Kingston University London

Economics Discussion Papers 2015-6

STOCK MARKET EFFICIENCY IN IRAN: UNIT ROOT TESTING WITH SMOOTH STRUCTURAL BREAKS AND NON-TRADING DAYS

Vince Daly

Kingston University
London, UK

Seyyed Ali Paytakhti Oskooe

Islamic Azad University
Tabriz, Iran

8 September 2015

Abstract

A 'flexible Fourier trend' unit root test, permitting smooth structural breaks of unknown form and dates, is used to test weak-form market efficiency in the Tehran stock market's TEPIX index. Monte Carlo experiments show that this test has low power when non-trading-day gaps in the daily data are filled with missing value codes. The test's properties for weekly returns and for data as published, with non-trading-day gaps suppressed, are better and similar to each other. Analysis of the full sample of TEPIX data as published supports a unit root null but indicates the presence of additional autocorrelation – questioning weak-form efficiency. Sub-sample analysis again finds evidence of a unit root, but also of complex autocorrelation. Support for the unit root increased in the years (2000-2004) following regulatory reform and has decreased since 2008. A Diebold and Mariano (1995) test is used to assess whether the revealed autocorrelation provides an effective basis for predicting price deviations from trend on the basis of their own history. Predictive effectiveness is found at a horizon of one trading day. We conclude that this market has not shown weak-form efficiency.

Keywords: Market efficiency; Unit root tests; Structural breaks; Non-trading days

JEL codes: G12; O16; C22

Acknowledgements: We are grateful for advice from Chris Stewart and for constructive criticism received when an earlier version of this paper was presented to the Economics research seminar at Kingston University.

Address for correspondence: Vince Daly, Department of Economics, Kingston University, Penrhyn Road, Kingston upon Thames, London, KT1 2EE, UK e-mail: vdaly@kingston.ac.uk
Seyyed Ali Paytakhti Oskooe, Department of Economics, Tabriz Branch, Islamic Azad University, Tabriz, Iran e-mail paytakhti@iaut.ac.uk

1. Introduction

Whether or not a developing economy's stock market is efficient is an important question. An allocatively efficient stock market, i.e. one in which the traded securities are appropriately valued, contributes significantly to sustainable growth and development by providing suitable conditions for savings and investment decisions. Where, as in the case of Iran, a stock market is not yet long-established then the question of whether it is developing as an effective institution is arguably particularly urgent. At the same time, the operating environment of an emerging stock market may bring a high risk of institutional instability, for example relatively frequent review and redesign of market microstructure and regulations. To the extent that emerging stock markets are often to be found in emerging market economies then such markets may additionally suffer instabilities that are a consequence of regime changes, e.g. liberalisation programmes, in the wider economy. Such considerations may intensify interest in the question of whether an emerging stock market is efficient but at the same time they point to the possibility of structural breaks, calling into question the applicability of statistical methods that have been employed to address the question of (weak-form) efficiency in well-established stock markets.

This paper examines the (weak form) efficiency of Iran's emerging stock market using a flexible trend unit root test proposed by Enders and Lee (2011) – hereafter E&L. This approximates structural breaks of unknown number and unknown date(s) by incorporating trigonometric components in the underlying data generating process. We use Monte Carlo simulations to assess the sensitivity of this test to the presence of non-trading days and apply the test to rolling sub-samples as well as to the full set of data, providing an extended table of critical values for this purpose. Where the test does not reject a unit root, we use the Diebold and Mariano (1995) comparison of predictive accuracy as a basis for confirming (or not) weak-form efficiency. The remainder of the paper is organized as follows: the empirical methodology is described in section 2; issues arising from the presence of non-trading days in the data are discussed in section 3; section 4 presents the empirical investigations and section 5 concludes.

2. Assessing market efficiency in the presence of structural change

Arguments based on Fama's (1970) Efficient Markets Hypothesis label a stock market as allocatively efficient – relative to a particular set of relevant information, if prices fully and accurately reflect that information. In such a situation there is no distortion in the pricing of capital and risk, in the sense that traders who only know the current price act as if they have full knowledge of the given information set. Given the unbounded nature of potentially relevant information, perfect efficiency may be typically an unrealisable ideal; more realistically a stock market is called 'weak form efficient' if it is allocatively efficient relative to the past history of stock prices. In this case, past prices contain no information that is not available from the current price and thus have no role to play in predicting future prices. Price behaviour can then be characterised for purposes of statistical analysis as a 'random walk', i.e. a series in which the non-deterministic component of future increments is unpredictable, viz: $\Delta P_t = P_t - P_{t-1} = \mu_t + \varepsilon_t$. Here, P_t is the natural logarithm of the stock price index at time t; μ_t is a non-random expected price change ('drift') at time t. The successive ε_t are independently and identically distributed random variables with zero mean and are unpredictable, in the sense that the unconditional density for each ε_t is no different to its density conditioned on the history of prices up until time t. A commonplace variant is the case where the deterministic drift is a constant: $\mu_t = \mu, \forall t$, so that

$$P_t = P_0 + \mu t + \sum_{s=1}^{s=t} \varepsilon_s \tag{1}$$

The random walk characterisation of weak form efficiency in equation (1) implies that P_t should be 'difference-stationary', i.e. integrated of order one: $P_t \sim I(1)$, a 'unit-root' process. Unit root tests have become well-established in financial econometrics literature as an approach to assessing market (weak) efficiency by testing whether a price series appears to be difference-stationary.

A variety of unit root tests have been developed; arguably the most commonly employed is the augmented Dickey–Fuller (ADF) test (Dickey and Fuller 1979). Perron (1989), however, argues that structural changes in time series, for example in the value of a drift parameter, can reduce the effectiveness of this and other tests for unit roots. This argument is relevant if unit root tests are used to assess market efficiency when there is a possibility of institutional changes within a market, or of structural changes in the external environment, and thereby a possibility of structural breaks in time series data. A market that is at a relatively early stage of development or is situated in an emerging market economy may be especially susceptible to such breaks, a perspective adopted here with regards to the Iran stock exchange.

Several methods have been developed to deal with the adverse consequences of structural breaks for unit root tests (see, for example: Perron 1989, Zivot and Andrews 1992, Lumsdaine and Papell 1997, Lee and Strazicich 2003). In most of the methods employed to date for unit root testing in the presence of structural breaks, assumptions must be made regarding the number, date or form of these breaks. The breaks are typically represented by pulse or shift dummies and are thus modelled as occurring instantaneously and inducing abrupt jumps in the values of level and / or trend parameters. Leybourne et al. (1998) argue that, since it is more likely that structural changes are not instantaneous, breaks should be approximated as smooth processes. To allow for this feature in structural breaks and to control for the effect of unknown forms of nonlinear deterministic terms, E&L propose a unit root test that employs a Fourier function approach to flexibly model the trend component, using trigonometric terms to capture nonlinearities in the trend, including structural change at unknown dates. Structural change is thus characterised as an ongoing transition process, rather than as multiple regimes separated by instantaneous structural breaks. E&L demonstrate that an effective Fourier approximation to a nonlinear trend can be achieved with a small number of contributing frequencies, possibly only a single frequency.

Figure 1. Tehran market index (natural logarithms)

Figure 1 shows, in logarithms, the Tehran Stock Exchange (TSE) all-share price index (TEPIX) over the period 2/January/2000 – 31/December/2012. The possibility of structural breaks in the trend is evident – whether this trend is deterministic or resulting from a random walk with drift. The following brief history points to a number of institutional and environmental developments that may have induced such breaks.

The TSE opened in 1967; trading was initially limited to corporate and public sector bonds but economic growth during the 1970s fuelled demand for a market in equities. A nationalisation programme following the Islamic revolution of 1979, together with restrictions on interest-bearing assets slowed market developments, which were further hampered by the 1980-1988 Iran/Iraq war. The TSE was re-launched after the war but the inherited regulatory framework proved inadequate. During the late 1990s the regulatory framework was reviewed and an automated trading system was introduced. Our data begins in 2000 following this institutional overhaul. The TSE then began to expand its activities until the election of an ant-reformist government in 2004 triggered economic uncertainties that were exacerbated in the following years by a sequence of U.N. and other sanctions prompted by the Iranian nuclear programme. During 2004-2006 many companies de-listed from the TSE; market size and trading volume both decreased. A brief bear market followed the global financial crisis of 2007-2008 but recovery was strong and the final period covered by our data shows an expanding market with a rising index – possibly in part a consequence of Iranian investors repatriating their assets because of the impact of sanctions on their overseas investments.

The possibility of a broken or otherwise non-linear trend motivates our use of the E&L flexible Fourier trend approach to unit root tests. This approach is constructed within the framework for Lagrange Multiplier (LM) unit root tests introduced in Schmidt and Phillips (1992) - hereafter S&P. The LM framework begins with a statement of the data generation process (DGP) as some variant of (2) below:

$$\begin{gathered} P_t = N_t + e_t, t = 1,2. \cdots, T \\ e_t = \rho e_{t-1} + \varepsilon_t \end{gathered} \tag{2}$$

In equation (2), the series $\{\mathrm{P_t}\}$ is to be assessed for stationarity; it is modelled as being driven by unobservable random shocks $\{\mathrm{e_t}\}$ and also a linear combination of observable exogenous variables: $\mathrm{N_t} = \alpha + \beta t + X_t, \mathrm{X_t} = \sum \gamma_j Z_{jt}$. The random shocks are autocorrelated, possibly with a unit root ($\rho = 1$). S&P develop in detail the case where $\{\mathrm{N_t}\}$ is a linear trend: $\mathrm{N_t} = \alpha + \beta \mathrm{t}$, or a polynomial time trend: $\mathrm{N_t} = \alpha + \beta t + \sum_{r=2}^{r=R} \gamma_r t^r$. E&L augment the linear time trend with trigonometric components intended to proxy non-linearities such as changes in the trend slope. These trigonometric terms may employ a single frequency (k), viz: $\mathrm{N_t} = \alpha + \beta \mathrm{t} + \left(\gamma_\mathrm{s} \sin(2\pi \mathrm{kt/T}) + \gamma_\mathrm{c} \cos(2\pi \mathrm{kt/T})\right), \mathrm{k} \leq \mathrm{T}/2$, or multiple frequencies.

LM tests in general begin by first estimating the model subject to the restrictions of the null hypothesis. Here, $H_0: \rho = 1$ makes equation (2) one in differences: $\Delta P_t = \Delta N_t + \varepsilon_t$, equivalently $\Delta P_t = \beta + \Delta X_t + \varepsilon_t$. Assuming the $\{\varepsilon_\mathrm{t}\}$ to be independently and Normally distributed means that Maximum Likelihood (ML) estimation of β and of the parameters within

$\{\mathrm{X_t}\}$ is achieved by application of OLS to this equation. ML estimation of the intercept (α) under the null hypothesis then proceeds by considering the first observation point, at which $P_1 = \alpha + X_1 + e_0 + \varepsilon_1$. The value of α and of the initial shock (e_0) cannot be separately identified under the unit root null; restricted ML estimation of their joint contribution is given by $(\widetilde{\alpha + e_0}) = P_1 - X_1$.

The second stage of an LM test is to evaluate how significantly the maximised likelihood could increase if the null hypothesis were to be relaxed. S&P show that for equation (2) this derivative is equivalent to the coefficient, ϕ, attaching to the term $\tilde{S}_{t-1}$ in OLS estimation of an auxiliary regression: $\Delta P_t = \phi \tilde{S}_{t-1} + \delta + \sum \delta_j \Delta Z_{jt} + u_t$. They show also that the LM test statistic can be calculated as the t-statistic for $\hat{\phi}$, denoted τ_{LM}. In this auxiliary regression, $\{\tilde{S}_t\}$ is the series of deviations around the fitted trend whose parameters were estimated in stage 1: $\tilde{S}_t = P_t - \widetilde{N}_t$. Critical values for τ_{LM} depend upon sample size and the specification of $\{N_t\}$: S&P provide critical values for the case where $\{N_t\}$ is a linear or polynomial time trend, E&L provide critical values for use when a linear trend is augmented with trigonometric terms and demonstrate that the asymptotic distribution of this test statistic depends only on the trigonometric frequency, k, being invariant with regards to all other parameters in the DGP.

Hence, application of the flexible Fourier trend unit root test with a single trigonometric frequency (k) involves

1. OLS estimation of $\beta, \gamma_s, \gamma_c$ in $\Delta P_t = \beta + \gamma_s \Delta sin(2\pi kt/T) + \gamma_c \Delta cos(2\pi kt/T) + \varepsilon_i$;
2. Estimation of $(\alpha + e_0)$ as $P_1 - \tilde{X}_1$, $\tilde{X}_t = \tilde{\beta} t + \widetilde{\gamma_s} sin(2\pi kt/T) + \widetilde{\gamma_c} cos(2\pi kt/T)$;
3. Construction of the fitted trend: $\widetilde{N}_t = (\widetilde{\alpha + e_0}) + \tilde{X}_t$, and the deviations from trend: $\tilde{S}_t = P_t - \widetilde{N}_t$
4. OLS estimation of the auxiliary regression: $\Delta P_t = \phi \tilde{S}_{t-1} + \delta + \delta_s \Delta sin(2\pi kt/T) + \delta_c \Delta cos(2\pi kt/T) + u_t$;
5. Comparison of τ_{LM} (the t-statistic for $\hat{\phi}$) with critical values tabulated in E&L

In practice, k is not given *a priori*. E&L recommend that steps 1 – 4 are repeated for each integer value of k in the interval $1 \leq k \leq 5$. The preferred value for k is the one that produces the smallest residual sum of squares for the auxiliary regression in step 4. E&L offer evidence to suggest that such a Fourier approximation – employing only the preferred single frequency, is able to capture the effects of a (small) unknown number of structural changes at unknown dates, particularly when these are not sudden breaks.

Before implementing this unit root testing procedure we consider some issues pertinent to the data series.

3. The consequences of non-trading days

Our data series gives the TEPIX daily closing values, obtained from the official TSE website (www.irbourse.com). The sample data (see figure 1) cover the time period from January 2, 2000 to December 31, 2012, offering 3145 observations. As is commonplace for stock market data, the data source publishes the observations as a contiguous sequence without explicit recognition of data gaps caused by non-trading days, such as weekends and other periods of market closure. Application of the selected unit root test thus begs the question of whether or not to read equation (2) as applying to the data in its published form – in which dates and lags are expressed by reference to sequential observation numbers rather than the passage of calendar time.

We take the position that market prices are a reflection of domestic and international fundamental forces that evolve largely independently of the market itself. In particular, we assume that these market fundamentals – which will include, for example, national and global macroeconomic fluctuations, proceed whether or not the market is trading. Consequently, we conclude that the passage of time is, in principle, better represented by calendar date than by observation number. The issue of whether to specify lag structures by reference to observation number or calendar time may be less clear cut. If behavioural forces are dominant in price setting then the last observed price may have relevance, even if a period of closure has intervened – a consideration that argues for specifying lag structures by reference to observation number instead of, or as well as, calendar time. In this study we maintain an assumption that price-setting reflects fundamental asset values and test whether or not it does so efficiently. Some of the factors influencing fundamental value are best modelled as a deterministic trend, $\{N_t\}$, other factors are best modelled as unpredictable 'news', $\{\varepsilon_t\}$. The trend and news flow exist whether or not the market is trading. From this perspective, we argue that equation (2) should be read as applying to all calendar dates; if the market is not trading then this equation models an unobservable 'what the price would have been'. The autoregression in equation (2) is therefore understood to be defined on a sequence of values indexed by calendar time rather than observation number. We will consider the implications of this understanding of the DGP when the unit root test of E&L is applied to data which has recurrent missing values because of non-trading days. Ryan & Giles (1999) investigate the implications of recurrent data gaps for the Dickey-Fuller style of unit root tests; we are not aware of any such investigation for tests conducted within the LM framework of S&P.

We will consider the properties of the E&L 'flexible Fourier trend' test in the context of three possible decisions by an investigator: a) the investigator records non-trading days as missing values in a daily data series indexed by calendar date; b) the investigator ignores the presence of non-trading days and works with the series as published – a daily series indexed by observation number; c) the investigator chooses to analyse weekly returns rather than daily returns. These alternatives are now discussed in more detail before a Monte Carlo experiment explores their finite sample consequences.

Case A: daily observations with weekly non-trading gaps in the data

When equation (2) describes a data series with no missing values, then – as is argued more completely in Appendix 1 of S&P, the contribution that the t^{th} observation makes to the joint likelihood is $\varepsilon_t = P_t - N_t - \rho(P_{t-1} - N_{t-1})$. At $t = 1$, however, $P_{t-1} = P_0$ is not observed and so we must use $\varepsilon_1 = P_1 - N_1 - \rho e_0$, making e_0 another parameter to be estimated. This need for incidental parameters repeats at any gap in the sequence of observations. If $t = s$ is the date of the first observation following a group of missing observations then P_{s-1} is unobserved. We should therefore employ $\varepsilon_s = P_s - N_s - \rho e_{s-1}$ in constructing the likelihood function implied by equation (2), making each e_{s-1} a parameter to be estimated. Where the data are the daily prices reached in a stock market then these normally follow an 'A:B' trading pattern, in which a period of A trading days is followed by a period of B non-trading days. Consequently, the number of non-trading periods and, therefore, the number of parameters requiring estimation within the LM framework increases at the same rate as the sample size, invalidating the arguments used by S&P to develop the LM test statistic, of which the τ_{LM} statistic (E&L, eq.11) is a special case. An investigator who simply applies the procedure documented in E&L to data with missing values fails to recognise the presence of these additional parameters.

Case B: ignoring the non-trading gaps in the daily observations

It is not unknown within the literature for an investigator to work with the daily observations from a market with a $5{:}\,2$ trading:non-trading pattern, treating these as if they are a contiguous series of observations. In this case, each observation point, except the first, contributes data for both current and lagged price, eliminating the problem of incidental parameters described in *Case a*. However, the first trading day of each week now employs the price from the final trading day of the previous week as if it were the price from the immediately preceding day. In this case, equation (2) – now referencing observation number rather than calendar time, should in principle be modified for all weekly first trading days to read $P_s = N_s + e_s$, $e_s = \rho^3 e_{s-1} + \rho^2 \varepsilon_{NT:s,1} + +\rho \varepsilon_{NT:s,2} + \varepsilon_s$, where s indicates the first trading day of some week, $\varepsilon_{NT:s,1}$, $\varepsilon_{NT:s,2}$ are the unpredictable news flows of the two preceding non-trading days, and e_{s-1} refers to the most recent preceding observation, which is from the final trading day of the previous week. This modification means that the joint likelihood function for the observed data is not exactly as is assumed in the LM framework introduced by S&P. Additionally, the investigator who is using stock market data in the form in which it is published may use the observation sequence number rather than the calendar date to indicate the passage of time when constructing deterministic trend contributions to $\{N_t\}$. This again constitutes a departure from equation (2). For example, the calendar date on the opening trading day of a week will be three calendar days ahead of the last trading day, whereas equation (2) assumes a gap of one day between all observations. Whilst this irregularity might be accommodated within equation (2) by suitable definition of deterministic contributions to $\{N_t\}$, it constitutes a departure from the assumptions underpinning the tabulated critical values in S&P and in E&L.

Case C: using weekly observations

Empirical investigations of financial markets for which daily prices, and thus daily returns, are observed may also report the behaviour of returns taken over a longer period, for example weekly. In the notation of equation (2), P_t then might be the mid-week log-price of the current week and P_{t-1} the mid-week log-price of the preceding week, making $\{\Delta\mathrm{P}_t\}$ a series of weekly returns. The passage of time can be measured in days or weeks without consequence for the test results since one measure is a simple rescaling of the other. Assuming that incidental market closures do not cause any missing mid-week values, every week excepting only the first, contributes both current and lagged observations of $\{\mathrm{P}_t\}$ – as was assumed in the development of the LM test by S&P. At the same time, however - since each week now provides only one observation, the sample size is dramatically reduced, with consequent power reduction for the test. Additionally, iteration of equation (2) shows that its autocorrelation component can be expressed as $e_t = \rho^7 e_{t-7} + \bar{\varepsilon}_t$, with $\bar{\varepsilon}_t = \sum_{i=0}^{i=7} \rho^i \varepsilon_{t-i}$, implying that, under the unit root null, the random shock to weekly returns ($\bar{\varepsilon}_t$) has higher variance than the shock to daily returns (ε_t), again weakening the test. On the other hand, whereas the autoregressive coefficient in equation (2), has a value of ρ for daily data, it has a value of ρ^7 for weekly returns.

Since $\rho < 1$ under the alternative hypothesis then $\rho^7 < \rho$, presumably making it easier for the test to detect departures from $\rho^7 = \rho = 1$. This consideration points to the possibility that, although weekly data offer fewer observations than do daily data, and with higher variance in the disturbance term, the power of the LM test for weekly data might still not be much reduced, if at all, relative to when it is applied to daily data over the same span of weeks.

The consequences of an investigator's response to weekend non-trading gaps in the data is now explored through Monte Carlo experiments. (The Monte Carlo simulations and the statistical investigations that follow were conducted using Eviews v8.) Each experiment consists of 50,000 trials in which we apply the flexible Fourier unit root test with multiple frequencies to a 'daily' data series, $\{P_t\}$, that is artificially generated according to (3).

$$P_t = N_t + e_t, t = 1,2. \cdots, T$$

$$e_t = \rho e_{t-1} + \varepsilon_t$$

$$\mathrm{N_t} = \alpha + \beta \mathrm{t} + \sum_{k=1}^{k=n} \gamma_{sk} sin(2\pi\, k/T) + \gamma_{ck} cos(2\pi\, k/T) \tag{3}$$

$$n = 2$$

Equation (3) is a variant of equation (2) employing multiple trigonometric frequencies. We set the maximum frequency to $n = 2$ since this is sufficient to generate data series similar in appearance to the TEPIX data shown above in figure 1. In each trial the γ_{sk} and γ_{ck} parameters are drawn independently from a uniform distribution on the range $(-5, +5)$, reflecting the range of parameter values considered in the power simulations in E&L. The linear trend parameter, β, is drawn from a uniform distribution on the range $(-0.25, +0.25)$; this produces a 'standardised drift', meaning the average value of the $\{\Delta P_t\}$ series expressed as a ratio to its standard deviation, which is on average close to 0.15, a value in the range that S&P (page 271) suggest might be expected for financial market data. The intercept parameter, α, is set to zero in each trial, as also is the initial deviation from trend, e_0. The random disturbances, $\{\varepsilon_t\}$, are generated as independent drawings from the standard Normal distribution.

In each trial, data is generated for a sample period spanning 100 seven-day 'weeks' - giving a total of 700 observation points. Each week has a 5:2 trading:non-trading pattern. For simulation of *case A*, which has $T = 700$, the unobservable values on day 6 and day 7 each week are replaced by missing value codes, leaving 400 useable observation points, i.e. days 2, 3, 4, 5 of each week. For simulation of *case B*, the missing value at day 7 each week is replaced by the observed value of the preceding day 5. This simulates an investigator using the data series as published – giving 499 useable observation points (every trading day except the first), with $T = 500$. *Case C* is simulated by restricting the sample to include only day 3 of each week, implying $T = 100$, with 99 useable observation points. In each trial, a test of $H_0: \rho = 1$ employing multiple frequencies ($k = 1$ jointly with $k = 2$) was conducted for each of cases A, B, C, following the method as described in E&L, with 5% critical values drawn from or interpolated from their table 2. This experiment of 50,000 trials was repeated for various values of the autoregressive root and a count was kept of the percentage of rejections; the results are summarised in table 1.

Table 1. Empirical rejection percentages using 5% critical values from Enders & Lee (2011) with 100 weeks of 5:2 daily data

$\rho =$	1	0.995	0.99	0.975	0.95	0.9	0.75	0.5
Case A: daily, with gaps	0.37	0.34	0.38	0.78	3.03	15.54	37.28	45.13
Case B: daily, no gaps	5.41	5.76	7.04	17.2	64.88	99.95	100	100
Case C: weekly returns	5.03	5.29	6.42	15.16	54.19	98.86	100	100

From table 1 it is evident that failure to recognise, within the likelihood function, the additional pseudo-parameters occasioned by the missing value gaps in *Case A* greatly reduces the power of this LM-based test when those gaps are occurring repetitively and frequently, as with weekend market closures. We should also note that the insertion of missing value codes means that contiguous data appear in groups of only five observations. This implies a greatly reduced sample size for any model including lagged variables. For example, the E&L test equation cannot be estimated on such data if it is augmented with lagged differences to a lag depth greater than three.

Table 1 also suggests that proceeding as in *Case B* (ignoring data gaps) gives marginally better power than is obtained when employing weekly returns. The price being paid for this slight advantage is that *Case B* produces a somewhat over-sized test, rejecting the unit root hypothesis more than 5% of the time when it is actually true. In the context of assessing market efficiency, this slight conservatism may be acceptable.

The suite of experiments reported in table 1 was repeated with other sample sizes (50 weeks, 200 weeks, 500 weeks). At the larger sample sizes *Case A* remains a weak performer, *Cases B, C* both improve in power, with Case B continuing to be slightly oversized and retaining a small power advantage relative to *Case C* when ρ is close to unity. At the smaller sample size (50 weeks) *Case C* apparently becomes substantially over-sized which may be a consequence of interpolating a critical value below the limit of applicability of table 2 in E&L, where the minimum sample size is $T = 100$.

Based on this Monte Carlo evidence, we will analyse the TEPIX daily data series as published, ignoring the gaps created by days when the market is closed and defining trend variables, trigonometric and linear, by reference to observation number rather than calendar time.

4. Empirical results

4.1 Unit root testing on the full sample

To model nonlinearity of unknown form by a Fourier approximation employing a single frequency, E&L recommend selecting whatever frequency in the range $k = 1$ to $k = 5$ yields the smallest sum of squared residuals in least squares estimation of equation (4). They also recommend pre-testing for non-linearity. We dispense with this pre-testing for the full sample of TEPIX data on the grounds that figure 1 provides sufficient evidence to discount a linear trend.

$$\Delta P_t = \phi \tilde{S}_{t-1} + \delta + \delta_s \Delta sin(2\pi kt/T) + \delta_c \Delta cos(2\pi kt/T) + u_t \qquad (4)$$

As described previously, in section 2, $\{\tilde{S}_t\}$ is a series of deviations from a non-linear trend that employs the nominated frequency and embodies the unit root null hypothesis. Once the preferred frequency is established, equation (4) may be augmented with lags of $\Delta\tilde{S}_t$ if this is necessary to remove residual autocorrelation. The test-statistic, τ_{LM}, is the t-statistic for $\hat{\phi}$ in the (possibly) augmented version of equation (4), with critical values for the single frequency case as tabulated in table 1 of E&L. For our data the frequency giving the lowest sum of squared residuals in equation (4) is $\hat{k}_{1st} = 4$, with $\hat{k}_{2nd} = 1$ being a close runner-up. Because their SSR values are similar, we report the test results for all frequencies in table 2, along with critical values interpolated from table 1 in E&L.

Table 2. Unit root tests for TEPIX, full sample, employing single frequencies (k)

			Critical values			
k	SSR	τ_{LM}	1%	5%	10%	lags
1	0.09981	-2.250	-4.56	-4.02	-3.77	10
2	0.10053	-1.146	-4.16	-3.54	-3.22	10
3	0.10105	-1.121	-3.94	-3.30	-2.98	10
4	0.09951	-1.056	-3.80	-3.19	-2.88	10
5	0.10114	-1.106	-3.74	-3.13	-2.83	10

The final column of table 2, reports the maximum lag length (p) in the augmented version of equation (4). Our specification search strategy for determining (p) was to begin with $p = 0$ and to apply increments, $p \rightarrow p + 1$ with the compound stopping rule 'Stop when there is no residual autocorrelation at the current maximum lag and also no AIC improvement would result from incrementing the maximum lag'. (Residual autocorrelation was deemed absent if the prob. values for the Ljung-Box Q-statistics were above 5% at each of lags 1 through 25.)

It is apparent in table 2 that the unit root null hypothesis, characterising a weakly efficient market, is not rejected by a test where a non-linear trend is approximated by a Fourier approximation employing any single frequency in the range $k = 1$ to $k = 5$. It is, however, well-known that inappropriately restricting the deterministic trend component of a series can bias unit root testing towards erroneous non-rejection of the unit root when the DGP is in fact stationary around a trend process more complex than can be accommodated within the testing framework. For this reason we also apply the flexible Fourier unit root test with multiple frequencies in the deterministic trend. E&L provide critical values for a test equation containing sine and cosine terms at multiple frequencies. The test statistic, denoted $\tau_{LM(n)}$, is now the t-statistic for the estimator of ϕ in

$$\Delta P_t = \phi \tilde{S}_{t-1} + \delta + \sum_{k=1}^{k=n} \gamma_{sk} sin(2\pi\, k/T) + \gamma_{ck} cos(2\pi\, k/T) + u_t \qquad (5)$$

Again, the $\{\tilde{S}_t\}$ are deviations from a fitted trend obtained as outlined in section 2 but now with multiple frequencies, $k = 1 \cdots n$, for the sine and cosine terms. Because additional trigonometric terms reduce the power of the test, E&L limit the cases considered to $n \leq 3$. In table 3, we report $\mathrm{\tau_{LM(n)}}$ for $\mathrm{n} = 1,2,3$. Critical values for our sample size are obtained by interpolating those provided by E&L. The test equation (5) is augmented with lagged terms, $\Delta\tilde{\mathrm{S}}_{\mathrm{t-j}}$, $\mathrm{j} = 1, \cdots \mathrm{p}$, with p selected by the same stopping rule as previously. It is evident that the more flexible trend specifications ($\mathrm{n} = 2,3$) do not reverse the previous conclusion of a weakly efficient market. One caveat that shall be noted here and considered more fully below is that the unit root testing has found evidence of a more complex autocorrelation structure than the simple AR1 process envisaged in equation (2). Such autocorrelation is potentially a basis for predicting price movements from their own past history and thus calls into question market efficiency, regardless of the results of unit root testing.

Table 3. Unit root tests for TEPIX, full sample, employing all frequencies $\leq n$

		Critical values			
n	$\tau_{LM(n)}$	1%	5%	10%	Lags
3	-3.281	-5.87	-5.37	-5.12	10
2	-3.067	-5.27	-4.76	-4.51	10
1	-2.250	-4.56	-4.02	-3.77	10

4.2 Sub-sample testing

We have discovered that the daily TEPIX data for the period from 2/1/2000 to 31/12/2012 satisfies the unit root requirement for a weakly efficient market when the unit root test assumes that non-linearities in the trend can be adequately approximated by Fourier series employing a single frequency or a small number of frequencies, as in equation (3) above. It seems appropriate, in the context of an emerging stock market, to investigate whether the evidence for or against market efficiency is uniform throughout the period of available data. By employing a rolling sub-sample window for unit root testing, we can potentially address questions such as (i) whether, following its regulatory reforms, the Tehran exchange required some operational experience in order to develop an institutional framework supportive of efficient price-setting, and (ii) whether the efficiency of price-setting was challenged by particular historical episodes.

We employ a sub-sample window with a width of 250 observations, approximately a year's trading, and roll this window through the available data, advancing the window with a step size of a single observation. For each position of the sub-sample window, we perform the test with a single frequency. The preferred single frequency for each sub-sample is selected as described in section 4.1. With the substantially reduced sample size we do not attempt the test with multiple frequencies, both because of power considerations and also because, in the light of the arguments and examples in E&L, we expect adequate approximation by a single frequency at this sample size. We pre-test for a null hypothesis of linearity, using the F-test described in E&L, with critical values interpolated from their table 1 (panel c). Whenever linearity is not rejected we employ the LM test with a linear trend, as described in S&P, in place of the LM test with a flexible Fourier trend.

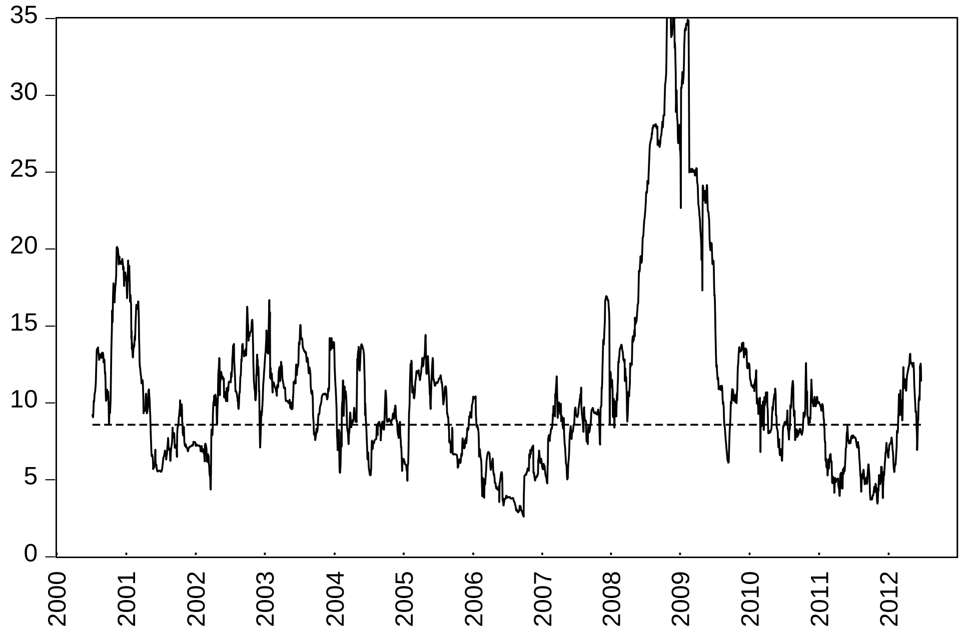

Figure 2. Sub-samples F-test, null hypothesis = 'linear trend'

Figure 2 shows the results of the pre-testing for linearity; the F-test statistic from each sub-sample window is plotted against the mid-point of that window. The horizontal dotted line shows the critical value for a test with 5% significance level. For some periods of time a linear trend appears to be sufficient - implying that unit root testing with a simple linear trend is to be preferred (see E&L), but these episodes are not in the majority and there are other periods where the rejection of linearity is emphatic.

Figure 3 plots the t-statistic obtained within each sub-sample window against that window's mid-point. The t-statistics result from employing the best single frequency Fourier trend for that sub-sample, or a simple linear trend, according to the results of the pre-testing for linearity. Because the critical values for the t-statistics vary according to the selected test, the results are reported as the deviation of each t-statistic from its median value under the null hypothesis, with this deviation expressed as a percentage of the distance between the median and the one-tailed critical value for a 5% significance level. So, for example, a test result that falls exactly on whatever is the appropriate critical value would be recorded as 100% and lie exactly on the dashed line in figure 3. Plotted values that fall above this line indicate sub-samples for which the unit root hypothesis is rejected. To support these calculations, table 1 in E&L was extended to include a more complete range of quantiles for the test-statistic under the null. These are appended as table A1 since they are potentially useful for any investigation employing the flexible Fourier trend unit root test.

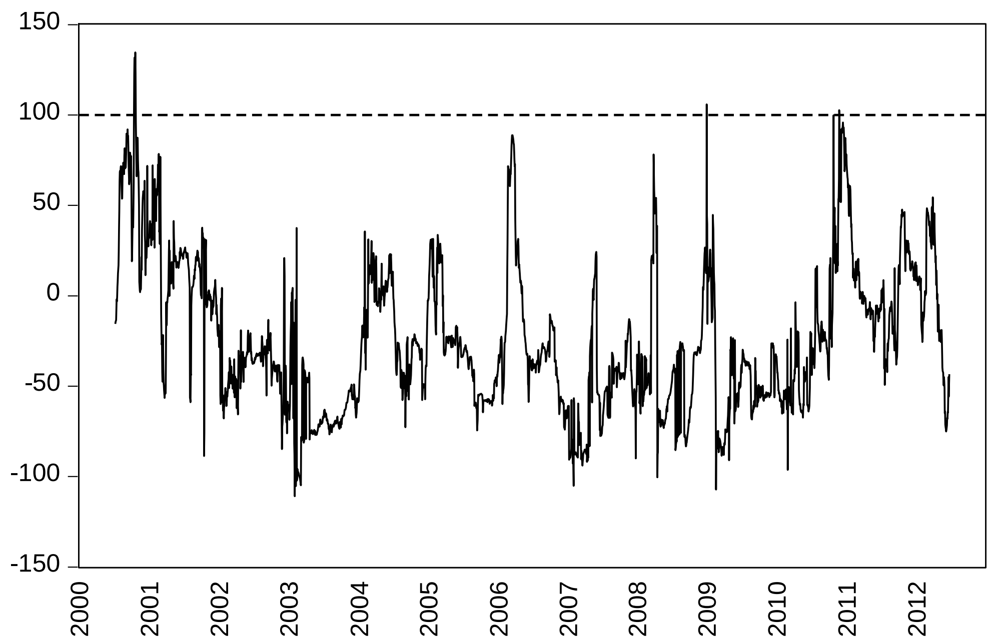

Figure 3. Deviation of t-stats from median, relative to 5% quantile

In figure 3 we see that there are only a small number of sub-sample windows for which the unit root test statistic enters the 5% rejection region. There are, however, a number of episodes where the evidence against a unit root has spiked – most recently, for example, in subsamples whose mid-point is in late 2010 / early 2011. Figure 3 is also suggestive of cumulative strengthening of support for the null hypothesis in the early years, up until 2004, and a cumulative weakening of support in recent years, from 2008 to 2012. It is tempting to associate the earlier of these trending episodes with a market that is developing its institutional practices and depth of trading. The more recent cumulative weakening of support for weak efficiency coincides with the aftermath of the global financial crisis and covers the period during which market activity and the TEPIX index have rapidly increased as nationals repatriate their overseas investments.

4.3 Assessing the predictability of prices

We noted earlier that the test equation (4) is augmented with lags of $\Delta\tilde{\mathrm{S}}_{\mathrm{t}}$ whenever there is evident need to remove autocorrelation from this equation's disturbance process. Such augmentation was required in over 95% of the sub-sample windows. The median lag-length was 2 trading days, with an upper quartile of 3 days, and almost half of the remaining sub-samples requiring a lag-length of 10 days. Such autocorrelation is evidence that the DGP In equation (2) should be amended to be $P_t = N_t + e_t$, $\mathrm{A}(L)e_t = \varepsilon_t$, with $\mathrm{A}(L)$ being a lag polynomial that expresses the pattern of the autocorrelation. In the case where we have accepted the presence of a unit root in this lag polynomial then we can factor it as $\mathrm{A}(L) = \mathrm{B}(L)(1 - L)$. Noting that the leading coefficient is normalised to unity, we can write $\Delta e_t = \sum_{j=1}^{j=J} b_j \Delta e_{t-j} + \varepsilon_t$ and recalling that S_t has been introduced earlier as a notation for deviations from trend when a unit root is assumed present allows the DGP in first differences to be written as:

$$\Delta P_t = \Delta N_t + \sum_{j=1}^{j=J} b_j \Delta S_{t-j} + \varepsilon_t,\ t = 2,3,\cdots,T \tag{6}$$

Equation (6) appears to be inconsistent with market efficiency inasmuch as deviations from trend in the one-day returns ($\Delta S_t = \Delta P_t - \Delta N_t$) are dependent upon their own past history. A counter argument is that equation (6) is only *de facto* inconsistent with weak efficiency if the b_j are known *a priori* and are thus a basis for predicting prices. Where the b_j must be estimated from market history the question to be addressed is then whether or not the past history contains sufficient information to make equation (6) an effective predictor. We therefore examine the possibility that 'the Tehran stock exchange is weakly efficient in the sense that past deviations of returns from their deterministic trend do not provide an effective basis for predicting future such excess returns'. We approach this as an exercise in comparing predictive effectiveness. The first of the two predictors which are to be compared is (i): $\Delta\tilde{S}_{t+h} = \sum_{j=1}^{j=J} \tilde{b}_j \Delta\tilde{S}_{t+h-j}$. This is an implication of equation (6) in which $\Delta\tilde{S}_t = \Delta P_t - \Delta\widetilde{N}_t$, where $\Delta\widetilde{N}_t$ is the ML estimator for the flexible Fourier trend under the unit root null hypothesis, as defined previously in section 2 and $\tilde{b}_j$ are OLS estimates. This is a dynamic predictor, meaning that, for h>1, it predicts $\Delta\tilde{S}_{t+h}$ by using its own predicted values for those $\Delta\tilde{S}_{t+h-j}$ that are outside of the estimation sample period. As such, it reflects the position of a market participant who only has access to current and past prices when forming expectations of future prices. The effectiveness of this dynamic predictor is compared with (ii): $\Delta\tilde{S}_{t+h} = 0$, which simply sets predictions equal to the unconditional mean, thus asserting that past prices have no useful predictive power.

The two predictors just defined will be compared by the 'DM test' (Diebold and Mariano 1995). This requires that we define a non-negative loss function to represent the cost of forecast error – we shall simply use squared error. The null hypothesis is 'no expected difference between the losses resulting from the two predictors', as is appropriate to a weakly efficient market. The appropriate alternative hypothesis for assessing market efficiency is one-tailed, namely that expected losses are less when employing the predictor utilising past market history.

We consider a range of forecast horizons, from h =1 to h=25 trading days. For each of the 2862 rolling sub-samples of 250 days, in which the unit root hypothesis was not rejected (see figure 3), we retrieve $\Delta\widetilde{N}_t$, the ML estimate of a flexible Fourier trend in differences using the best single frequency, as discovered in the unit root testing for that sub-sample. This provides $\Delta\tilde{S}_t$ within that sub-sample and also within its immediately following forecast period. We then use those deviations from trend within the sub-sample as data for OLS estimation of the coefficients in $\Delta\tilde{S}_t = \sum_{j=1}^{j=J} \tilde{b}_j \Delta\tilde{S}_{t-j}$, with the maximum lag ($J$) set to the lag length previously determined in unit root testing for that sub-sample. This autoregression then provides a dynamic forecast for each of $\Delta\tilde{S}_{T+h}; h = 1,2 \cdots 25$, where T is the last date in that rolling sub-sample. For each of these horizons, the rolling sub-samples generate a sequence of forecast errors for the two predictors being compared. A sequence of 'loss differentials' for any given forecast horizon is given by subtracting the sequence of squared errors for predictor (ii) from those for predictor (i). The null hypothesis of equal forecast accuracy proposes that these loss differentials have an expected value of zero; an alternative hypothesis that predictor (i) is superior implies a negative expected value.

Table 4. Diebold-Mariano test of equal forecast accuracy at various forecast horizons

h	1	2	3	4	5	6	7	8	9	10
DM-tstat	-2.839	-0.635	0.011	-0.293	-0.742	-0.516	-1.730	-1.048	-0.494	-0.425
prob	0.0023	0.2628	0.4955	0.3847	0.2292	0.3028	0.0418	0.1475	0.3108	0.3355

h	11	12	13	14	15	16	17	18	19	20
DM-tstat	0.525	0.771	0.823	1.072	0.551	0.763	0.196	0.275	-0.083	0.716
prob	0.2998	0.2205	0.2054	0.1419	0.2909	0.2229	0.4224	0.3916	0.4667	0.2371

h	21	22	23	24	25
DM-tstat	0.789	0.996	-1.001	0.486	0.381
prob	0.2151	0.1598	0.1584	0.3137	0.3518

Diebold (2012) notes that the DM test statistic can be conveniently calculated as the t-statistic in an intercept-only regression model for the loss differential, estimated by OLS with HAC-robust standard errors. These are presented in table 4, together with their left-tail p-values, showing that the null hypothesis that past market history gives no useable information for predicting future market behaviour is supported at most forecast horizons in $h = 1,2 \cdots 25$. There are two exceptions. The hypothesis of equal forecast accuracy is clearly rejected (p=0.0023) for $h = 1$; recent deviations of market price from its deterministic trend are relevant predictors for this deviation on the next trading day. For $h = 7$, the rejection of the null is borderline (p=0.0418) and less easy to rationalise (The passage of time is measured in trading days so $h = 7$ should not be read as a 'week-day effect'.) We can note, however, that one false rejection of a null hypothesis in twenty five applications of a test would not be surprising.

5. Summary and conclusions

We have investigated whether pricing in the Tehran stock market is weak-form efficient, using unit root testing and also an assessment of the predictability of the non-deterministic component of the market index. The history of the market index is such that unit root testing should admit the possibility of structural breaks. The ‘flexible Fourier trend’ unit root test of Enders and Lee (2011) allows the timing and also the degree of abruptness of such non-linearities to be data-determined.

The theoretical development of this test assumed a data generation process with no missing values but stock market data at daily frequency typically has repeated missing value gaps because of weekend market closure. We argue that the fundamental forces driving asset values play out in calendar time, whether or not the market is trading and discuss the implications for the test procedure when applied to data with non-trading gaps. By Monte Carlo simulation of a market with regular weekend closure we establish that the test is grossly underpowered if the non-trading gaps are filled with missing value codes and has much better power when applied to either (i) weekly returns dated in calendar time, or (ii) daily returns as published, in which observation number replaces calendar time. The latter appears to be slightly more powerful and so was employed, despite its disrespect of the test’s assumptions regarding data generation.

We apply the test first to the full sample of available data and conclude in favour of a unit root, whether employing a single trigonometric frequency or multiple frequencies for the Fourier trend. We note the need to augment the test equation with lagged terms, and argue that this begs the question of whether a unit root is sufficient evidence of weak-form efficiency.

We also conduct unit-root testing on rolling sub-samples of 250 observations. Pre-testing for linearity of trend in each sub-sample determines whether unit root testing follows E&L or S&P. We again conclude in favour of a unit root in almost all sub-samples but note that the evidence of a unit root strengthens during the period 2000 – 2004, suggesting improvement in market efficiency as operational experience accumulated following the earlier regulatory reforms. We also note a weakening in the support for a unit root after 2008 and conjecture that this may be associated with very rapid expansion in market activity when international sanctions forced repatriation of funds that had been previously invested overseas.

As with the full-sample, the application of unit root testing in sub-samples typically required augmentation with lagged differences. These might provide a basis for prediction of prices, thus questioning weak-form market efficiency, even with a unit root present. We argue that a market is *de facto* weak-form inefficient only if the past prices provide a basis for *effective* prediction of future prices. We use the Diebold and Mariano (1995) test to assess whether this is the case, discovering that the empirical evidence suggests that past price deviations from trend are an effective predictor for future deviations from trend at a forecast horizon of one trading day. To this extent, our analysis suggests that the Tehran Stock Exchange has not displayed weak-form efficiency.

References:

Dickey, D. A. and Fuller, W. A., Distribution of the estimators for auto-regressive time series with a unit root. *J. Am. Stat. Assoc.*, 1979, 74, 427-431.

Diebold, F.X. and Mariano, R.S., Comparing predictive accuracy. *J. Bus. Econ. Stat.*, 1995, 13, 253-263.

Diebold, F. X., Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. Working Paper #18391, NBER, 2012.

Enders, W. and Lee, J., A unit root test using a Fourier series to approximate smooth breaks. *Oxf. Bull. Econ. Stat.*, 2011, 74(4), 574-599.

Fama, E. F., Efficient capital markets: A review of theory and empirical work. *J. Financ*., 1970, 25(2), 383–417.

Lee, J. and Strazicich, M.C., Minimum Lagrange multiplier unit root test with two structural breaks. *Rev. Econ. Stat.*, 2003, 85(4), 1082–1089.

Leybourne, S., Newbold, P., and Vougas, D., Unit roots and smooth transitions. *J. Time Ser. Anal.*, 1998, 19(1), 83–97.

Lumsdaine R. L. and Papell, D.H., Multiple trend breaks and the unit root hypothesis. *Rev. Econ. Stat*., 1997, 79, 212-218.

Perron, P., The great crash, the oil price shock and the unit root hypothesis. *Econom*., 1989, 57, 1361–1401.

Ryan, K. F. and Giles, D. A., Testing for unit roots with missing observations. In *Messy Data – Missing Observations, Outliers and Mixed-Frequency Data, Advances in Econometrics 13*, edited by T. Fomby and R. Carter-Hill, pp. 203-242, 1999 (JAI Press Inc.: Stamford, CT.)

Schmidt, P. and Phillips, P., LM tests for a unit root in the presence of deterministic trends. *Oxf. Bull. Econ. Stat*., 1992, 54, 257-287.

Zivot, E. and Andrews, D.W.K., Further evidence on the great crash, the oil price shocks and the unit root hypothesis. *J. Bus. Econ. Stat*., 1992, 10, 251–270.

APPENDIX

Table A1 reports the simulated distribution of the Enders & Lee (2011) ‘τ_{LM}’ test statistic for sample sizes quoted in their table 1 and with an extended selection of quantiles. Each column is derived from 100,000 simulations of equation (2) with $\rho = 1$; $\{N_t\}$ a null series; $\{\varepsilon_t\}$ iid $N(0,1)$, $t = 0,1,\cdots T$ and $e_0 = \varepsilon_0$. The 5% quantiles reported here do not all exactly match those in Enders & Lee (2011) but are all within the range of empirical variation that can be expected with 100,000 repetitions of the DGP.

TABLE A1. Quantiles for τ_{LM}

	T=100					T=200					T=500					T=2500				
	k=1	k=2	k=3	k=4	k=5	k=1	k=2	k=3	k=4	k=5	k=1	k=2	k=3	k=4	k=5	k=1	k=2	k=3	k=4	k=5
1.0%	-4.69	-4.24	-3.97	-3.84	-3.74	-4.60	-4.20	-3.96	-3.83	-3.73	-4.57	-4.14	-3.94	-3.80	-3.74	-4.55	-4.15	-3.93	-3.80	-3.72
2.5%	-4.37	-3.87	-3.61	-3.48	-3.40	-4.31	-3.85	-3.61	-3.48	-3.40	-4.29	-3.82	-3.59	-3.47	-3.41	-4.26	-3.82	-3.59	-3.46	-3.39
5.0%	-4.10	-3.58	-3.30	-3.18	-3.12	-4.07	-3.56	-3.32	-3.19	-3.11	-4.05	-3.54	-3.30	-3.20	-3.13	-4.03	-3.54	-3.30	-3.18	-3.13
10.0%	-3.82	-3.23	-2.96	-2.85	-2.80	-3.79	-3.22	-2.98	-2.87	-2.81	-3.77	-3.21	-2.98	-2.88	-2.83	-3.77	-3.22	-2.98	-2.88	-2.83
20.0%	-3.49	-2.82	-2.57	-2.50	-2.46	-3.47	-2.82	-2.60	-2.51	-2.47	-3.47	-2.82	-2.60	-2.52	-2.49	-3.46	-2.83	-2.61	-2.53	-2.49
30.0%	-3.27	-2.53	-2.32	-2.26	-2.23	-3.25	-2.54	-2.34	-2.27	-2.25	-3.25	-2.54	-2.35	-2.29	-2.27	-3.24	-2.55	-2.36	-2.29	-2.27
40.0%	-3.08	-2.29	-2.12	-2.07	-2.04	-3.07	-2.30	-2.14	-2.08	-2.06	-3.07	-2.31	-2.15	-2.10	-2.08	-3.07	-2.33	-2.16	-2.11	-2.08
50.0%	-2.91	-2.09	-1.94	-1.90	-1.89	-2.91	-2.10	-1.97	-1.92	-1.91	-2.91	-2.11	-1.97	-1.93	-1.92	-2.91	-2.12	-1.99	-1.94	-1.92
60.0%	-2.76	-1.90	-1.78	-1.75	-1.74	-2.75	-1.91	-1.81	-1.77	-1.76	-2.76	-1.92	-1.81	-1.78	-1.77	-2.75	-1.93	-1.82	-1.79	-1.77
70.0%	-2.60	-1.71	-1.63	-1.61	-1.59	-2.59	-1.73	-1.65	-1.62	-1.61	-2.60	-1.75	-1.66	-1.64	-1.63	-2.59	-1.75	-1.67	-1.64	-1.63
80.0%	-2.42	-1.53	-1.47	-1.45	-1.44	-2.42	-1.54	-1.49	-1.47	-1.46	-2.43	-1.56	-1.49	-1.48	-1.47	-2.42	-1.56	-1.50	-1.48	-1.47
90.0%	-2.20	-1.32	-1.28	-1.27	-1.27	-2.19	-1.34	-1.30	-1.28	-1.28	-2.20	-1.34	-1.30	-1.30	-1.29	-2.19	-1.35	-1.31	-1.30	-1.29
95.0%	-2.03	-1.18	-1.15	-1.14	-1.14	-2.03	-1.19	-1.16	-1.16	-1.15	-2.03	-1.20	-1.17	-1.16	-1.16	-2.03	-1.20	-1.17	-1.17	-1.16
97.5%	-1.90	-1.08	-1.06	-1.05	-1.05	-1.90	-1.09	-1.06	-1.06	-1.06	-1.90	-1.09	-1.07	-1.06	-1.07	-1.89	-1.10	-1.07	-1.07	-1.07
99.0%	-1.76	-0.98	-0.96	-0.96	-0.95	-1.76	-0.99	-0.97	-0.96	-0.96	-1.75	-0.99	-0.97	-0.97	-0.97	-1.75	-0.99	-0.98	-0.97	-0.97