CONSTRUCTIVE ENSEMBLES FOR TIME SERIES IN ECONOMETRICS AND FINANCE

H. D. Vinod, Fordham University, Bronx, New York, 10458. Vinod@fordham.edu

Abstract

The ensemble plays a key role as the notional 'population' for the observed time series. I propose a new method of constructing the ensemble by using maximum entropy (ME) methods. The ME distribution satisfies the mean preserving constraint by construction and is computer intensive. My seven-step algorithm for constructing ensembles is designed to satisfy the ergodic theorem and Doob's theorem, without assuming stationarity and without using complicated asymptotics. Proposed methods are particularly convenient for short nonstationary economic time series and can potentially simplify several inference problems in time series analysis. Three examples illustrate them. A consumption function example explicitly shows that: (i) the constructed ensemble retains the basic shape and dependence structure of autocorrelation function (acf) and partial autocorrelation function (pacf) of the original time series, (ii) one can avoid shape-destroying transformations (differencing) and the underlying need for achieving stationarity, and (iii) one can provide confidence intervals for coefficients of lagged dependent variables. A Finance example shows that new methods can provide a confidence interval for the popular tool called value at risk or VaR.

JEL Classification: C10, C22, C32
Key Words: Nonstationary series, Dependent bootstrap, Value at Risk, Lagged dependent variable.

1. Introduction

In standard statistical analysis one attempts to study the properties of a population parameter from the available sample data. In time series analysis, we have only one observation x_t on the random variable X at time t. How can we have a probability theory when there is only one realization? Researchers including Wiener (1930), Kolmogorov (1931) and Khintchine (1934), (WKK), among others developed the ingenious solutions to this problem by using two routes. Kolmogorov focused on Markov processes, extrapolations and joint distributions. Wiener focused on constructing stochastic processes mostly as linear combinations of simpler independent and identically distributed (iid) processes (mostly white noise), as building blocks and on explicit extrapolations. Khintchine considered stationary processes and nonnormal families and gave a weak law of large numbers that does not assume bounded variances. All regarded x_t as a random variable and the entire time series as a random function, Yaglom (1962).

The WKK model considers x_t as arising from a random experiment, which can (at least in principle) be repeated endlessly. The performance of the random experiment will yield distinct values of x_t and hence a distinct time series, each time the experiment is performed. The WKK theory imagines a collection an infinite set of time series called the **ensemble**. Khintchine's law of iterated logarithm provides bounds (involving the log(log(T)) term) within which paths of partial sum sequences can fluctuate. A member of this ensemble set is the observed time series x_t. This paper offers a computer intensive construction of a plausible ensemble, which has been crucial for the traditional probability model for time series. My ensemble is created fromthe maximum entropy (ME) 'empirical' cumulative distribution function (CDF) and I cite Kolmogorov's 'distance theorem' to claim that it is asymptotically valid. For example, I can construct explicit bounds within which 9999 ensemble paths themselves, not just partial sums, lie. Such constructions were impossible to imagine in 1930's without modern computers.

Progress in time series inference was possible by assuming that the observed time series can be converted into stationary series by simple operations such as differencing. The stationary distribution is viewed as the equilibrium distribution, not affected by the initial conditions, when $t \to \infty$. In physical sciences it is plausible that the process becomes stationary once the equilibrium is achieved. Physical diffusions of heat and ocean waves indeed can have some kinds of equilibrium asymptotic properties independent of the initial conditions. However, since human reactions are generally sensitive to initial conditions (endowments), stationarity assump tion is too strong for time series arising in social sciences.

Even in engineering and natural sciences researchers avoid assuming 'strictly stationary' processes (which require that all ensemble moments to be independent of the origin). The 'weakly stationary' (or covariance stationary) process requires only the mean and autocovariance to be independent of the origin. I avoid stationarity altogether by proposing a new seven-step algorithm described in the Section 2.3 to create J (e.g., J=999) time series ω_j as elements of the ensemble Ω. Although the algorithm randomizes the order statistics $x_{(t)}$ of x_t, it uses double sorting to recover the original ordering of x_t. Next, I use the J values to directly construct approximate numerical sampling distributions for many pivotal statistics. There is no need to know their possibly multi-modal, highly non-normal functional form. Their confidence intervals simplify statistical inference in evolutionary systems or short nonstationary economic time series.

Consider the traditional probability model with a simple experiment of tossing a coin two times. Let the outcomes be 1 (head) or 0 (tail). The sample space Ω consists of a rectangle in the positive quadrant from 0 to 1 in the usual Cartesian coordinate system. In time series analysis we imagine the two tosses as being attached to T=2 points in time leading to four possible x_t with t=1,2: (1,0), (0,1), (0,0) and (1,1). The occurrence of (1,1) when one observes heads at both t=1 and t=2 is called an 'event,' and regarded as a subset of Ω. Note that the events as random elements $\omega_j \in \Omega$ are actually time series. In this example, ω_j with j=1,..,4 exhaust the finite ensemble Ω. More generally, we let the date $t \in \{1,2, \ldots, T\}$ be a subset of an arbitrary index set T'. As $T \to \infty$ we have a long time series of 1's and 0's and Ω becomes an infinite dimensional space to characterize an ensemble containing ω_j.

In general, ω_j are time series of all kinds of real or imaginary random experiments. There exist some pathological subsets of an infinite dimensional space for which it is impossible to assign any probability. Hence mathematicians use a σ-field (or σ-algebra) of subsets denoted by A. We can assign probability to only those events that belong to A. In this general set up, we denote by P a function, which assigns a 'probability measure,' (a number in the closed interval [0,1]) to all elements of A. Thus (Ω, A., P) represents the probability space. The observed time series may be more precisely denoted by $X(t, x_t)$, where t is the time subscript and x_t is the magnitude. Clearly, $X(t, x_t)$ is a real valued function defined on $T' \times \Omega$ such that for each fixed t, the $X(t,x_t)$ is a random variable on (Ω, A, P), which is called a realization ω_x from the ensemble.

If we consider thousands of coin tosses over thousands of dates the entire historical time series is actually a single realization ω_x from the ensemble. When we consider infinite sequences of 0's and 1's comprising Ω, it is quite plausible to think of some sub sequences of 1's and 0's repeating themselves. The WKK theory exploited the possibility of repeating sequences to detach from the time dependence and focused on stationary time series. Their model used a measure-preserving backshift or lag transformation operating on X_t which gives X_{t-1}. That is, $L^k X_t = X_{t-k}$ when k=1 for one lag. More generally, k can be a large positive or negative integer to represent backward or forward shift. Note that L^k is a "1-1 onto" mapping $L^k: \Omega \to \Omega$. For harmonic or spectral analysis of time series, the function exp(iu[x+h])=exp(iux)exp(iuh), where $i=\sqrt{-1}$ is important. It says that a time displacement of such a function merely multiplies it by a complex number and is exploited by the spectral analysis of time series.

Let $\{t_1, t_2, .., t_T\} \in T'$. Strictly stationary time series have the following property: Their joint distribution $f\{X(t_1), X(t_2), .., X(t_T);\theta\}$ remains the same if we shift each point in T by any constant k. If the mean and variance are not functions of time, it is covariance stationary, where the autocovariance of lag k can still depend on |k|, but not on time. Stationary times series are integrated of order zero, x_t~I(0). Many economic applications involve a mixture of I(0) and nonstationary I(d) series, where the order of integration d can be different for different series and even fractional, where the strong stationarity assumptions are difficult to verify. The WKK theory mostly needs the zero memory I(0) white noise type processes to build virtual ensembles. Some technical results are true only for circular processes where the points in T are arranged on a circle, Priestley (1981, p.260). In economic applications, circularity assumption means we can go back in history, (e.g., undo the SEC, FCC, or go back to horse and buggy, etc.) and is quite unrealistic. Economists bring realism by testing and allowing for finite "structural changes," often with *ad hoc* tools. See references in Maddala and Kim (1998, part IV), Sen (2003) and other articles in that issue. Similarly it is hard to accept the notion of *infinite* memory of the random walk I(1) when the very definitions of economic series (e.g., GDP) change over time.

The order statistics $x_{(t)}$ are obtained by reordering the magnitudes x_t. For example, if T=3 and x_t={3,5,2}, reordering them leads to $x_{(t)}$={2,3,5} and the original time subscripts t={1,2,3} go to the new set of locations τ={2, 3, 1}. In general, the location set τ is unique, except for ties and the corresponding mapping may be denoted by $O^{rd}\omega_x = O^{rd} X(t, x_t) = X[\tau, x_{(t)}]$, where $\tau \in T'$. We abandon stationarity and replace the related requirement that: $L^k:\Omega \rightarrow \Omega$ by the requirement that the unique <u>location set τ</u> based on transforming the observed data into order statistics should remain <u>fixed</u> for all elements of the ensemble. For example, consider another time series y_t with T time periods representing the element $\omega_y \in \Omega$. We require that $O^{rd}\omega_y = O^{rd} Y(t, y_t) = Y[\tau, y_{(t)}]$, where it obeys the order in the set τ from the index set. In other words, we require $O^{rd}:\Omega \rightarrow \Omega$, with a fixed sequence of locations $\tau \in T'$, defined by the observed time series x_t linking the *relative* magnitudes to the time subscript t. This mapping O^{rd} will be further clarified and illustrated in the sequel.

Remark 1 (ergodic theorem): Let $\bar{x}$ denote the sample mean of the time series x_t over the data range t=1, .., T and let μ denote the unknown population mean over the ensemble Ω. If ω_j denotes one realization of the ensemble with j=1,2,…, J, J+1,…(with J in thousands), then μ is the ensemble mean for the ω_j time series over all possible j in Ω. I apply the maximum entropy (ME) principle to the empirical cumulative distribution function (ECDF) of the x_t data to help preserve essential features of the true CDF. Section 2 explains two specific features preserved: (i) preserving the probability mass around each x_t, and (ii) preserving the expectation $E(\bar{x})$. These are called mass and mean preserving constraints. In my application, $\bar{x}$ is a 'time average' and $E(\bar{x})=\mu$ is the ensemble average. In statistical mechanics, ergodic processes immediately permit "replacing ensemble averages by their corresponding time averages." Hence the corresponding theorem is called the ergodic theorem. Birkhoff (1931) proved the ergodic theorem rigorously for stationary processes. See Dhrymes (1998) or Spanos (1999, p. 424) for the mathematical details (e.g., Cesaro sums) needed for the infinite dimensional Ω, where the very existence of μ needs to be assured. Since I construct thousands of $\omega_j \in \Omega$, all of which satisfy the mean preserving constraint, the existence of μ can be reasonably assumed. If the observable sample mean for j-th sample is not close enough (i.e., within the user defined tolerance, T_{ol}) to $\bar{x}$ due to the presence of extreme (outlier) elements from the tails of the ME distribution, we delete that sample and get a new j-th sample. In short, the limit of the time average in the ensemble exists by construction and the convergence to μ is assured by the "mean preserving constraint" satisfied in the construction of each time series in my ω_j. We shall see that this can be placed in a rigorous axiomatic framework.

Economists have long recognized that economic time series are often nonstationary. Considerable effort is devoted to devising tools for creating stationary series after appropriate testing and adjustments. For example, we are asked to first assess the order of integration *d* of each series. However, a long-memory fractionally integrated process has a fractional *d*, and most tests for stationarity (the null hypothesis that *d*=0), which ignore the possibility of fractional d, are known to have low power, Maddala and Kim (1998), Vinod (1994, 2002b). If the alternative hypothesis allows for breaks, the unit root test conclusions can be reversed, (See, Sen, 2003 and his references). Recent econometric literature includes α, ϕ, and ψ mixing notions of asymptotic independence and mixingales to replace martingales, Spanos (1999), to help the real world economic facts fit the classical WKK theory. I claim that the L^k operator and all such unpleasant adjustments needed by the WKK theory point to a need for a fresh and simpler alternative. My computer intensive algorithm is a first step in the direction of achieving it.

The plan for the remaining paper is as follows. Section 2 explains my constructive approach to ensemble creation as a potentially powerful and yet simple alternative for studying nonstationary time series. The reader will recognize that since the algorithm in section 2.3 creates resamples, it is similar to Efron's bootstrap. In fact, Vinod (2002) presents it only as a bootstrap. Apart from the use of the maximum entropy distribution, my double sorting to recover the time-series dependence properties seems to be important. Section 2.4 explains how to create the joint probability distribution to satisfy the conditions of Doob's theorem for a valid creation of Ω. In short, I convert the data into a set ordered in size, record what this transformation O^{rd} is, use Maximum entropy to construct a population around the data, draw random sample from this ME-density, apply the inverse transformation and construct a random time series. These steps can be viewed as those of a new dependent data bootstrap.

Section 3 starts with an example with T=5 to understand the ensemble creation. Section 3.1 considers two familiar macroeconomic quarterly time series for the US: Y_t = the GDP, and C_t =personal consumption, where T=223 from 1929 to 2002. The WKK fiction imagines an ensemble coming perhaps from other solar systems in the universe. It further imagines a probability model of different outcomes of Y_t and C_t, where the actual series is a realization ω from among the infinite possibilities. Section 3.2 uses an example from Finance. With these examples I show that assuming stationarity is not the only way and transformations like the L^k operator are not the only tools to construct the infinite possibilities. Section 4 contains my final remarks.

2. Ensemble Creation from the ME distribution

My object in this section is to develop a new method for creating as many realizations as needed from $\{x_t\}$ to comprise the potentially thousands of generally distinct elements ω_j of Ω. Hence, I insert the additional subscript j and denote the elements of the j-th realization of x_t by x_{jt} for j=1, .., J, where J is a large number. I use the principle of maximum entropy (ME) in creating x_{jt}. The ME method is similar to Efron's traditional bootstrap. The traditional bootstrap possesses the following three properties, which are especially unsuitable when x_t represents economic or financial time series. We shall see that my resampling algorithm based on the ME distribution avoids all three properties simultaneously.

(**P1**) The traditional bootstrap sample repeats some x_t values while not using as many others. We are considering applications where there is no reason to believe that values near the observed x_t are impossible. For example, consider x_t=49.2, there is no reason why it cannot be 49.19 or 49.24, both of which round to 49.2.

(**P2**) The traditional bootstrap resamples must lie in the closed interval [min(x_t), max(x_t)]. In most economic series this is an artificial restriction with no economic justification. The observed range is random and somewhat smaller or larger x_t cannot be ruled out.
(**P3**) The traditional bootstrap resample shuffles x_t such that any dependence information in the time series sequence ($x_1, \ldots, x_t, x_{t+1}, \ldots, x_T$) is lost. If we try to restore the original order to the shuffled resample of the traditional bootstrap, we end up with essentially the original set $\{x_t\}$, except that some dropped x_t values are replaced by the repeats of adjacent values. Hence, it is impossible to generate a large number J of sensibly distinct resamples with the traditional bootstrap. The next remark is actually a digression explaining why some fancy bootstrap confidence intervals are not used in this paper.

Remark 2 (Why not percentile-t?): Consider a typical bootstrap where one has estimates of a statistic b, its estimates from resamples denoted by b* and wishes to study the parameter β. When bootstrap was first proposed there was a direct comparison of b with β leading to the so-called naïve confidence intervals (CIs). Later bootstrap methods suggested replacing a study of (b−β) with the observable (b−b*) ultimately leading to percentile and percentile-t CIs. Berkowitz and Kilian (2000) note that the percentile and percentile-t methods have important shortcomings for time series applications. Hence, it is convenient to use the naïve CIs in this initial study.

I shall avoid properties P1 to P3 listed above without introducing unnecessary arbitrariness. The Appendix in Vinod (1985) was perhaps the first to use the maximum entropy (ME) arguments to extend the range of the income data. The concept of maximum entropy has been widely used in both social and natural sciences as a measure of uninformativeness in a system. Let $f(x)$ denote its density. The entropy is defined by the mathematical expectation of the Shannon information:

(1) $\text{Entropy} = H = E(-\log f(x))$

Let us impose mass and mean-preserving constraints and select the $f(x)$ which maximizes H. Such an $f(x)$ is called the 'ME distribution' in the literature, and consists of a combination of T pieces joined together. Figure 1 gives an illustrative graph later. First, we sort the data in an increasing order of magnitude, denote the order statistics by $x_{(t)}$ and compute new 'intermediate points' defined as the averages of successive order statistics:

(2) $z_t = 0.5\,(x_{(t)} + x_{(t+1)}) \quad t = 1, \ldots. T-1$

The intermediate points help define the ME intervals from pairs of successive z's. For t=2 to t=T−1 we define the intervals as: $I_2=(z_1, z_2), \ldots, I_{T-1}=(z_{T-2}, z_{T-1})$. Since we are considering a continuous random variable, both open (.,.) and closed [.,.] intermediate intervals are equivalent. The key novelty here is to also include the open-ended intervals I_1 $(-\infty, z_1)$ for t=1, and I_T (z_{T-1}, ∞) for $t=T$. Thus, we have exactly T intervals, each of which contains exactly one $x_{(t)}$.

The **mass-preserving** constraint says that, on an average, a fraction 1/T of the mass of the probability distribution must lie in each interval. The ME resampling algorithm achieves this by giving an equal chance to each interval I_1 $(-\infty, z_1)$, $I_2(z_1, z_2), \ldots, I_T$ (z_{T-1}, ∞) of being included in the resample. The traditional bootstrap selects uniform pseudorandom numbers between [0,1], transforms them into a sequence of random integers in [1,T] to define the shuffled selection from the set $\{x_t\}$. The bootstrap gives each observation an equal chance (1/T) of being included in the resample. The advantages of the ME resampling algorithm are: (i) it uses the interval that contains the observation and not necessarily the observation itself, and (ii) it uses the pseudorandom numbers, not transformed integer values.

The **mean-preserving** constraint (on order 1 moments of $f(x)$) is $\Sigma x_t = \Sigma x_{(t)} = \Sigma m_t$, where m_t denote the mean of $f(x)$ within the interval I_t. This property is satisfied by the traditional bootstrap, since its $f(x)$ is a delta function with the mass (1/T) concentrated at x_t in I_t. The mean preserving requirement is a bit complicated for the ME resampling algorithm, since it requires that the mean m_t in the interval I_t is equal to a weighted sum of the order statistic $x_{(t)}$ with weights from the set {0.25, 0.50, 0.75} explained below.

The $f(x)$ which maximizes our 'ignorance' defined by the entropy (1) alone is called the ME distribution. The problem of finding such $f(x)$ was solved long ago in the statistical literature dealing with so-called characterization problems, Kagan et al (1973). A reference familiar to economists in a different context is Theil and Laitinen (1980). The solution states that:

(i) $f(x)=(1/\theta)\exp([x-z_1]/\theta)$ with $\theta = m_1 ==0.75x_{(1)}+0.25x_{(2)}$ in I_1 the first interval.

(ii) $f(x)=(1/\theta)\exp([z_{T-1}-x]/\theta)$ with $\theta = m_T=0.25x_{(T-1)}+0.75x_{(T)}$ in I_T the last interval.

(iii) $f(x)=1/(z_k- z_{k-1})$ is a uniform density where the mean of the uniform $(z_{k-1}+ z_k)/2$ also equals $m_k=0.25x_{(k-1)}+ 0.50x_{(k)}+0.25x_{(k+1)}$ in the intermediate intervals I_k (z_{k-1}, z_k) for k=2, …, T−1. The following example illustrates how these weights from the set {0.25, 0.50, 0.75} help satisfy the mean-preserving constraint $\Sigma x_t = \Sigma m_t$. Thus the ME distribution avoids arbitrary choices and is completely known from the data.

For further clarity I use a simple example where there are T=5 data points (36, 20, 12, 8, 4). Hence the order statistics are $x_{(t)}$ =(4, 8, 12, 20, 36). From consecutive averages of two order statistics we compute z_t ={6, 10, 16, 28}. At the left extreme, the support of the ME distribution in the left tail is the interval I_1 (−∞, 6) and the density $[(1/\theta)\exp([x-6]/\theta)]$ should have the parameter θ=0.75*4+0.25*8=5. The next interval I_2 is (6, 10) and the ME distribution in it is uniform with the sample mean 0.25*4+ 0.5*8+ 0.25*12 =8. Similarly the mean of the uniform over the interval (10, 16) is 13 and over (16, 28) it is 22. Over the last open ended interval I_T (28, ∞), the ME distribution is again $(1/\theta)\exp([28-x]/\theta)$ with θ=0.25*20+0.75*36=32.

Now verify that the mean preserving constraint, $\Sigma x_t = \Sigma m_k$= 5+8+13+22+32=80, based on the judicious use of the weights 0.75, 0.50 and 0.25, is indeed satisfied. The elements x_{jt} of the j-th resample will not, of course, add up to 80 in each realization. Thus the ME distribution is continuous, artificially joined at apparent points of discontinuity. It consists of T attached parts over the entire Real line (−∞, ∞) and obviously avoids properties P1 and P2 of the traditional bootstrap. We shall see that it requires double sorting to avoid the property P3.

We need not explicitly impose any variance-preserving constraint, because it can be shown, Theil and Laitinen (1980), that the variance of the ME distribution is

(3) $$\text{Var}_{ME}=\text{sample variance} - \frac{1}{4T}\sum_{i=1}^{T-1}(x_{(i+1)}-x_{(i)})^2 - \frac{1}{24T}\sum_{i=2}^{T-1}(x_{(i+1)}-x_{(i-1)})^2 .$$

Since the ME distribution *reduces* the sample variance, there is no danger that $\text{Var}_{ME} \to \infty$. The amount by which Var_{ME} is reduced in (3) is negligible when T is large. Hence, it does not seem necessary to impose a variance preserving constraint. If one wants to preserve the sample variance, I suggest multiplying all members of the resample by a constant larger than unity based on (3).

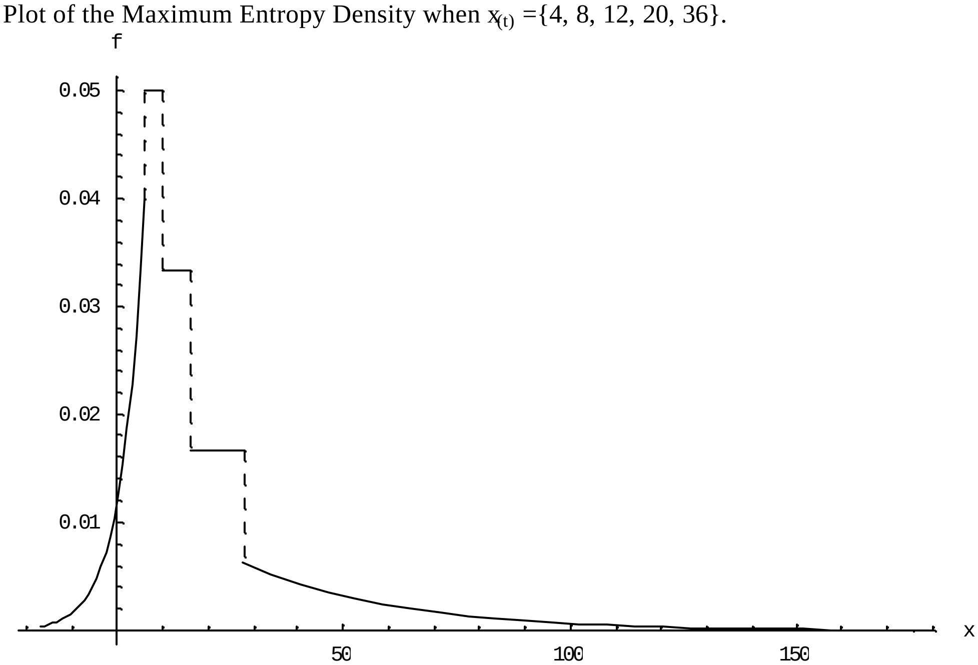
It is fortunate that the ME density involves only the exponential distribution in the two end intervals. We have simple analytical expressions for its CDF as $[1-\exp(-\theta x)]$ and its 'inverse CDF' as $[-|\theta| \ln(1-p_r)]$, where p_r denotes the probability. Since the logarithm of a fraction is always negative, $\ln(1-p_r) <0$, but the mean θ in the tails need not be always positive. Hence we use |θ| instead of θ in the expression to obtain a positive result. Thus pseudorandom numbers from the ME density are easy to obtain on a computer, despite the presence of the exponential. Each member of the constructed ensemble is a time series having T elements and we plan to create a large number j=1, 2, …J of such time series x_{jt} from the inverse CDF of ME-density as quantiles. Since the extrapolation in the left and right tails needs some care, we explain them first and also mention the algorithm for outlier rejection in separate subsections.

2.1 Tail Extrapolations:

First consider the right tail of the ME density similar to the one plotted in Figure 1, which begins at z_{T-1} and goes to infinity. The interval $[0, \infty)$ of the usual (standard) exponential with mean $\theta = m_T = 0.25x_{(T-1)}+0.75x_{(T)}$ is mapped on the interval $[z_{T-1}, \infty)$ needed here and assigned the probability mass (1/T). Hence we shift the starting point of the exponential to z_{T-1}. Consider T uniform pseudo-random draws denoted by p_s lying in the [0,1] interval. They yield j-th ensemble elements x_{jt} for t=1,..,T as quantiles x_{jt} of the ME density. If the random draw exceeds (T−1)/T, x_{jt} will belong to the right tail and is given by:

(4) $z_{T-1} - |\theta| \ln(1-p_r)$.

Figure 1. Plot of the Maximum Entropy Density when $x_{(t)}$ ={4, 8, 12, 20, 36}.



The left tail extrapolation is similar to that for the right tail, except that the exponential is in the reverse direction, so that $(1-p_r)$ in (4) will become (p_r). The left tail begins at $-\infty$ and goes to z_1. Let us first compute the quantile as if it were in the right direction and then adjust for the direction. The quantile of the ME density will belong to the left tail only if the random draw is less than 1/T. Since the mean of the exponential in the left tail is $\theta = m_1 = 0.75x_{(1)}+0.25x_{(2)}$, which may well be negative. The desired quantile similar to (4) is:

(5) $z_1 - |\theta|\,|\ln(p_r)|$.

2.2 Outlier Identification and Rejection of Quantile sets:

The ME density for intermediate intervals is uniform and the corresponding quantiles require simple interpolations. Collecting all T quantiles $\{x_{jt}\}$ and averaging them we check for outliers as follows. If $|(1/T)\Sigma_t\, x_{jt} - \bar{x}\,| > T_{ol}$, where T_{ol} is a tolerance number (defined by the user), the entire set $\{x_{jt}\}$ is considered an outlier. In some applications the smallest plausible value for the series cannot be negative (e.g., price) and this may require rejection of negative quantiles. In general, statisticians have developed tools defining the outliers. For example, elementary textbooks describe the low end outlier as any value less than Q_1–1.5*IQR, where Q_1 is the first quartile, Q_3 is third quartile and IQR=Q_3–Q_1= the inter-quartile range. Similarly upper end outlier is defined from Q_3+1.5*IQR. Although the user is assumed to be able to choose the appropriate outlier rejection scheme and specify T_{ol}, I am not asking from the user a detailed Bayesian specification of a prior distribution. I do not claim that my algorithm is the best.

Gilboa and Schmeidler (2003) have developed a rigorous axiomatic framework for inductive inference using fuzzy information, which does not fit the Bayesian paradigm. What they call 'memory' consists of a database from past cases and our original x_t data can be regarded as memory. Gilboa and Schmeidler imagine a person who uses this memory to "predict" some eventuality and create a qualitative plausibility ranking without having to envision all possible memories. In our case the 'eventuality' is the mean of j-th sample. Using $|(1/T)\Sigma_t\, x_{jt} - \bar{x}\,|$, my algorithm generates plausibility rank of the j-th sample leading to a rejection of implausible ensembles. Thus a rigorous axiomatic framework is available.

2.3 List of Steps in the Seven-Step Algorithm:

The entire set of steps needed to create a random realization of x_t is as follows.

[1] Define a T×2 sorting matrix called S_1 and place the observed time series x_t in the first column and the set of integers a={1, 2, …, T} in the second column of S_1.

[2] Sort the matrix S_1 with respect to the numbers in its first column. This sort yields the order statistics $x_{(t)}$ in the first column and a vector a^s of sorted a in the second column. The vector as was called τ earlier and will be needed later. From $x_{(t)}$ construct the intervals I_t defined from z_t, from (2), and m_t with certain weights on the order statistics $x_{(t)}$ defined above.

[3] Choose a seed, create T uniform pseudorandom numbers p_s in the [0,1] interval, and identify the range $R_t=(t/T, (t+1)/T]$ for t=0 to T–1 wherein each p_s falls.

[4] Match each R_t with I_t. Use equations (5) and (4) if $p_s \in R_0$ or R_{T-1}, respectively. Otherwise, use linear interpolation and obtain a set of T values $\{x_{jt}\}$ as the j-th resample. Recall that each mean of the uniform equals the correct mean m_t. If $|(1/T)\Sigma_t\, x_{jt} - \bar{x}\,| > T_{ol}$, where T_{ol} is a tolerance number (defined by the user), the entire set $\{x_{jt}\}$ is an outlier. Some users may reject outlier sets based on: Q_1–1.5*IQR or Q_3+1.5*IQR, where Q_1 is the first quartile, Q_3 is third quartile and IQR=Q_3–Q_1. If $\{x_{jt}\}$ is judged to be an outlier, go back to step 3 and select a new $\{x_{jt}\}$.

[5] Define another T×2 sorting matrix S_2. Reorder the T members of the set $\{x_{jt}\}$ for the j-th resample obtained in step 4 in an increasing order of magnitude and place them in column 1. Also place the sorted set a^s of step 2 in column 2 of S_2.

[6] Sort the S_2 matrix with respect to the second column to restore the order {1, 2, .., T} there. Denote the jointly sorted column 1 of the elements by $\{x_{sjt}\}$. These represent the ME resample, where the additional subscript s reminds us of the sorting step, which has restored the time dimension to correspond with the original data.

[7] Repeat steps 1 to 6 a large number of times for j=1, 2, …, J (J=999, say).

2.4 Doob's Theorem and constructive approximation to Ω, the complete ensemble.

This subsection shows why the seven-step algorithm gives us the ensemble Ω. Its validity is based on the assumption of 'perfect rank matching' between the observed data and any ω_j time series we construct. In effect, I am assuming that the rank correlation coefficient between the observed time series and any element ω of the constructed ensemble Ω should be unity. This is the implication of the steps 2 and 6 of Section 2.3 and is claimed to be less demanding than commonly used stationarity assumption. I will have created a completely specified Ω only when I identify a suitable joint probability distribution in order to satisfy the conditions of Doob's (1953) theorem. Following Priestley (1981, p.104) Doob's theorem is:

Doob's Theorem: For any positive integer T, let $\{t_1, t_2, .., t_T\}$ be any admissible set of values of T. Then under general conditions, the probabilistic structure of the random process $\{X(t)\}$ is completely specified, if we are given the joint probability distribution of $\{X(t_1), X(t_2), .., X(t_T)\}$ for all values of T and for all choices of $\{t_1, t_2, .., t_T\}$.

Since Doob's theorem is designed to facilitate an application to stationary processes, its statement focuses on *all possible* choices of $\{t_1, t_2, .., t_T\}$ subject to time shifts. In my design, {1,2,. ., T} is the only admissible choice for my dependent non-stationary time series, that is, no time shifts are admissible. Let $X_{(t)}$ denote the random variable associated with the order statistic at (t) and let its marginal density $f(X_{(t)})$ be the ME density. Recall that the support for the ME densities is $I_1(-\infty, z_1)$, $I_2(z_1, z_2)$, …, $I_T(z_{T-1}, \infty)$ based on the order statistics $x_{(t)}$. For example, let T=5 and the time spent in hours by a law firm on a case during weeks t=1,2,..,5 is observed to be: x_t=(4, 12, 36, 20, 8). Figure 1 plots the ME density for this example. The number of ME densities $f(X_{(t)})$ are exactly T, just enough for the T data points in x_t. Now I use the 'perfect rank matching' assumption to identify the $f(X_t)$ from each known $f(X_{(t)})$. Thus the joint probability distribution describing the Ω is given by the product $\Pi f(X_t)$ defined over t=1 to t=T. In practice, one chooses random draws and relies on the probability integral transformation to create each element ω of Ω. In constructing my Ω algorithm I do not require that exactly one observation come from each $f(X_t)$, although such a requirement can be imposed. Since I have satisfied the conditions of Doob's theorem, I claim that the seven-step algorithm provides a good approximation to the ensemble Ω.

3. Numerical Examples

My first example uses the illustrative data with T=5 plotted in Figure 1 with law firm hours now in the descending order: x_t=(36, 20, 12, 8, 4). I create a very large set Ω of elements ω_j for j=1,2,…,9999, comprising the ensemble. The first seven elements ω_j are reported in Table 1. These numbers are seen to have a great variety and are not too close to the original data and yet not too far. Hence they represent a plausible Ω. Since T=5, no asymptotic theory can apply, and yet it is remarkable that we are able construct such a large ensemble. The largest number 56.561 among the 9999 is larger than $x_{(T)}$=36 of the data. Due to the smallness of T (=5) the smallest number in the ensemble, 4.8772, is not smaller than $x_{(1)}$=4. This property does not hold for the example of the next section where T>223.

Table 1: Original x_t in the first column and the ensemble elements $\omega_j \in \Omega$, for j=1, .., 7.

x_t	j=1	j=2	j=3	j=4	j=5	j=6	j=7
36	54.375	16.352	43.8	16.478	29.049	39.986	10.225
20	10.225	6.1481	16.405	16.441	16.295	39.736	10.143
12	10.102	6.1209	16.394	6.024	16.128	37.785	10.112
8	6.1287	6.1008	16.107	5.6902	5.6549	36.54	6.1219
4	6.0765	6.0227	6.0415	5.0358	5.6471	5.4338	6.0903

The smallest realization in 9999 resamples is 4.8772 and the largest is 56.561. The original average is 16 and the grand mean of all means over the 9999 resamples is 15.909.

The 9999 standard deviations have: min=0.025146, max=6.968, mean=12.223 and median=13.142.

Column 1 of Table 1 has a declining time series, whereas the order statistics used in ME calculations are increasing. Yet the algorithm correctly retains the declining shape for all J. As a further test, I started with the shape (4, 12, 36, 20, 8), which has a mode in the middle and created J series ω_j by using the double sort ('perfect rank matching'). The algorithm correctly retained the hump shape in the middle of the series for each j.

Remark 3 (shape retention): Economists familiar with the ordinal utility theory can appreciate why perfect rank matching (i.e., double sorting) retains the basic shape of x_t, the observed time series. We simply pretend that x_t is a bundle of T commodities where only the ordering matters and each ω_j obeys its partial ordering. The next sub-section shows that the basic shape is retained even for a much larger data set with T=223. Figure 2 plots x_t for the consumption data (solid line) and four ω_j of 999 (dotted and dashed lines), which are visually seen to retain the basic shape of the solid line. Figure 3 is a similar plot for the plausible reincarnations of GDP. Table 2 shows similar patterns for autocorrelation functions (acf) and partial autocorrelation functions (pacf) of consumption. I emphasize that differencing, de-trending, spectral decomposition, moving blocks, and many other operations destroy the basic shape. By contrast, I retain the basic shape without imposing explicit parametric conditions on the acf or pacf.

The mean preserving constraint holds for the example of Table 1, since the grand mean 15.909 of the 9999 resamples is very close to the original sample mean of 16. The average of 9999 standard deviations is 12.223, which is slightly smaller than the sample standard deviation of 12.649, as predicted by the formula (3). It appears from Table 1 and Figures 2 and 3 that my seven-step algorithm is doing what I have claimed.

The constructed large ensemble is potentially useful for time series inference. It is not my purpose here to suggest a new bootstrap *per se*. However, I have removed all three undesirable properties of the traditional bootstrap listed at the beginning of Section 2. The bootstrap literature is vast, and there are piecemeal attempts to remove one or more of P1 to P3, but not all three. For example, the so-called 'smooth bootstrap' deals with P2, whereas the 'moving blocks' bootstrap deals with P3 by assuming 'm-dependent' time series (See Berkowitz and Kilian, 2000, for references). A simulation comparison of my bootstrap will have to include both the smooth and moving block methods. Although interesting, such a simulation is beyond the scope of this paper.

I used GAUSS software version 4.1 on a Pentium III, 850Mhz processor, 'Windows Me' OS machine. Replacing initial seed=321 by seed+1 at each iteration, all bootstraps took less than six minutes.

3.1 Keynesian Consumption Function:

Now that we have successfully created an algorithm for creation of time series ensembles, our next step is to show its usefulness for inference purposes. For illustration, consider the simple Keynesian consumption function:

$$(6) \quad C_t = \alpha + \beta\, Y_t + \gamma\, C_{t-1} + \varepsilon_t,$$

where Y denotes income, C denotes consumption and time is denoted by a subscript. My data for Y is the gross domestic product (GDP) and for C it is personal consumption. Both data series come from the National Income accounts by the Bureau of Economic Analysis (BEA) in Washington, DC. My two time series of 223 data points start from the first quarter of 1929 and end in the third quarter of 2002, the latest available quarter when this empirical example was estimated.

First, I estimate a version of (6) when the lagged dependent variable is absent ($\gamma\equiv 0$). Using the seven-step algorithm of Section 2.3, I create ω_j for j=1,..,999 realizations of Y and C from the ME distribution. Figures 2 and 3 illustrate ω_j for C and Y, respectively, where j=1,..4 and solid lines represent the original data. Are the ω_j retaining the autocorrelation properties (acf, pacf) of the original series? To study this, autocorrelations ρ_k for lag orders k=1,6 for C are reported in the upper panel of Table 2. Partial autocorrelations are reported in the lower panel. Although the ρ_k in Table 2 are not identical for the consumption data and different ω_j, their essential patterns are very similar. Partial autocorrelations ϕ_{kk} are, by definition, regression coefficients of the last term in AR(k), the autoregression of order k. In Table 2 the ϕ_{kk} are close to zero after the first order. These calculations were repeated for the Y variable and for ω_j when j=101, 102, .., 105. I do not report them for brevity, since they continue to show appropriate pattern similarities. Thus, Table 2 supports my claim in Remark 3 that the basic pattern of dependence is captured by the artificial ensembles. In fact if one wants to assess the sampling properties of sample estimates of ρ_k or ϕ_{kk} from x_t, I suggest computing their 999 realizations for any $k<T-1$. Similarly, if x_t~I(d) it is possible to get 999 ensemble estiamtes of d even if it is fractional for new insights on inference for the *d* parameter. In short, I can provide new confidence intervals for ρ_k, ϕ_{kk} or *d*.

Table 2: Autocorrelations and partial autocorrelations of order k for C

ρ_k	k=1	k=2	k=3	k=4	k=5	k=6
x_t	0.984	0.969	0.953	0.938	0.922	0.907
ω_1	0.986	0.971	0.957	0.943	0.929	0.914
ω_2	0.984	0.969	0.953	0.938	0.923	0.908
ω_3	0.985	0.97	0.955	0.94	0.925	0.911
ω_4	0.982	0.965	0.949	0.933	0.916	0.9
ϕ_{kk}						
x_t	0.984	-0.00479	-0.0134	-0.00869	0.00254	-0.0131
ω_1	0.986	-0.0163	0.0136	-0.00606	-0.00785	-0.0101
ω_2	0.984	-0.00444	-0.00687	0.00904	-0.00746	-0.00609
ω_3	0.985	-0.00054	-0.00924	0.00343	-0.0157	-0.00278
ω_4	0.982	0.00547	0.0173	-0.0139	-7.20E-05	-0.00678

The respective ordinary least squares (OLS) estimates of α and β are: a= −0.14962 and b=0.69245. Their t-statistics are: −18.96 and 447.06, respectively and the R^2 adjusted for degrees of freedom is 0.999, suggesting an excellent fit. The usual I(1) asymptotic arguments should hold, since a relatively large number (=223) of observations is available. On the basis of such arguments, the econometric literature suggests that:

(i) C and Y are integrated of order 1, or I(1),

(ii) the regression is spurious leading to underestimated standard errors,

(iii) the OLS is T-consistent rather than $\sqrt{T}$ consistent, and

(iv) equation (6) should be supplemented by the identity Y=C+ι , where ι denotes investment, creating a system of two simultaneous equations. Otherwise, the direct OLS estimates of (6) are inconsistent.

Clearly, not all four claims can be valid at the same time. The ensemble method often avoids situations rendering the OLS inconsistent or spurious. Let us exclude the OLS regression reported above from the observed 223 data points and focus on the 999 realizations ω_j from the ensemble Ω created by the seven-step algorithm. By construction, my J=999 realizations are not too apart from the observed data. Indeed, the observed data x_t may be reasonably regarded as one sample from Ω. The mean preserving constraint ensures that the ergodic theorem holds. Next, I run J separate regressions to estimate α and β coefficients J times by the OLS method. Legendre introduced least squares in 1805 as a mathematical approximation, without assuming any parametric distribution for the error term ε at all. Hence the J estimates of the parameters α and β from the ensemble are semiparametric and remain meaningful irrespective of the distributional properties of error terms.

Heteroscedasticity and autocorrelation consistent (HAC) estimation of standard errors of regression coefficients is not attempted, since we have explicitly constructed a large ensemble for direct inference. Of course, with J time series ω_j we can construct e_j vectors, each having a $T\times 1$ vector of regression residuals. Now the outer product $V_j = e_j e'_j$, is a rank one estimate of V, the $T\times T$ error covariance matrix. Following Vinod (2000, p. 380) a suitable linear combination of V_j can yield a suitable symmetric, nonsingular estimate of V. Next, use $b_V = (Z'V^{-1}Z)^{-1} Z'V^{-1}y$, where Z is a matrix of regressors, as a new feasible GLS estimator for efficient estimation. A comparison of b_V with other methods discussed in the large HAC estimation literature is left outside the scope of this paper.

Now let us turn to inference based on confidence intervals. We have j=1,…,J ensemble estimates of α and β denoted by a_j and b_j respectively. I simply rank order the J estimates a_j in an increasing order of magnitude and denote the order statistics by $a_{(j)}$. Similarly I can construct $b_{(j)}$. A simple first approximation to the 95% confidence interval (CI95) for α is $[a_{(25)}, a_{(975)}]$ or [−0.13027, −0.18222]. The CI95 for β is: $[b_{(25)}, b_{(975)}]$ or [0.68781, 0.69846]. Since consumption and income are closely linked, it is not surprising that these naïve confidence intervals are narrow (See Remark 2 above).

Another important issue for inference arises when the lagged dependent variable is included in a regression, where the OLS coefficients are inconsistent, Fuller (1976, sec. 9.8). The inconsistency of OLS arises because C_{t-1} might be correlated with ε_t. However, Durbin (1960) proved the optimality of the OLS estimator despite the presence of lagged dependent variable. See also Vinod (2000). We have a lagged dependent variable in (6) if we estimate γ instead of setting $\gamma \equiv 0$. Table 3 reports the results for this case. When we compute ensemble confidence limits, we have a direct approximation to the quantiles of the sampling distribution of any pivotal statistic. Thus the ensemble confidence limits reported in the last two columns of Table 3 should suffice for a valid inference, despite the presence of a lagged dependent variable C_{t-1}. A note to Table 3 reports that for consumption data, $\bar{x}$ =3.0088 and μ'=2.9952. These are obviously very close to each other and a similar closeness holds for the GDP data between sample mean and the finite ensemble mean.

I conclude this example with the claim that my constructive ensemble offers potentially simple solutions to hard inference problems where the asymptotic normality is known to fail, or where we are not sure about achieving stationarity. In this subsection my algorithm was applied separately to C_t, Y_t, and C_{t-1}. When there are several regressors and it is cumbersome to construct ω_j time series for each, we can construct the ω_j time series just from regression residuals, provided the OLS specification is valid.

Table 3. Estimation of $C_t = \alpha + \beta\, Y_t + \gamma\, C_{t-1} + \varepsilon_t$

	OLS Coefficient	t-statistic	Ensemble lower limit	Ensemble upper limit
Intercept	–0.014942	–3.077931	–0.050671	–0.00783
Income	0.073495	4.252101	0.046547	0.24766
Lagged Consumption	0.90115	35.84485	0.64645	0.93961

Note to Table 3: Ensemble mean μ' for C_t is 2.9952 (for Y_t it is 4.5426), original sample mean $\bar{x}$ for C_t is 3.0088 (for Y_t it is 4.5617). Clearly μ' are very close to $\bar{x}$ in both cases.

3.2 Application in Finance

This subsection shows that new methods are potentially useful in Finance. Wall Street investors and bankers often want a dollar figure on the potential loss in a worst-case scenario. Value at risk (VaR) is one such measure which can be obtained from a low (e.g., 1%) quantile of a parametric or nonparametric f(X). We use the ensemble to construct a 95% confidence interval around VaR. We use for illustration a mutual fund named Alliance All-Asia Investment Advisors Fund, with the ticker symbol AAAYX, for the period of T=132 months from January 1987 to December 1997 from Morningstar (2000). Descriptive statistics reveal that the density f(X) is a negatively skewed.

Table 4: Autocorrelations ρ_k and partial autocorrelations ϕ_{kk} of order k for original data x_t and for the first four elements of the ensemble Ω.

ρ_k	k=1	k=2	k=3	k=4	k=5	k=6
x_t	-0.00451	-0.0433	-0.151	-0.0869	0.0684	-0.0142
ω_1	-0.051	0.0264	-0.0584	-0.0117	0.0166	-0.015
ω_2	-0.0389	-0.00836	-0.175	-0.0872	0.0388	-0.0454
ω_3	-0.062	0.031	-0.0578	-0.0161	0.0232	-0.0178
ω_4	-0.0467	0.0195	-0.0454	-0.0108	0.017	-0.0103
ϕ_{kk}						
x_t	-0.00451	-0.0434	-0.151	-0.0935	0.0537	-0.0441
ω_1	-0.051	0.0239	-0.056	-0.0181	0.018	-0.016
ω_2	-0.0389	-0.00988	-0.176	-0.105	0.0258	-0.0794
ω_3	-0.062	0.0272	-0.0545	-0.0238	0.0241	-0.0172
ω_4	-0.0467	0.0173	-0.0439	-0.0152	0.0174	-0.0103

Note that due to limited liability laws, investors can never lose more than all 100% of their invested capital. Hence, the worst outcome occurs when the risk-free return is at its maximum or max(Tb3) whereas the mutual fund is liquidated (has a 100% loss). For our application we choose T_{ol}=3 and also reject the sets $\{x_{jt}\}$ in Step [4] if any element is outside the range (–110, 100). For the chosen J=999 members of the ensemble, the grand mean of 1.0194, is not too far from the time average 0.7387, and hence it satisfies the ergodic theorem. The ensemble minimum –109.97 is smaller than the sample minimum –21.147, suggesting that we are getting a sufficient range of realizations on the down side, despite the outlier rejections. Similarly, the maximum 99.944 over J values exceeds the sample maximum of 10.724 confirming good variety of realizations on the up side as well. The quantiles 0.01 for the original data is −9.629. For a $100,000 capital, the potential loss VaR= $9,629 is such that we expect that the probability that the actual loss to be *less* than $9,629 is 0.99. Table 4 is similar to table 2 above and makes the same point for these data.

Vinod and Morey (2002) note that financial economists usually ignore the estimation risk. This is also true of the commonly calculated VaR, which focuses only on market investment risk, not VaR estimation risk. Given adequate computer resources, we can extend the algorithm to joint portfolios of mutually dependent but varied classes of assets. By contrast, traditional VaR based on asymptotic theory has great difficulty in handling joint dependencies and complicated return computations involving taxes and state-contingent gains and expenses. We claim to provide a direct and simple method of computing the VaR, which incorporates estimation risk in addition to market risk and should be useful in practical Finance. A 95% confidence interval for VaR is: [−15.388, −5.1515]. Thus, when the ensemble is taken into account, the VaR = $15,388, instead of the VaR=$9,629 from a single realization x_t in the data. This is claimed to be a new way to allow for investment risk, as well as, estimation risk.

4. Final Remarks

The classical time series methods by WKK and others imagine that the observed time series is one realization ω_j from a large ensemble Ω. No one has used collections of white noise and other processes to construct an approximation to the Ω. This paper takes the first step in that direction. For valid inference under the WKK methods it is necessary to first make sure that each available series is converted into a stationary series by differencing or other shape-destroying transformations. Since the WKK methods were developed well before computers and even before Shannon's concept of information (or entropy as its expectation) were developed, I offer a fresh look using computer intensive tools.

Remark 1 shows how the ME density satisfies the ergodic theorem by construction. The joint probability distribution is explicitly given by the product $\Pi f(X_t)$ derived from the ME density, and hence the conditions for Doob's theorem are satisfied. I provide a seven-step algorithm to construct J time series of ω_j as elements of Ω. The use of the ME density and double sorting are both new and crucial for my algorithm. Gilboa and Schmeidler's (2003) axiomatic framework is applicable to my construction.

I show with three numerical examples that the maximum entropy density can be explicitly computed for any time series. I explicitly verified the ergodic theorem for all examples. A simple example of T=5 points was shown to have 9999 mostly distinct resamples. Table 1 exhibits ω_j for j=1, ..,7 and shows that their range is reasonable and they may be viewed as members of a plausible ensemble.

In general, I claim that computer-generated ω_j yield plausible reincarnations of any economic variable. There is a sense in which the true object of statistical inference is to assess what the estimated parameter values might have been under such reincarnations. Hence my algorithm offers a new tool for inference. This is useful for nonstationary series, since the algorithm retains the dependence properties (acf and pacf) of the original time series without imposing parametric constraints (see Tables 2 & 4).

For the consumption function example, I illustrate in Figures 2 and 3 the original data quarterly consumption and GDP, respectively, by solid lines and the ensemble elements ω_j for j=1, 2, .., 4. These figures visually show the unique achievement of my algorithm in retaining the basic shape of the time series without imposing parametric restrictions on the acf or pacf. Traditional differencing and other methods do not retain the shape at all. This example also shows that the computer intensive algorithm readily solves the inference problem of nonnormality and inconsistency when the regressors contain a lagged dependent variable. Similarly the Finance example shows that my algorithm can readily incorporate estimation risk in addition to investment risk in the popular worst-case scenario calculation VaR. It can be extended to GARCH-based VaR calculations, McCullough and Vinod (2003).

Another appealing feature for short economic time series is that it is not necessary to test stationarity, nor is it necessary to use differencing to initially transform the time series into a stationary I(0) series before regular estimation and testing can begin. We can work directly with any I(d) series, even if d is fractional and unknown. The new ME bootstrap for dependent data is also applicable elsewhere. The constructive ensemble can potentially simplify several inference problems in time series analysis. Among possible extensions I have noted new confidence intervals for ρ_k, ϕ_{kk} or *d* and a new feasible GLS estimator to avoid both autocorrelation and heteroscedasticity. Since I offer a fresh look at the elegant theory established over a century, I hope that my methods are viewed as a feeble first attempt and receives further critical study and extensions.

Acknowledgment.
I thank Colin Rose for help with Figure 1 and Bruce McCullough for detailed comments and helpful suggestions.

References

Berkowitz, J. and L. Kilian. 2000. Recent developments in bootstrapping time series. econometric Reviews 19(1), 1-48.

Birkhoff, G.D. 1931. Proof of the ergodic theorem. Proceedings of the National Academy of Sciences 17, 656-660.

Dhrymes, P. J. 1998. Time Series, Unit roots and Cointegration. New York: Academic Press.

Doob, J. L. 1953. Stochastic processes. New York: J. Wiley

Durbin, J., 1960, Estimation of parameters in time-series regression models. Journal of the Royal Statistical Society, Ser. B, 22, 139-153.

Fuller, W. 1976. Introduction to statistical time series analysis. New York, J. Wiley
Gilboa, Itzak and D. Schmeidler (2003) Inductive inference : An axiomatic approach. Econometrica 71 (1), 1-26

Kagan, A. M., Yu.V. Linnik, and C. R. Rao., 1973. Characterization problems in mathematical statistics. J. Wiley and Sons, New York, USA.

Khinchin, A. I. 1934. Korrelationstheorie derr stationaren stochstischen prozesse. AMath Ann. 109, 604-615.

Kolmogorov, A. N. 1931 Über die analytischen methoden in der Wahrscheinlichkeitsrechnung. Math Ann. 104, 415-458.

Maddala, G. S. and In-Moo Kim. 1998. Unit roots, Conitegration and Structural Change. New York, Cambridge University Press.

McCullough, B. D. and H. D. Vinod. 2003. "Comments: Econometrics and Software" *Journal of Economic Perspectives*, 17 (1) (winter 2003), 223-224.

Morningstar, 2000. On-Disk, Principia, Principia-Plus, and Principia-Pro manuals 1991-2000, Morningstar Publications, Chicago, IL.

Parzen, E., 1962. Stochastic Processes, Holden Day, San Francisco.

Sen, Amit, 2003. On unit root tests when the alternative is trend-break stationary process. Journal of Business and Economic Statistics 21, 174-184.

Spanos, A. 1999. Probability theory and statistical inference: econometric modeling with observational data. New York, Cambridge University Press.

Theil, H. and Laitinen, K., 1980. Singular moment matrices in applied econometrics. in P. R. Krishnaiah, (Ed.) Multivariate Analysis – V, North-Holland Publishing Co., New York, USA, 629-649.

Vinod, H. D., 1985. Measurement of economic distance between blacks and whites. Journal of Business and Economic Statistics 3, 78-88

Vinod, H. D., 1994. Economic equilibria, fractional cointegration and multivariate AR-FIMA models. 1993 Proceedings of the Business and Economics Section, Washington, DC: American Statistical Association, 302-307.

Vinod, H. D., 2000. Foundations of multivariate inference using modern computers. Linear Algebra and Its Applications, 321, 365-385.

Vinod, H. D., 2002. Ranking mutual funds using unconventional utility theory and stochastic dominance, Journal of Empirical Finance (submitted)

Vinod, H. D., 2002b. A looser cointegration concept using fractional integration parameters and quantification of market responsiveness," Journal of Statistical Planning and Inference. 100, 399-410.

Vinod, H. D. and M. Morey, 2002. Estimation risk in Morningstar fund ratings. Journal of Investing 11 (4), 67-75.

Wiener, N. 1930. Generalized harmonic analysis. Acta Math. 55, 117-258.

Yaglom, A. M. 1962. An introduction to the theory of staionary random functions. (translated by Richard A.Silverman) New York, Dover Publications.

Figure 2: Observed Consumption, C

and Four Ensemble Elements

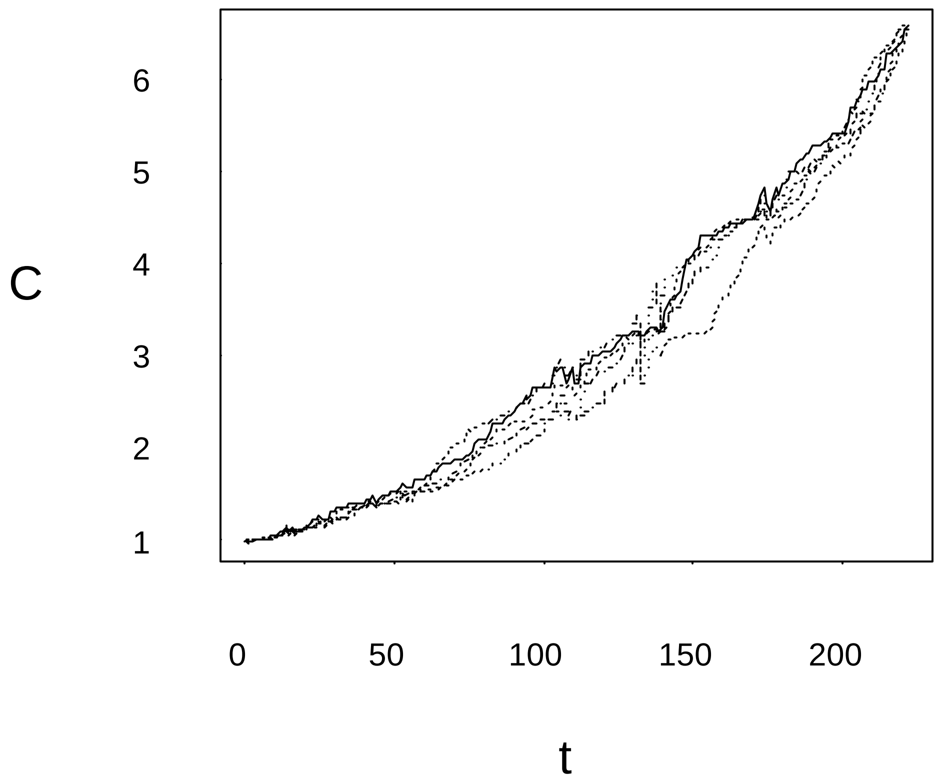


Figure 3: Observed GDP, Y (solid line) and Four Ensemble Elements