outside the non-rejection regions given by (8a) and (8b).

Table 2. Numbers of exceptions of (7) with $z_t \sim N(0,1)$ in (7b)

Days ahead	1	2	3	4	5	6	7	8	9	10
VaR^{95}	54	47	47	44	42	41	33*	35*	36*	34*
VaR^{99}	4*	7	9	8	5	7	9	7	6	7

As can be seen, the model (7) is misspecified. The misspecification can be attributed to standard normally distributed z_t in (7b) that in fact found to be J distributed. After replacing $z_t \sim N(0,1)$ for $\sim J(0,14.8)$ in (7b), the following backtesting results have been obtained.

Table 3. Numbers of exceptions of (7) with $\sim J(0,14.8)$ in (7b)

Days ahead	1	2	3	4	5	6	7	8	9	10
VaR^{95}	57	51	47	47	42	43	38	38	38	36*
VaR^{99}	9	10	11	12	8	7	10	9	8	8

The model has become almost correctly specified.

5. Conclusion

In the study the distribution of conditional on realised volatility financial returns has been studied. It has been demonstrated that conditional returns are not standard normally distributed, as argued in Andersen et al. [6], but follow a specific previously undocumented distribution. The probability density that represents this distribution has been derived. Since the shape of the new density depends on two parameters it was referred to as $J(\gamma,\alpha)$. It has been demonstrated how $J(\gamma,\alpha)$ can be used in the financial model for Value at Risk forecasting. It should be noted that J does not have a probability density function and therefore is simulated but not a theoretical distribution. It should also be noted that bimodal $J(\gamma,\alpha)$ is the simplest in the J^m family that includes probability densities with any desired number of peaks *m+1*. We are confident that time will indicate additional applications as well as limitations of J^m.

References

1. E. Fama, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 1970, pp. 383-417.
2. J.Campbell, A. Lo and C. MacKinlay, The econometrics of financial markets, (Princeton University Press, New Jersey, 1997).
3. T. Andersen and T. Bollerslev, Answering the critics: yes, ARCH models do provide good volatility forecasts, National Bureau of Economic Research, 1997, Working Paper No. 6023.
4. T. Andersen, T. Bollerslev, F. Diebold and P. Labys, The distribution of realised exchange rate volatility, *Journal of the American Statistical Association* 96, 2001, pp. 42-55.
5. T. Andersen, T. Bollerslev, F. Diebold and P. Labys, Realised volatility and correlation, Working Paper, University of Pennsylvania, 1999.
6. T. Andersen, T. Bollerslev, F. Diebold and P. Labys, Exchange rates returns standardised by realised volatility are (nearly) Gaussian, *Multivariate Finance Journal* 4, 2000, pp. 159-179.
7. P. Clark, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 1973, pp. 135-156.
8. M. Moldovan, Stochastic modelling of random variables with an application in financial risk management, Masters Thesis, School of Economics and Finance, Queensland University of Technology, 2003.
9. http://www.esh.lv/r&d/Matlab_codes/
10. Basel Committee on Banking Supervision, Overview of the amendment to the capital accord to incorporate market risks, January, 1996.
11. P. Jorion, Value at Risk, (McGraw Hill, 2001).
12. P. Kupiec, Techniques for verifying the accuracy of risk management models, *Derivatives* 2, 1995, pp. 73-84.

Appendices

Appendix 1. Data description

One main dataset is used in this study. This is the foreign futures exchange rates between the Australian dollar (AUD) and American dollar (USD). AUD/USD futures exchange rates series is obtained from Tick Data, Inc (www.tickdata.com) and covers the period from 2 January 1990 to 31 March 2000 ($T = 2586$ trading days). The duration of each trading day t is 400 minutes. Trading is open Monday to Friday, from 7.20 a.m. to 2.00 p.m. Data is recorded tick by tick and contains 234,905 observations of prices on AUD/USD contracts.

Appendix 2. Volatility signature plot

For selecting the optimal sampling frequency for realised volatility estimation the volatility signature plot approach has been used in the study. A volatility signature plot has been suggested in Andersen et al. [5] and is a graphical diagnostic, designed to provide some guidance in selecting the optimal return sampling frequency for realised volatility series computation. The problem is to find the balance between a measurement error, which is the most severe in the case of using squared returns as volatility estimates, and microstructure noise, which distorts realised volatility estimates as sampling frequency increases. AUD/USD futures exchange rates series, containing tick-by-tick data, is resampled with different frequencies. Eight sampling frequency intervals have been used: 5, 10, 20, 40, 50, 100, 200 and 400 minutes. The corresponding numbers of intraday prices are $k =$ 80, 40, 20, 10, 8, 4, 2 and 1. For return calculation, the last prices that were observed over a sampling period have been used. If no transactions occurred over a sampling period, then the latest recorded prices were taken.

Realised volatility in day t based on k intraday price observations is calculated as follows.

$$v_{t,k} = \sum_{n=1}^{k} x_{t,n}^2 \tag{A2.1}$$

As a result, eight realised volatility series $v_{1,80},\ldots,v_{T,80};\ldots;v_{1,1},\ldots,v_{T,1}$ were obtained. To define a volatility signature plot, the average values of each of eight series were calculated.

$$\overline{v}_k = T^{-1}\sum_{t=1}^{T} v_{t,k} \tag{A2.2}$$

The average realised volatilities $\overline{v}_k$, $k = 80,40,\ldots,1$ were plotted against the lengths of sampling intervals in minutes $5,10,\ldots,400$. Figure A2.1 shows the result.

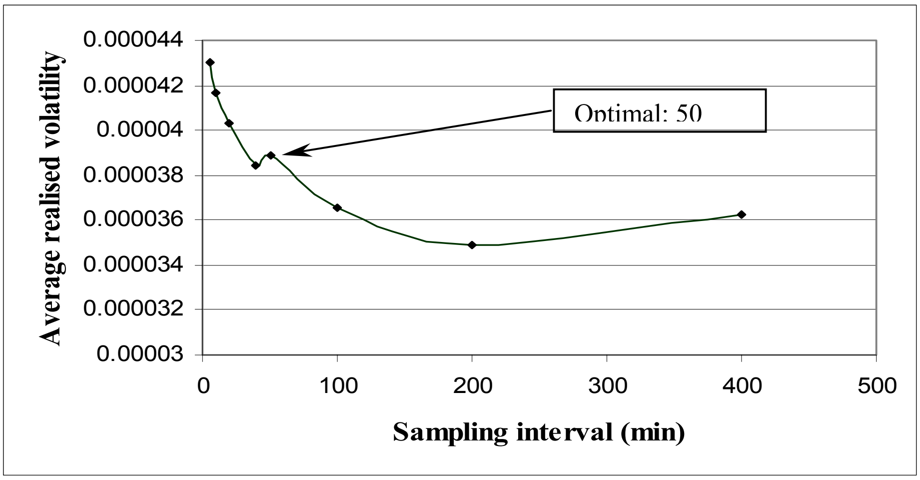

***Figure A2.1**. Realised volatility signature plot for AUD/USD series*

Remembering that microstructure effects are of the main concern in series with the highest sampling frequency, one should follow the volatility signature plot from the left to the right. Andersen et al. [5] argued that the point where the volatility signature plot is "stabilised" should be selected as optimal. In the AUD/USD series plot, this point corresponds to the 50 minute sampling interval ($\overline{v}_8$). This means that for obtaining the optimal realised volatility estimates, returns should be sampled 8 times per trading day. This sampling frequency was used in the study for estimating the realised volatility series.

THE SIMULATION-FREE SIMPLE TEST FOR NORMALITY: SOME COMPARISONS OF SIZE AND POWER

Max V. Moldovan[1], Konstantin N. Nechval[2] and Nicholas A. Nechval[2]

[1] ESH Group Ltd.,
6-450 Ezermalas Street, Riga, Latvia, LV-1016
Ph: (+371) 9 82 69 66. Fax: (+371) 7 518 519. E-mail: max@esh.lv

[2] Mathematical Statistics Department, University of Latvia,
19 Raina Blvd, Riga, Latvia, LV-1050
Ph: (+371) 7034701. Fax: (+371) 7034702. E-mail: nechval@junik.lv

Abstract

The question of whether or not a set of random variables follows a specific theoretical distribution often arises in many theoretical and applied disciplines. When true population parameters of a hypothesised distribution are known the test is known as simple. However, more often one is not able to deal with a whole population because of reasons such as the absence of access to an entire dataset or unacceptable cost of information collection. In this case, parameters have to be estimated from a sample and the test that is based on parameter estimates is known as composite. Inferences that are based on a composite test are in general less reliable than in the case with a simple test. In the following paper we suggest the analytical simple test for normality that allows performing the testing without knowing mean and variance parameters of a population. The presented test can be especially effective in the time-critical applications such as automatic control. Size and power of the presented test are given and compared with the traditional composite test for normality.

Keywords: Decision support systems; Simple test for normality; Change-point detection

1. Introduction

The question of whether or not a set of random variables follows a specific theoretical distribution often arises in many theoretical and applied disciplines. When true population parameters of a hypothesised distribution are known the test is referred to as *simple*. However, more often one does not have the ability to deal with an entire population because of reasons such as the absence of access to a complete dataset or unacceptable cost of information collection. In this case, parameters have to be estimated from a sample and the test that is based on parameter estimates is known as *composite*. Inferences that are based on a composite test are in general less reliable than in the case with a simple test. To understand the reason for that, it should be noted that estimates obtained from a sample will often differ from parameters of a population. The sample estimates are the 'best fit' parameters, or parameters that fit especially well to a sample that they are estimated from. We can say that a sample and estimates are 'relatives' and therefore composite tests are too liberal, failing to reject H_0 of a hypothesised distribution reliably enough.

The normal distribution is certainly the most widely used by academics and practitioners theoretical distribution. Some authors suggest simulation-based techniques that allow overcoming the weaknesses of composite tests for normality. For example, Dufour et al. [1] use the Monte Carlo simulation for controlling size of the several tests. Becker and Hurn [2] also used the simulation-based bootstrap technique for correcting the normality test size. Although with the power of modern computers simulation-based techniques are already not excessively computationally intensive, there are still some areas where the time consuming simulation methods are not affordable. In the following paper we suggest the analytical simple test for normality that allows performing the testing without knowing values of population mean and variance parameters. The test is based on the method first suggested by Nechval [3]. This method is originally designed for change-point detection in regression relationship and especially efficient in small samples of data.

The paper is structured as follows. In Section 2 the parameter-free transformation and the simple test for normality are presented in the explicit algorithmic form. In Section 3 power and size of the suggested test are presented and compared with the traditional composite test. Section 4 summarises the study.

2. Parameter-Free Transformation and Simple Test for Normality

The presented test is based on the technique originally suggested in Nechval [3]. This technique allows transforming a set of random variables to a slightly smaller set of random variables (less than a number of eliminated parameters) without knowing any parameters of the initial sample. The idea behind the method can be stated as follows. A stochastic process observed up to time t can be described by an invariant statistic that fully defines the character of the process with respect to known assumptions. This statistic follows a known theoretical distribution. At time t+1 the next outcome of the statistic becomes available. Since distribution of the statistic is known, the confidence interval $1-\alpha$ can be specified. If a new outcome of the statistic falls outside the confidence interval it can be concluded that the process is broken with confidence $1-\alpha$. The analytical derivation of the method is given in Moldovan et al. [4]. Here the simple test for normality, which follows directly from the method, is presented in the explicit algorithmic form.

Suppose a sample of random variables X_t ($t = 1, 2, ..., T$) is drawn from a population with an unknown stochastic character. It is required to test $H_0 : X_t \sim N(\mu, \sigma^2)$ against H_A : X_t is not $N(\mu, \sigma^2)$. The strategy that employed below transforms X_t to a set of random variables u_t ($t = 3,4...,T$) distributed on the interval $(0,1)$. Note that a resulted set u_t ($t = 3,4...,T$) is two variables less than the initial sample X_t because two parameters, mean and variance, are eliminated from consideration. $X_t \sim N(\mu, \sigma^2)$ if and only if u_t ($t = 3,4...,T$) is identically uniformly distributed $U(0,1)$. The test in the algorithmic form is presented below.

1. Remove the unconditional mean from the initial sample X_t:

$$x_t = X_t - \frac{\sum_{t=1}^{T} X_t}{T} \qquad (1)$$

Now the resulted sample x_t ($t = 1,2,...,T$) has an unconditional expectation 0.

2. Take the first sub-sample x_t ($t = 1,...,n$, $n = 2$) and calculate the sum of squared residuals:

$$s\_sq_1 = \sum_{t=1}^{n} \dot{x}_t^2 , \qquad \dot{x}_t = x_t - \frac{\sum_{t=1}^{n} x_t}{n} \qquad (2)$$

3. Take the second sub-sample x_t ($t = 1,...,n$, $n = 3$) and calculate the sum of squared residuals:

$$s\_sq_2 = \sum_{t=1}^{n} \dot{x}_t^2 , \qquad \dot{x}_t = x_t - \frac{\sum_{t=1}^{n} x_t}{n} \qquad (3)$$

4. Calculate the first characteristic variable:

$$z\_sq_n = n-2\left(\frac{s\_sq_2}{s\_sq_1}-1\right),\ n=3 \tag{4}$$

5. Repeat Step 3 with the next sub-sample x_t ($t=1,...,n$, $n=4$) and calculate the next characteristic variable:

$$z\_sq_n = n-2\left(\frac{s\_sq_3}{s\_sq_2}-1\right),\ n=4 \tag{5}$$

6. Repeat Step 5 until $n=T$. As a result the set z_sq_t ($t=3,4...,T$) of characteristic variables is obtained.

7. Transform characteristic variables to variables that lie on the interval $(0,1)$ by applying the cumulative distribution function of the central F distribution with 1 and $t-2$ ($t=3,...,T$) degrees of freedom to the characteristic variables z_sq_t ($t=3,4...,T$):

$$u_t = 1-F_{1,t-2}(z\_sq_t),\ (t=3,4...,T) \tag{6}$$

8. Test if the sample u_t ($t=3,4...,T$) is identically uniformly distributed $U(0,1)$. Any test for uniformity, such as Kolmogorov-Smirnov (K-S) test (Smirnov [5]) can be used for that. The initial sample X_t is distributed $N(\mu,\sigma^2)$ if and only if u_t ($t=3,4...,T$) is identically uniformly distributed $U(0,1)$. The end of the algorithm.

Note that no information about the mean and variance of the initial sample X_t has been used in the test indicating that the presented test is simple.

3. Comparisons of Size and Power

The size and power of the proposed test have been measured with application of the K-S test at Step 8 of the presented algorithm. The size of the test is based on $M=10000$ Monte Carlo trials and reported in Table 1.

Table 1. Size of the simple test

sample size T =	**50**	**100**	**200**	**300**	**400**	**500**
α = 0,01	0,0127	0,0086	0,0084	0,0108	0,0107	0,0100
0,05	0,0560	0,0491	0,0443	0,0497	0,0498	0,0499
0,10	0,1052	0,0957	0,0978	0,1044	0,0979	0,0989

Note: α is the level of significance

As can be seen, the size of the presented test is asymptotically correct. For comparison, the size results of the ordinary K-S test for normality have been computed. Note that this alternative test is composite since mean and variance parameters must be estimated from the samples and used for testing. The size of the alternative test is reported below.

Table 2. Size of the composite test

sample size T =	**50**	**100**	**200**	**300**	**400**	**500**
α = 0,01	0,0075	0,0077	0,0079	0,0085	0,0085	0,0084
0,05	0,0395	0,0356	0,0430	0,0435	0,0422	0,0449
0,10	0,0751	0,0844	0,0860	0,0849	0,0920	0,0916

Note: α is the level of significance

It is obvious that the size of the composite test is substantially distorted due to deviations of mean and variance estimates from the true parameters of a population. In other words, the composite test is less reliable.

Next, the power of the presented simple test is measured against some common theoretical distributions and reported in Table 3.

Table 3. Power of the simple test

	Distribution						
	Uniform	$\chi^2(2)$	$t(2)$	$t(4)$	$t(6)$	$t(8)$	$t(10)$
α = 0,01	0,1135	0,4596	0,8127	0,2778	0,1277	0,0713	0,0511
0,05	0,3027	0,6514	0,8919	0,4477	0,2621	0,1740	0,1474
0,10	0,4354	0,7451	0,9258	0,5411	0,3569	0,2591	0,2233

Note: Power is measured with samples of length $n = 100$

Compare with the power of the composite K-S test reported in Table 4.

Table 4. Power of the composite K-S test

	Distributions						
	Uniform	$\chi^2(2)$	$t(2)$	$t(4)$	$t(6)$	$t(8)$	$t(10)$
α = 0,01	0,0035	0,5519	0,0551	0,0128	0,0103	0,0101	0,0069
0,05	0,0070	0,8995	0,2097	0,0668	0,0526	0,0186	0,0162
0,10	0,0367	0,9696	0,3540	0,1277	0,1056	0,0431	0,0424

Note: Power is measured with samples of length $n = 100$

With exemption of $\chi^2(2)$ distribution, the suggested simple test is more powerful than the alternative composite K-S test. Although there are some composite tests with greater power than the considered K-S test, it should be noted that these tests can also be used at Step 8 of the presented in Section 2 algorithm, thus increasing a relative power of the suggested simple test. Moreover, to the best of our knowledge, no simple or composite tests exist that have considerable power in detection of such hardly distinguishable from a normal distributions as Student $t(v)$ with $v \geq 4$. As can be seen from Table 3 the presented simple test for normality keeps a substantial power with these distributions.

4. Conclusion

In the paper the simple, as opposed to composite, test for normality was presented. The technical derivation of the method that the suggested test is based on is given in Moldovan et al. [4]. In this paper the eight-step explicit algorithm is given as an alternative to the mathematical representation. Size and power of the suggested simple test are reported and compared with the traditional composite K-S test. It is shown that in opposite to the composite test, the simple test has the correct size, or more reliable. Additionally, with exception of $\chi^2(2)$ distribution, the simple test is more powerful than the composite K-S test. Specifically, the presented simple test detects distributions that are only slightly deviate from a normal in the tail regions, such as Student $t(v)$ with $v \geq 4$, much better than any alternatives that we are aware of. This feature makes the suggested simple test especially useful in areas where extreme outcomes of the process are of the main concern, such as financial risk management. Finally, it should be noted that the suggested simple test, as opposed to the known simulation-based simple alternatives, is analytical and therefore can be used even in the time-critical applications, such as automatic control systems.

References

1. J.-M Dufour, A. Farhat, L. Gardiol and L. Khalaf, Simulation-based finite sample normality test in linear regressions, *Econometrics Journal* 1, 1998, pp. C154-C173.
2. R. Becker and S. Hurn, A bootstrap test for uniformity using the Gini coefficient, Working Paper, Queensland University of Technology, 2002.
3. N.A. Nechval A general method for constructing automated procedures for testing quickest detection of a change in quality control, *Computers in Industry* 10, 1988, pp. 177-183.
4. M. Moldovan, N. Nechval, E. Vasermanis, U. Rozevskis and K. Rozite, Testing for two-phase multiple regressions, in Proceedings of the International Conference on Reliability and Statistics in Transportation and Communication, 2002, pp. 73-81.
5. N. Smirnov, Table for estimating goodness of fit of empirical distributions, *Annals of Mathematical Statistics* 19, 1948, pp. 279-281.

TRANSPORT and TELECOMMUNICATION

ISSN 1407-6160

The journal of Transport and Telecommunication Institute (Riga, Latvia)
The journal is published since 2000.

Articles and review are presented in the journal in English (preferably), Russian and Latvian languages (at the option of authors).

EDITORIAL CORRESPONDENCE

Transporta un sakaru institūts (Transport and Telecommunication Institute)
Lomonosova iela 1, LV-1019, Riga, Latvia. Phone: (+371)-7100594. Fax: (+371)-7100535.
E-mail: kiv@tsi.lv, http:// www.tsi.lv

The K. Kordonsky Charitable Foundation

The International Conference
RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION (RelStat'04)

15-16 September 2004. Riga, Latvia

PURPOSE

The purpose of the conference is to bring together academics and professionals from all over the world to discuss the themes of the conference:

- Theory and Applications of Reliability and Statistics
- Reliability and Safety of Transport Systems
- Rare Events and Risk Management
- Software Reliability and Testing
- Modelling and Simulation
- Intelligent Transport Systems
- Transport Logistics and Economics
- Education Programmes and Academic Research in Reliability and Statistics

DEDICATION

The Conference is devoted to the memory of Prof. Kh.Kordonsky.

OFFICIAL LANGUAGES

English and Russian will be the official languages of the Conference.

SUPPORTED BY:

Transport and Telecommunication Institute (Latvia) and
The K. Kordonsky Charitable Foundation (USA) in co-operation with:
Latvian Transport Development and Education Association (Latvia)
Latvian Academy of Science (Latvia)

SPONSORED BY

Transport and Telecommunication Institute (Latvia)
The K. Kordonsky Charitable Foundation (USA)
ALTA - Consulting Company in the Air Navigation Services (Latvia)
Latvian Operations Research Society
PAREX bank (Latvia)

HOSTED BY

Transport and Telecommunication Institute (Latvia)

SECRETARIAT

Prof. Igor Kabashkin, Latvia - Chairman
Ms. Inna Kordonsky-Frankel, USA - Co-Chairman
Mr. Nikolay Gudanets, Latvia – Secretary

DEADLINES AND REQUIREMENTS

Abstracts Submissions:	**June 1, 2004**
Camera-ready final manuscript:	**September 16, 2004**
Conference start:	**September 16, 2004**

Abstracts submitted for review should be a maximum of 600 words in length, should present a clear and concise view of the motivation of the subject, give an outline, and include a list of references.

The abstracts should reach the Secretariat before June 1, 2004. Authors should provide a maximum of five key words describing their work. Please include the full name, affiliation, address, telephone number, fax number, and e-mail address of the corresponding author.

Camera-ready documents must be handed in at the registration desk. Papers presented at the conference will be included in the conference proceedings. Instruction for papers preparing can be found on the conference WWW page: www.tsi.lv.

REGISTRATION FEE

The registration fees will be Euro 30 before the deadline, and Euro 40 after the deadline. This fee will include: hard copy of the Conference Abstracts, hard copy of the Conference Proceedings (the proceedings will be mailed to the delegates after the conference), coffee breaks, Conference Dinner.

Each registration fee might include just one paper, which presentation will be included in the conference program and published in the conference proceedings.

VENUE

Riga is the capital of the Republic of Latvia. Thanks to its geographical location, Riga has wonderful trade, cultural and tourist facilities. Whilst able to offer all the benefits of a modern city, Riga has preserved its historical charm. It's especially famous for its medieval part – Old Riga.

Old Riga still preserves many mute witnesses of bygone times. Its old narrow streets, historical monuments, organ music at one of the oldest organ halls in Europe attract guests of our city. In 1998 Old Riga was included into the UNESCO list of world cultural heritage.

ACCOMMODATION

A wide range of hotels will be at the disposal of participants of the conference and accompanying persons.

FURTHER INFORMATION

Contact:

Nikolay Gudanets, Secretary, RelStat'04
Transport and Telecommunication Institute
Lomonosova iela 1, Riga, LV-1019, Latvia
Telephone: +(371)-7100651
Fax: +(371)-7100660
E-mail: nlg@tsi.lv
WWW: www.tsi.lv