University *of* Leicester

**DEPARTMENT OF ECONOMICS**

# The behaviour of Dickey Fuller test in the case of noisy data: to what extent we can trust the outcome

**Stephen G. Hall, University of Leicester, UK**
**Sahar S. Qaqeesh, Univerersity of Leicester, UK**

**Working Paper No. 09/18**
**September 2009**
**Updated 18/09/2009**
**Updated 21/09/2009**

# The Behaviour of Dickey Fuller Test in the Case of Noisy Data: To What Extent We Can Trust the Outcome

**Stephen G. Hall**
Department of Economics
University of Leicester

**Sahar S. Qaqeesh**
Department of Economics
University of Leicester

**September 18, 2009**

**ABSTRACT**

We examine the behaviour of Dickey Fuller test (DF) in the case of noisy data using Monte Carlo simulation. The findings show clearly that the size distortion of DF test becomes larger as the noise increases in the data.

Keywords: Hypothesis testing; Unit root test, Monte Carlo Analysis.

JEL classification: C01, C12, C15.

## 1. INTRODUCTION

Various unit root tests have been employed in empirical work to identify the order of integration of economic variables. So far Dickey Fuller (DF) test remains the most famous one. A substantial body of research examines the main characteristics of DF test and particularly its main shortcomings such as low power, large size distortion and sensitivity to the true data generating process (DGP). Diebold and Rudebusch (1989a, 1991) and DeJong, Nankervis, Savin, and Whiteman (1992) examine the power of DF test when the process has short memory with a unit root close to unity and provide strong evidence that DF test has low power when the process is fractionally integrated.

The Size distortion of DF test is also examined heavily in the literature. Perron (1989), Hamori and Tokihisa (1997), Montañés and Reyes (1998), Leybourne and Newbold (2000), Sen (2001, 2003, 2008) and Kim, Lybourne and Newbold (2004) examine the behaviour of DF test in the case of structural breaks. Cheung and Lai (1998), and Cook and Manning (2004) examine the influence of the lag selection process using standard information criteria on the size distortion of ADF test. Granger and Hallman (1991), Kramer and Davies (2002) examine the robustness of DF test in the case of improper transformations of the data. Phillips and Perron (1988), Schwert (1989), Agiakloglou and Newbold (1992) analyze the performance of DF test when the process that generates the time series contains moving average term. The main findings for all these studies show that the distribution of the unit root test statistics is different from the distribution proposed by Dickey-Fuller. Accordingly a severe size distortion occurs and the power of DF test becomes questionable.

In this paper, we examine the performance of DF test when the process that generates the time series contains noise. The main focus will be on the size distortion of DF test as the noise increases in the data. This paper is organized as follows: section 2 the motivation, section 3 the full design of Monte Carlo experiment that is employed to illustrate the behaviour of DF test and the size distortion that occurs as a result of the measurement errors in the data, Section 4 the empirical results and section 5 the conclusion.

## 2. THE MOTIVATION

The motivation for this study comes from some simple tests on the order of integration on Jordanian inflation. Figure 1 shows monthly inflation in Jordan from 1997 to 2007, if we consider the spectrum for this series it suggests a non-stationary process, but as table 1 shows the DF test strongly suggests that the series is stationary. In order to check this we then defined the inflation rate over a 12 month period, shown in figure 2 and this clearly looks non-stationary and indeed, as table 1 shows, the DF test strongly suggests non stationarity

Table 1: Dickey-Fuller tests for stationarity of Jordanian Inflation

| **DF Test using (AIC), Exogenous: Constant** | |
|---|---|
| | **t-statistics, no. of lags** |
| **INFLATION RATE (H=1 MONTH)** | ****-9.368042(1)** |
| **INFLATION RATE (H=12 MONTH)** | ***-0.899845(12)** |

*** The null hypothesis of non-stationarity can't be rejected at (1%,5%,10%) significance levels.**

**** The null hypothesis of non-stationarity is rejected at (1%, 5%,10%) significance levels.**

Of course these two tests are in clear contradiction of each other as,

$$\pi_t^{12} = P_t - P_{t-12} = (P_t - P_{t-1}) + (P_{t-1} - P_{t-2}) + \ldots + (P_{t-11} - P_{t-12}) = \pi_t^{1} + \pi^{1}{}_{t-1} + \ldots + \pi^{1}{}_{t-11}$$

Where $\pi^{j}{}_{t}$ is the rate of inflation at time t over j periods. So if monthly inflation is stationary then annual inflation which is simply the sum of monthly inflation must also be stationary. We believe that the explanation for this contradiction lies in the presence of measurement error. This is evident in figure 1 as we can clearly see here that many months exhibited negative inflation, this is not a phenomenon which is observed by Jordanians and hence it would seem to be a problem with the actual measurement of the price level.

If this is the case then we can also show why taking a longer period for the inflation calculation would give a more meaningful test statistic. We begin by assuming that the true price level in logs is a random walk

$$P^*_t = P^*_{t-1} + \varepsilon_t$$

But the observed log of the price level is subject to measurement error

$$P_t = P^*_t + v_t$$

Where both $\varepsilon_t \; and \, v_t$ are IID noise processes and observed inflation is given by

$$\pi^j_t = P_t - P_{t-j}$$

Now

$\pi^1_t = \varepsilon_t + v_t - v_{t-1}$ and hence $\mathrm{var}(\pi^1_t) = \sigma^2_\varepsilon + 2\sigma^2_v$

And

$\pi^{12}_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_{t-11} + v_t - v_{t-12}$ and hence $\mathrm{var}(\pi^{12}_t) = 12\sigma^2_\varepsilon + 2\sigma^2_v$

It is therefore evident that by measuring inflation over a longer period the size of the random walk variance grows relative to the size of the measurement error and hence measurement error has less effect on measured inflation.

To examine the influence of the noisy data on the size distortion and the empirical power of DF test more completely, we now turn to a formal Monte Carlo Simulation.

### 3. Monte Carlo Analysis/ Experiment Design

The Monte Carlo experiment begins with Data Generating Process. The steps of data generating are as follows:-

Step one: Generate a data set using the simplest model of time series which is a non-stationary normal random walk. So the first data set ($x_t$) is a random walk process without a drift and it is generated by an AR (1) model of the form:-

$x_t = \alpha x_{t-1} + \varepsilon_t$, t= 1, 2, ……., T **Eq. (1)**

Where,

$\alpha = 1$

$x_0$ = 0 (the initial value)

$\varepsilon_t$ = Random disturbance is generated from normal distribution with zero mean and constant variance ($\sigma^2$) equals one, i.e. $\varepsilon_t$~ N(0,1).

Step two: Create noises in the data set ($x_t$) by adding random disturbances with zero mean and fixed variance and create a measured variable y.

$y_t = x_t + \omega_t$, t= 1, 2, ……., T **Eq. (2)**

Where we may vary the variance of this error to investigate the effects of different levels of measurement error relative to the random walk component,

$\omega_t$ ~ N(0,0), N(0,0.5), N(0,1), N(0,1.5), N(0,2), N(0,3), N(0,4), N(0,5), N(0,6), and N(0,7).

We consider the following samples sizes T= 25, 50, 100, 150, and 200 observations and we perform 50,000 replications for each sample size and for each variance. We chose 50,000 replications on the grounds that for each sample size using variance zero ($\omega_t$ ~ N(0,0)) we needed 50,000 replications to exactly replicate the standard Dickey Fuller critical values.

In order to show the size distortion of DF test as noise increases in the data, we calculate the percentage of rejection of the null hypothesis at 5% level of significance using the normal critical values of DF test with constant model[1].

## 4. Empirical Results

In table 2 we show clearly that the percentage of rejection of the null hypothesis at 5% level of significance increases dramatically as the noise increases in the data. The null hypothesis of non stationrity is rejected more often in favour of the alternative (the stationarity).

[1] **DF test asymptotic critical values at 5% level of significance under sample sizes (25, 50, 100, 150, and 200) are as follows: -2.99%, -2.92%, -2.89%, -2.88% and -2.88%.**

The benchmark case in this experiment is the one where the variance equals zero (no noise embedded) and both generated data sets ($x_t$) and ($y_t$) are equal. The percentage of rejection of the null hypothesis for all samples' sizes is exactly 5% (see table 2) which means that DF test is able to identify the truth about the unit root using the normal critical values (95%) of the time. When the noise increases, from variance 0.5 till variance 7, the size distortion becomes larger and the percentage of rejection increases dramatically which means that DF test provide misleading results using the same normal critical values.

The results also show that the size distortion becomes larger when the sample size increases even at lower variances which implies that even very large samples containing measurement error will give incorrect inference. Figure 3 demonstrates that under sample size 50, 100, 150 and 200 the percentage of the rejection of the null hypothesis increases faster than the case of sample size 25. The influence of the noise appears more quickly when the sample size is big, for example the percentage of rejection reaches 100% at variance six for both sample size 150 and 200 and at variance seven for sample size 100 while in the case of sample 25 we need to add more noises to reach 100%.

It is crystal clear that the distribution of the t-statistic when the data set contains noises is different from the distribution proposed by Dickey Fuller where the process is a pure random walk. In this paper we propose a new set of critical values that can be used as an indication to identify the unit root in noisy data. The proposed critical values in table 3 are derived from the distribution of t-statistic values across the replications. The critical values at 1%, 5% and 10% are calculated as the first and fifth and tenth percentile of the t-statistic distribution. As a benchmark, the critical values at variance zero equal exactly the asymptomatic critical values under the DF test. It is obvious that the critical values become bigger (in absolute values) when the noise increases in the data and this mean that the new t-statistic distribution will have heavier and fatter tails than normal fat tails.

## 5. Conclusion

The main objective of this experiment is to prove that the rejection of the null hypothesis of unit root under DF test in some cases should not be taken without further investigation.

We prove by Monte Carlo simulation that the size distortion of DF test becomes larger as the noise increases in the data and faster as the sample size becomes bigger.

We believe that DF normal critical values can be misleading and implausible when the data set contains noise. Instead the proposed critical values (table 3) can be more reliable in identifying the truth about unit root properties.

## References


Agiakloglou, C. and Newbold, P. (1992). "Empirical Evidence on Dickey-Fuller-Type Tests", Journal of time Series Analysis, Vol. 13, No. 6, pp. 471-483.

Cheung Yin-Wong, Lai Kon S., (1998). " Power of the augmented dickey-fuller test with information-based lag selection", Journal of Statistical Computation and Simulation, Vol. 60, pp. 57-65.

Cook S., Manning N., (2004), "Lag optimisation and finite-sample size distortion of unit root tests", Economics Letters, Vol. 84, pp. 267-274.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992), "The Power Problems of Unit Root in Time Series with Autoregressive Errors", Journal of Econometrics, Vol. 53, pp. 323-343.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992), "Integration Versus Trend Stationary in Time Series", Econometrica, Vol. 60, No. 2, pp. 423-433..

Dickey,D.A., and Fuller, W.A., (1979). "Distribution of the estimators for autoregressive time series with a unit root". Journal of the American Statistical Association, Vol. 74, No. 366, pp. 427-431.

Dickey,D.A., and Fuller, W.A., (1981). " Likelihood ratio statistics for autoregressive time series with a unit root'. Econometrica, Vol. 49, No. 4, pp. 1057-1072.

Diebold, F. X., and Rudebusch, G. D., (1989), "Long memory and persistence in aggregate output", Journal of Monetary Economics, Vol. 24, pp. 189-209.

Diebold, F. X., and Rudebusch, G. D., (1991), "On the Power of Dickey-Fuller tests against fractional alternatives", Economics Letters, Vol. 35, pp. 155-160.

Granger, C.W.J., and Hallman, J., (1991), "Nonlinear transformations of integrated time series", Journal of Time Series Analysis, Vol. 12, No. 3, pp. 207-224.

Hamori, S., and Tokihisa, A, (1997), "Testing for a unit root in the presence of a variance shift" Economics Letters, Vol. 57, pp. 245-253.

Kim, T., Leybourne, S. J., and Newbold, P., (2004), "Behaviour of Dickey-Fuller Unit-Root Tests Under Trend Misspecification, Journal of Time Series Analysis, Vol. 25, pp.755-764.

Kramer, W., and Davies, L.,(2002), "Testing for unit roots in the context of misspecified logarithmic random walks", Economics letters, Vol. 74, pp. 313-319.

Leybourne, S. J., and Newbold, P., (2000), "Behavior of Dickey-Fuller t-tests when there is a break under the alternative hypothesis", Econometric Theory. Vol. 16, pp. 779-789.

Montañés, A., and Reyes M., (1998), "Effect of a Shift in the Trend Function on Dickey-Fuller Unit Root Tests". Econometric Theory, Vol. 14, pp. 355-363.

Perron, P., (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", Econometrica, Vol. 57, No. 6, pp. 1361-1401.

Phillips, P. C. B., and Perron P., (1988), "Testing for a Unit Root in Time Series Regression", Biomètrika, Vol. 75, No. 2, pp. 335-346.

Rudebusch, G.D.,(1992), "Trends and random walks in macroeconomic time series: A re-examination", International Economic Review, Vol. 33, No. 3, pp. 661-680.

Sen, A. (2001), " Behaviour of Dickey-Fuller F-tests under the trend-break stationary alternative", Statistics and Probability Letters, Vol. 55, pp. 257-268.

Sen, A. (2003), " On unit-root tests when the alternative is a trend-break stationary process", Journal of Business and Economic Statistics, Vol. 21, No.1, pp. 174-184.

Sen, A. (2008), " Behaviour of Dickey-Fuller tests when there is a break under the unit root null hypothesis", Statistics and probability Letters, Vol. 78, pp. 622-628.

Schwert, G. W., (1989)."Tests for unit roots :a Monte Carlo Investigation", Journal of Business and Economics Statistics, Vol. 7, No. 2, pp. 147-159.

**Figure (1)**

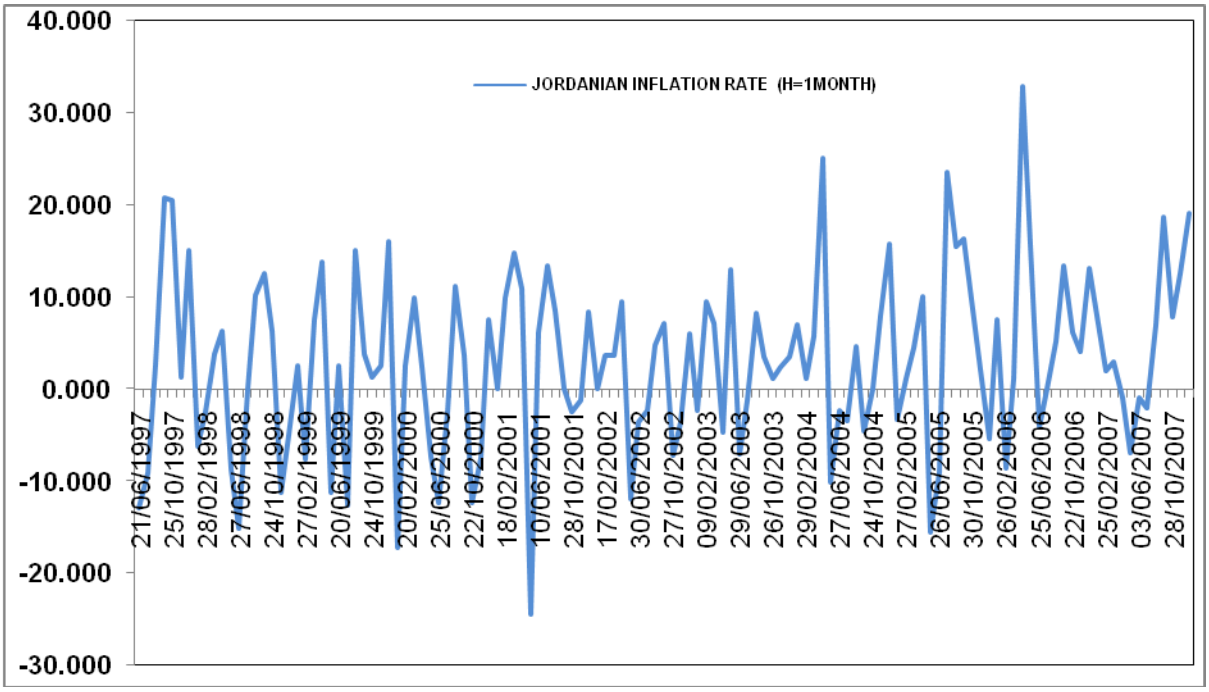
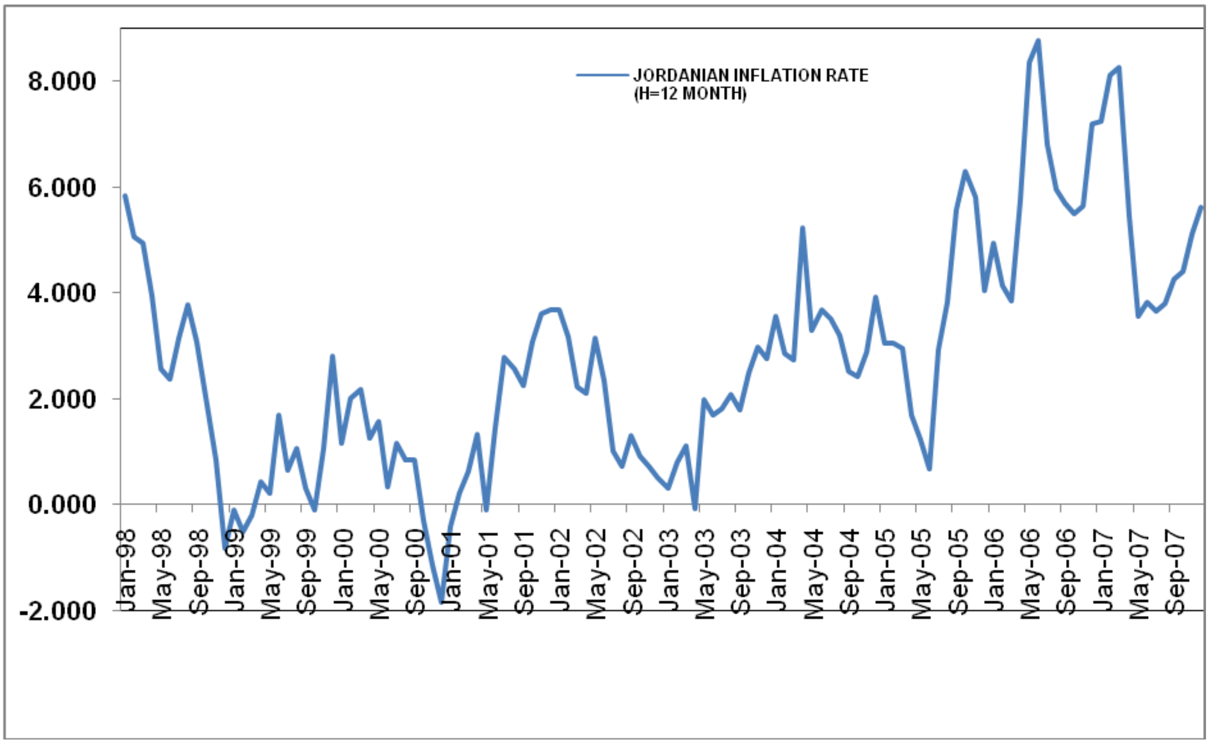
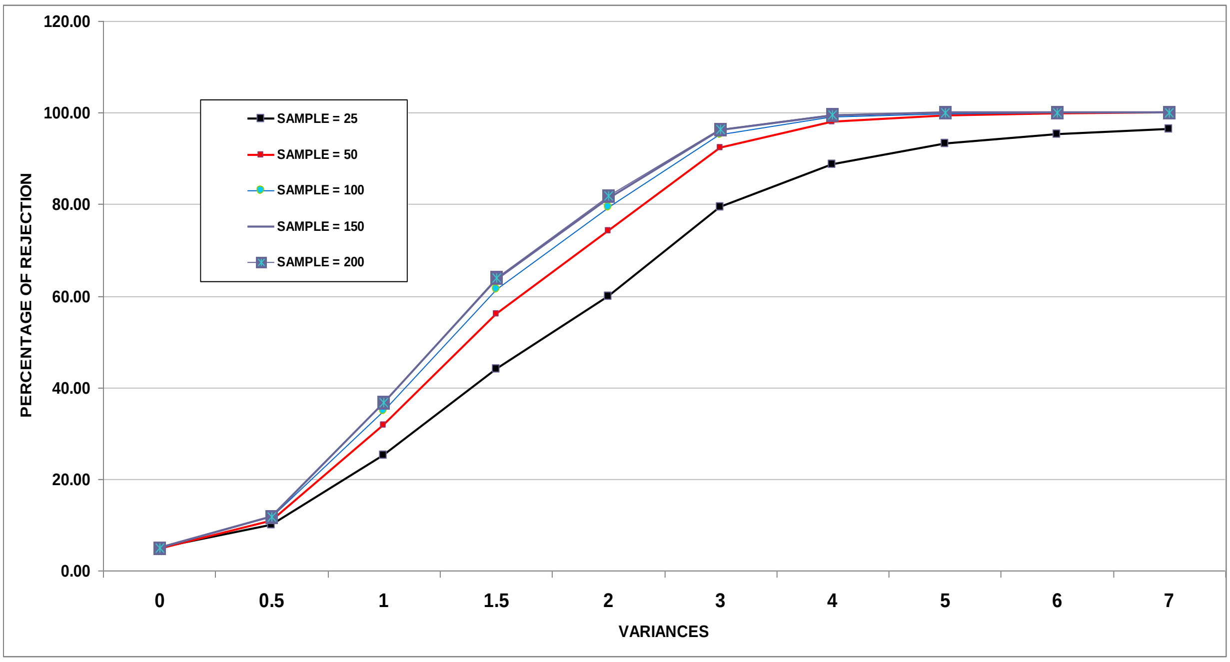



**Figure (2)**



## Figure (3)




## Table 2

| REPLICATIONS =50,000 | THE PERCENTAGE OF REJECTION OF THE NULL HYPOTHESIS OF UNIT ROOT AT 5% LEVEL OF SIGNIFICANCE SAMPLE SIZES = 25, 50, 100, 150, AND 200 | | | | |
|---|---|---|---|---|---|
| | 25<br>t-statistic<-2.99 | 50<br>t-statistic<-2.92 | 100<br>t-statistic<-2.89 | 150<br>t-statistic<-2.88 | 200<br>t-statistic<-2.88 |
| VARIANCE = 0 | 5.04% | 4.79% | 5.01% | 5.02% | 4.96% |
| VARIANCE = 0.5 | 9.90% | 10.89% | 11.73% | 11.82% | 11.85% |
| VARIANCE = 1 | 25.10% | 31.77% | 34.92% | 36.59% | 36.67% |
| VARIANCE = 1.5 | 44.00% | 55.92% | 61.39% | 63.48% | 63.86% |
| VARIANCE = 2 | 60.00% | 74.08% | 79.38% | 81.23% | 81.96% |
| VARIANCE = 3 | 79.47% | 92.35% | 95.27% | 96.10% | 96.43% |
| VARIANCE = 4 | 88.78% | 97.90% | 99.11% | 99.46% | 99.54% |
| VARIANCE = 5 | 93.18% | 99.44% | 99.87% | 99.94% | 99.96% |
| VARIANCE = 6 | 95.23% | 99.86% | 99.98% | 100.00% | 100.00% |
| VARIANCE = 7 | 96.45% | 99.96% | 100% | 100.00% | 100.00% |

**Table 3**

| REPLICATIONS =50,000 | THE CRITICAL VALUES FOR ALL SAMPLES' SIZES | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 25 | | | 50 | | | 100 | | | 150 | | | 200 | | |
| CRITICAL VALUES | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| VARIANCE = 0 | -3.77 | -2.99 | -2.63 | -3.56 | -2.90 | -2.59 | -3.49 | -2.89 | -2.58 | -3.48 | -2.88 | -2.57 | -3.46 | -2.88 | -2.57 |
| VARIANCE = 0.5 | -4.21 | -3.36 | -2.98 | -4.02 | -3.33 | -2.97 | -4.00 | -3.34 | -2.98 | -4.02 | -3.32 | -2.97 | -4.00 | -3.32 | -2.97 |
| VARIANCE = 1 | -5.08 | -4.15 | -3.69 | -5.12 | -4.27 | -3.83 | -5.29 | -4.41 | -3.94 | -5.36 | -4.44 | -3.97 | -5.40 | -4.45 | -4.00 |
| VARIANCE = 1.5 | -5.77 | -4.80 | -4.33 | -6.12 | -5.18 | -4.70 | -6.58 | -5.57 | -5.02 | -6.78 | -5.71 | -5.13 | -6.94 | -5.79 | -5.21 |
| VARIANCE = 2 | -6.29 | -5.26 | -4.80 | -6.87 | -5.91 | -5.41 | -7.63 | -6.57 | -5.99 | -8.05 | -6.87 | -6.24 | -8.35 | -7.06 | -6.40 |
| VARIANCE = 3 | -6.87 | -5.84 | -5.36 | -7.80 | -6.84 | -6.36 | -9.09 | -8.04 | -7.45 | -9.87 | -8.68 | -8.03 | -10.47 | -9.13 | -8.42 |
| VARIANCE = 4 | -7.18 | -6.16 | -5.67 | -8.34 | -7.40 | -6.91 | -9.98 | -8.97 | -8.41 | -11.03 | -9.92 | -9.29 | -11.88 | -10.62 | -9.92 |
| VARIANCE = 5 | -7.36 | -6.36 | -5.86 | -8.70 | -7.75 | -7.28 | -10.52 | -9.56 | -9.05 | -11.80 | -10.76 | -10.18 | -12.83 | -11.66 | -11.01 |
| VARIANCE = 6 | -7.50 | -6.49 | -5.99 | -8.93 | -7.98 | -7.52 | -10.91 | -9.99 | -9.49 | -12.32 | -11.35 | -10.81 | -13.51 | -12.41 | -11.82 |
| VARIANCE = 7 | -7.59 | -6.57 | -6.07 | -9.11 | -8.15 | -7.70 | -11.19 | -10.30 | -9.81 | -12.72 | -11.77 | -11.26 | -13.98 | -12.95 | -12.41 |