Monte Carlo Experiments on Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference

Allison Jennifer Ames

Virginia Polytechnic Institute and State University
Master of Science
Agricultural and Applied Economics

Dr. Tia Hilmer
Dr. Aris Spanos
Dr. Daniel Taylor

Keywords: Maximum Entropy, Bootstrap, Time Series, Monte Carlo Simulations

Monte Carlo Experiments on Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference

Allison Jennifer Ames

ABSTRACT

In econometric analysis, the traditional bootstrap and related methods often require the assumption of stationarity. This assumption says that the distribution function of the process remains unchanged when shifted in time by an arbitrary value, imposing perfect time-homogeneity. In terms of the joint distribution, stationarity implies that the *date* of the first time index is not relevant. There are many problems with this assumption however for time series data. With time series, the *order* in which random realizations occur is crucial. This is why theorists work with stochastic processes, with two implicit arguments, w and t, where w represents the sample space and t represents the order. The question becomes, is there a bootstrap procedure that can preserve the ordering without assuming stationarity? The new method for maximum entropy ensembles proposed by Dr. H. D. Vinod might satisfy the Ergodic and Kolmogorov theorems, without assuming stationarity.

AKCNOWLEDGEMENTS

I would like to thank Dr. Tia Hilmer for all of her help, not only on my thesis, during my master's program at Virginia Tech. She was a wonderful role model and being her teaching assistant taught me so much.

Dr. Aris Spanos was both an important professor and mentor over the time I spent here. He taught me to read critically and reinforced the importance of statistics in economics. Without his help I would still be searching for an interesting, challenging topic.

Dr. Daniel Taylor provided the guidance that new graduate students could not do without. I am forever grateful for how he and his family made all of my class feel welcome. I would also like to thank Dr. Deborah Mayo, Dr. Anya McGuirck, Andreas Koutris, and Jason Bergtold for their contributions as well.

My parents and my family offered so much support during this time and I greatly appreciate it.

CONTENTS

CHAPTER I

Introduction

A statistical model *S* comprises a set of internally consistent probabilistic assumptions from three broad categories: Distribution, Dependence, and Heterogeneity. *S* also defines a stochastic generating mechanism giving rise to a stochastic process $\{Y_t , t \in T\}$ where Y_t can be a vector. If the statistical model *S* is correctly specified, then the data $y = (y_1, \dots , y_n)$ is a highly likely realization of the process $\{Y_t , t \in T\}$. Therefore, the data $y = (y_1, \dots , y_n)$ is truly typical of model *S*, if it exhibits chance regularity patterns which reflect the probabilistic structure of the prespecified model *S* (Spanos 2002).

Inference propositions concerning optimal estimators, tests, or predictions are derived assuming that the prespecified statistical model is true. Establishing the truth of the premises is of paramount importance. This allows the modeler to assess the soundness of the deductive inference propositions and the reliability of any deductive inference based on the observed data $y = (y_1, \dots , y_n)$. In cases where the statistical model is not true for the particular data, the associated inference cannot be valid and thus no learning from the data is possible.

We see then that inductive inference based on model *S* is reliable when data $y = (y_1, \dots , y_n)$ is truly typical of the process defined by model *S*. Using misspecification testing one can then assess whether data $y = (y_1, \dots , y_n)$ is truly typical of the process defined by model *S*. The misspecification testing will test the probabilistic assumptions of model *S*. The presence of misspecification errors suggests that the data $y = (y_1, \dots , y_n)$ is not truly typical of *S*, rendering the model unreliable. Following this argument, the model

should be respecified and then undergo another round of misspecification testing. This process continues until all the probabilistic assumptions are met.

Resampling techniques are important to model validation. This is because misspecification testing and reliability of inference is assessed relative to the particular sample realization represented by the *observed* data, not potential or future data.

Being able to produce accurate replicas of data $y = (y_1, \dots, y_n)$ will allow the modeler to: 1) Enhance misspecification errors and render their presence easier to detect, and 2) Assess the reliability of inductive inference. The successful application of resampling techniques in statistical inference depends on two important conditions:1) The assumptions of the prespecified statistical model are valid for data $y = (y_1, \dots, y_n)$, and 2) The Resampling method used will give rise to accurate replicas of the original data. Any departures from these conditions are likely to lead the resampling results down the wrong path.

Problem Statement

In econometric analysis, the traditional bootstrap and related methods often require the assumption of stationarity. This assumption says that the distribution function of the process remains unchanged when shifted in time by an arbitrary value, imposing perfect time-homogeneity. In terms of the joint distribution, stationarity implies that the *date* of the first time index is not relevant. There are many problems with this assumption however for time series data. With time series, the *order* in which random realizations occur is crucial. This is why theorists work with stochastic processes, with two implicit arguments, w and t, where w represents the sample space and t represents the order. The question becomes, is there a bootstrap procedure that can preserve the ordering without

assuming stationarity? The new method for maximum entropy ensembles proposed by Dr. H. D. Vinod might satisfy the Ergodic and Kolmogorov theorems, without assuming stationarity.

Objectives

The objective of this research is to ensure, or disprove, that Dr. Vinod's methods proposed in *Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference* (2004) satisfy the Ergodic and Kolmogorov theorems. The second objective is to ensure that Dr. Vinod can do this without assuming stationarity. These objectives will be completed using the five proposed experiments, described in the methodology section. These five experiments are Monte Carlo simulations of the Maximum Entropy bootstrap proposed by Dr. Vinod.

This paper will first introduce the relevant literature and background on the following areas: Time Series Analysis, The Stationarity Assumption, Bootstrapping, and Maximum Entropy. The third chapter states the hypotheses. Next will be a discussion of the methodology, the proposed Monte Carlo experiments. Chapter five is a discussion of the results and analysis of the experiments. Then a brief conclusion follows as a summary of the results and includes proposed future work in this area.

CHAPTER II

Conceptual Framework: Time Series Analysis

Time series analysis has been a long-debated topic in econometrics. Currently, two different approaches have been adopted by workers in time series: the stochastic approach and the functional approach. The stochastic approach has been the one generally adopted by probability theorists and statisticians. On occasion it may make sense to think of a particular r vector-valued time series X(t) as being a member of an ensemble of vector time series which are generated by some random scheme. We can denote such an ensemble {X(t,θ);θ€Θ and t=0,1,…} where θ represents a random variable taking values in Θ. If X(t,θ) is a random variable, we can go on to discuss its finite dimensional distributions. Once θ has been generated (in accordance with its probability distribution), the function X(t,θ), with θ fixed, will be described as a realization, or sample path, of the time series.

The Probabilistic Reduction perspective is complementary and can be viewed as choosing an appropriate statistical model which amounts to choosing an appropriate statistical model which accounts for all the systematic information in an observed data series. This is the data at hand, not future or past data. This systematic information comes to the modeler in the form of chance regularity patters exhibited by the time series data. An a priori assessment of t plots, scatter plots, and histograms. The success of modeling relies upon recognizing the chance regularity and then choosing the appropriate probabilistic concepts to capture this information. These concepts come in the form of reduction assumptions of the joint distribution. The probabilistic assumptions impose reductions of the joint distribution in the three categories: Distribution, Dependence, and

Heterogeneity. These models are not viewed as data generating processes, but rather reductions from the joint distribution of the process.

A given time series is regarded as being selected stochastically from an ensemble of possible series. We have a set Θ of r vector functions $\theta(t)$, and after defining a probability measure on Θ, we obtain a random function $X(t,\theta)$, whose samples are the given functions $\theta(t)$. Alternatively, given $X(t)$, we can set up an index $\theta=X(\cdot)$ and take Θ to be the set of all θ. We may then set $X(t,\theta)=X(t, X(\cdot))$. In any case, we find ourselves dealing with measure theory and probability spaces.

For a given observable process {yt, t€T}, the modeler begins by specifying the model exclusively in terms of this process. In particular, it views the process as an orthogonal decomposition of the form: $y_t=E(y_t|\sigma(Y_{t-1}^o)) + u_t$, where $Y_{t-1}^o = (y_{t-1}, y_{t-2},\ldots y_0)$. In order to yield an autoregressive model of order one (AR(1)), the reduction assumption imposed on the joint distribution are Normal, Markov, and Stationary. This yields the model $yt=\alpha_0 + \alpha_1(y_{t-1}) + u_t$, t€T. The importance of these assumptions will be expanded upon further in this section. It is the stationary assumption that can provide difficulty in resampling techniques.

In the functional approach, a given r vector series is interpreted as a mathematical function and the basic ensemble takes the form $\{X(t,v)=X(t+v)|v=0,1\ldots\}$. This approach is taken by Wiener (1930) for example, and is called general harmonic analysis. The distinction between these approaches comes in the different mathematical tools required, and the different limiting processes involved (Brillinger).

Conceptual Framework: The Stationarity Assumption

As previously mentioned, one of the problems with the simple bootstrap, discussed below in detail, is the imposition of Stationarity. A traditional discussion of stationarity follows. Suppose $\{X_n\}$ is a sequence of random variables satisfying the equation $X_n=h(X_{n-1},\ldots,X_{n-p}; \beta) + e_n$, where $n>p$, β is a $q\times1$ vector of parameters, h: is a known measurable Borel function, and $\{e_n\}$ is a sequence of IID random variables with common distribution F that are independent of the random variables $X_1,\ldots,X_n$. Noting that the process $\{X_n\}$ is driven by the innovations e_i's that are IID, the IID bootstrap method can be extended to the dependent model. When the sequence is *stationary*, under suitable regularity conditions, a version of the IID-innovation bootstrap approximation (Lahiri 23-24) to the sampling of the standardized least square estimator is more accurate than the Normal approximation. For *nonstationary* cases, one must be considerably more careful when applying the IID-innovations bootstrap. Lahiri then goes on the confirm that the moving blocks bootstrap, nonoverlapping blocks bootstrap, generalized block bootstrap, subsampling, transformation-based, and the Sieve bootstrap all require stationarity assumptions (Lahiri).

The probabilistic reduction perspective can clarify some of the problems associated with traditional stationarity discussion. Of the three main probabilistic assumptions, stationarity is considered a homogeneity restriction. The most restrictive form of homogeneity for an independent process $\{Y_t, t€T\}$ is that of complete homogeneity. This is more commonly referred to as the ID, identical distributions, assumption. This independent stochastic process is said to be *strictly stationary* if the joint distribution remains unchanged if we shift each point (1, 2,…,T) by a constant τ. If

the stochastic process is strictly stationary, then the transformed sequence $(X_{t1},\ldots,X_{tn},\ldots\}$ (where $X_{tk}=g(Y_{tk})$, k=1,2,…, and where g(·) is a well-behaved Borel function) is also strictly stationary. That is, strict stationarity implies Identical Distribution for all of the marginal distributions. In this sense, it is possible to extend the Identical Distribution homogeneity assumption to the case where there is some dependence (Spanos 1999). Examples of strictly stationary series include a series of IID r vector-valued variates and a series that is a deterministic function.

An r vector-valued series X(t), t=0,… is called *second-order stationary* if:

$$c_a(t)=E(X_a(t))=c_a$$

$$c_{ab}(t+u,t)=\text{cov}\{X_a(t+u), X_b(t)\}=c_{ab}(u)$$

for t, u=0,… and a,b=1,…,r.

Note that a strictly stationary series with finite second-order moments is second-order stationary. In the link between the reduction and model assumptions, stationarity implies the model assumption of t-homogeneity.

Many examples of stationary time series are given as they are the basis of analysis up to this point. Some are as follows: First, a pure white noise series. Let e(t) be a sequence of IID r vector-valued random variables. Another is a linear process. Let e(t) be the r vector-valued pure noise series, as illustrated above. Let $X(t) = \sum a(t-u)e(u)$ where $\{a(u)\}$ is a summable filter. If only a finite number of the a(u) in an expression are nonzero, then the series X(t) is commonly called a moving average process. If a(0), $a(m)\neq 0$, the process is said to be of order m.

One of the more well-known stationary time series is the stationary Markov process. An r vector time series X(t) is said to be an r vector Markov process if the

conditional probability $P\{X(t)\leq X|X(s_1)=x_1,\ldots,X(s_n)=x_n, X(s)=x\}$ for any $(s_1<\ldots<s_n<s<t)$ is equal to the conditional probability $P\{X(t)\leq X|X(s)=x\}=P(s,x,t,X)$ for $s<t$. The function $P(s,x,t,X)$ is called the transition probability function. It is important to realize that Markovness is not a direct restriction on the memory, but rather a restriction on the conditional memory. That is, for predicting the future of the process, the present provides all of the relevant information. The Brownian motion process, with a history in physics, is one of the more well-known Markov processes.

Finally, we see the illustration of autoregressive schemes. The above leads us to consider r vector processes X_t that are generated by schemes of the form $X_t+ \alpha_1 X_{t-1}+\ldots+\alpha_m X_{t-m}=e_t$ where e_t is an r vector pure noise series and $\alpha_1 \ldots \alpha_m$ are $r\times r$ matrices. If the roots of the determinant $A_z=0$ lie outside the unit circle where $A_z=I+\alpha_1 z+\ldots A_m z^m$, it can be shown that the above has a stationary solution. Such an X_t is referred to as an r vector-valued autoregressive process of order m. Still others include Mixing Moving Average and Autoregressive Process and Functions of Stationary Series (Lahiri). However there are many problems with the idea of unit roots to begin with. The different implicit parameterizations must be considered.

Conceptual Framework: Bootstrap Techniques

For all data, not only time series, one attempt at an accurate replica of the data is the resampling technique known as the bootstrap, originally proposed by Efron in 1979. This method gives rise to accurate replicas of the original data, but only in the case where the data were realizations of an Independent and Identically Distributed (IID) process. Efron developed the bootstrap as an alternative to the jackknife because the latter did not perform well in finite samples. The goal of the bootstrap is to use the sample data to

mimic the overall population distribution and to use resampling to calculate an estimated sampling distribution.

For an initial sample data set of size n, this is accomplished by drawing a random sample of size n *with replacement* from the initial sample data set, calculating the parameter of interest for the sample drawn, and repeating the process many times. The bootstrap estimator is then the average across all of the parameter estimates from the different bootstrap samples. The intuition is that the distribution of the parameter estimates from the bootstrap samples mimics the traditional asymptotic sampling distribution of parameter estimates for samples drawn from the population distribution. Hence, the bootstrap technique allows the researcher to generate an estimated sampling distribution in cases in which he/she only has access to a single sample rather than the entire population (Hilmer).

There are many extensions of Efron's original bootstrap. Block resampling allows the modeler to resample using non-IID data, especially dependent data. Consider instead exchangeable blocks of consecutive observations rather than single data points. A resampling scheme based on long enough blocks, whose lengths increase with sample size, can preserve the original dependence in the resampled series. There are three main types of block resampling. The first is the moving blocks bootstrap with non-overlapping blocks. Another is the moving blocks bootstrap with overlapping blocks. This approach constructs pseudo-time series whose structure mimics the structure of the original data by resampling blocks of consecutive observations of length l and then joining them together. Finally, there is sub-sampling with random length blocks: views the resampled blocks as sub-series of the original data. The key difference between the two is that sub-sampling

recomputes the statistic on blocks of size $l < n$ while the moving blocks bootstrap recomputes the statistic on pseudo-time series.

These block resampling methods are far from perfect in preserving the probabilistic structure of the original data because of the distortions caused by breaks between blocks (Spanos 2002). This is where many developments in research have been pursued, including the paper by H. D. Vinod on the Maximum Entropy Constructive Ensemble (2004). However, these new resampling techniques must be examined in order to gauge their reliability.

Assessing how well a resampling technique behaves is based upon the faithful replication of observed data structure and reliability of inference. In order to assess how faithfully the resampling method replicates the probabilistic structure of the observed data we compare the resampling distribution of τ_n under H_o generated by a Monte Carlo simulation with that generated by the resampling method in question (where τ_n is the test statistic and H_o is the null hypothesis). The reliability of the hypothesis test is assessed by estimating:1) The density of the resampled test statistic, 2) The error rejection probability for H_o, and 3)The skewness and kurtosis coefficients. To test these conditions, five experiments will be conducted on: Normal IID data, heterogeneity in the mean, heterogeneity in the variance, dependence, and heteroskedasticity. These experiments will be expanded upon later in the methodology section.

Despite its appeal for simplicity, Efron's original bootstrap algorithm (1979) was designed for data which are, in population, independent and identically distributed (IID). This simplest, but most widely, used sampling model is the one based on the idea of a random experiment and is called a random sample. A set of random variables (X_1,

$X_2,\ldots,X_n$) is a random sample from f(x;θ) if the random variables $X_1, X_2,\ldots,X_n$ are IID. In this case, the distribution of the sample takes the form:

$$F(x1, x2, \ldots, xn;\theta) = \Pi_{i=1\ldots n} f(x_i;\theta) = [f(x_i;\theta)]^n$$

Where the first equality stems from independence and the second from identical distributions.

For the IID case, one can create artificial repeated samples by random resampling, *with replacement*, from the data. If the data are not IID, say they display some heteroskedasticity or serial correlation, a randomly resampled set of data will not preserve these properties. Therefore, statistics calculated from this resampled data will not be valid. Thus, the traditional bootstrap fails for time series data.

S.N. Lahiri provides an in-depth discussion and formal proofs for the failure of the traditional bootstrap for time series data (2003). The IID bootstrap method of Efron (1979), being very straightforward and general, has found application to a multitude of statistical problems. However, the general notion that is a catch-all is misleading. A prime example is that provided by Singh (1981), which in addition to demonstrating the superiority of the IID bootstrap, also brought its inadequacy for dependent data to light.

Suppose $(X_1, X_2,\ldots)$ is a sequence of m-dependent random variables with $E(X_1)=\mu$ and $E(X_1^2)<\infty$. Recall that $\{X_n\}$ is called m-dependent for some integer $m\geq 0$ if $\{X_1,\ldots,X_k\}$ and $\{X_{k+m+1},\ldots\}$ are independent for all $k\geq 1$. Also, let $\sigma_m^2=\text{var}(X_1)+2\sum\text{cov}(X_1,X_{1+i})$. If $\sigma_m^2 \in (0,\infty)$, then by the CLT for m-dependent variables, $\sqrt{n}(X_n\text{-bar} - \mu)$ converges in distribution to $N(0, \sigma_m^2)$. Now suppose that we want to estimate the sampling distribution of a random variable T_n using the IID bootstrap. For simplicity, assume that the resample size equals the sample size. Then the bootstrap

version of T_n is given by: $T_{n,n}^* = \sqrt{n}(X_n^*\text{-bar} - X_n\text{-bar})$. The conditional distribution of $T_{n,n}^*$ under the IID bootstrap method, still converges to a Normal distribution, but with the "wrong" variance. Following the proofs (Lahiri, pg. 21), we see that for all $x \neq 0$, the IID bootstrap estimator of the level-2 parameter has a mean squared error that tends to be nonzero in the limit and the bootstrap estimator is not consistent. Therefore, the IID bootstrap method fails for dependent data. It follows that resampling individual X_i's from the data X_n ignores the dependence structure of the sequence $\{X_n\}$ completely, and thus, fails to account for the lag-covariance terms in the asymptotic variance.

There are many proposed methods to address the above problems. One way to reduce time dependent data to an IID structure is to fit a parametric model. Some authors have focused on adapting the familiar residual-based resampling approach of Efron (1979) for bootstrapping time series. More recently, the focus has shifted toward residual-based nonparametric methods such as the Sieve or Cholesky bootstrap, which treat the underlying population model as unknown. Another approach has been the block resampling method, while still others have developed algorithms which operate in the frequency domain. The advantage of the latter approach is that in the frequency domain, there are IID variables which can be exploited for bootstrapping even when/if the original data are non-IID. However, which algorithm should be employed? These algorithms all differ by the extent to which they impose parametric structure on the data. Those methods which condition on a particular parametric model offer higher precision.

Early applications assumed that the underlying process follows a stationary finite-order autoregression. In practice, almost all of these AR(p) models are estimated by least squares. The problem with this method is that it requires the estimated process to be

stationary. For nonstationary coefficient estimates, the procedure breaks down. Broader classes of models are linear autoregressions of infinite order. If the true model is not finite ordered, the asymptotic justification of the bootstrap approximation is no longer valid either. The Sieve and Cholesky bootstrap were both developed to address this problem. These are all algorithms that transform stationary time series data in a way that gives rise to IID residuals. These residuals may then be resampled with replacement. Differencing and detrending are the most common way to transform non-stationary data into stationary samples. The more common technique is that of differencing. A traditional explanation follows. First, consider the solution for a random walk plus drift model:

$$y_t = y_o + \alpha_o t + \sum_{i=1\ldots t} e_i,$$

where taking the first difference yields $\Delta y_t = \alpha_o + e_t$. The $\{\Delta y_t\}$ is seen as stationary because it is equal to a constant plus a white noise process. $E(\Delta y_t) = \alpha_o$ and $var(\Delta y_t) = \sigma^2$, where the $cov(\Delta y_t, \Delta y_{t-s}) = 0$. Because the mean and variance are constants, and the covariance between Δy_t and Δy_{t-s} depends only on s, the sequence is stationary. This process only leads to a circular argument, because imposing a unit root is involved.

Detrending entails regressing a variable on time and saving the residuals. Consider an example model that is the sum of a deterministic trend and a pure noise component:

$$y_t = y_0 + \alpha_o t + e_t.$$

The first difference of y_t is not invertible in the sense that Δy_t cannot be expressed in the form of an autoregressive process. Instead, a common practice is to regress y_t on a deterministic polynomial trend such as:

$$y_t = y_0 + \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_n t^n + e_t.$$

The appropriate degree of the polynomial can be determined by t-tests, F-tests, and the AIC. The difference between the estimated values of $\{y_t\}$ from the actual values yields an estimate of the stationary sequence $\{e_t\}$. The problem with this practice is including regressors which may not capture the chance regularity accurately. In addition, there is a problem with considering $\{e_t\}$ as stationary as opposed to the observable process $\{y_t\}$.

A different strategy has been to focus on resampling blocks of contiguous time series observations. These are called *moving blocks* by researchers. Resampling overlapping blocks may provide somewhat higher bootstrap estimation efficiency than non-overlapping blocks, although available evidence indicates the efficiency gains may be small. Usually theoretical work on the moving blocks bootstrap assumes short-range dependence. That is, the observations are assumed to satisfy some sort of mixing conditions with a rapidly decaying mixing coefficient. Some of the related problems which must be addressed are selecting the block size and what to do when parameters are on a boundary.

Yet another area of research is resampling in the frequency domain. This area is motivated by noting that, even for non-IID data, there are IID relationships in the frequency domain which can be used for bootstrapping. Many of these methods are distinguished by the statistics they are designed to bootstrap. Unfortunately, they also require a good estimate of the spectral density function (sdf) and thus a bandwidth choice. The effect of the bandwidth choice on the performance of the bootstrap remains an open question.

If the model is known to contain an exact unit root, resampling remains valid if we impose that unit root in estimation. This is done through recognizing the different

implicit parameterizations between the unit root model (UR(1)) and the autoregressive model (AR(1)). It is of paramount importance to recognize these differences because any attempt to impose $\alpha_1=1$ on an AR(1) model leads to a degenerate stochastic process (Spanos and McGuirk).

From above, we have seen that in univariate models, the stochastic trend has been removed (or attempted to be removed) by differencing. The resulting "stationary" series is then estimated using various techniques. For multivariate models, the conventional wisdom was to difference all nonstationary variables used in a regression analysis. However, the appropriate method with which to treat nonstationary variables is not as straightforward as many have proposed. Some claim it is possible that there could be a linear combination of integrated variables that is stationary; such variables are said to be cointegrated (Enders). In particular, the absence of cointegration means that, for example, AR(1) processes can be estimated in first differences and the coefficients converted to the level representation. Similarly, non-parametric methods continue to be valid if resampling is based on first-differenced data. They are also valid if a known cointegrating vector is imposed on the data. Without such transformations, nonparametric resampling cannot be expected to preserve cointegration relationships in the data. Alternatively, parametric methods may be used. In applied work, the exact unit root or cointegration is seldom known. And not all inference problems involve a unit root null hypothesis either (Berkowitz).

All of the previous methods have tried to impose some form of stationarity. However, if the true process is nonstationary, many of the standard results of the asymptotic validity no longer apply for the standard bootstrap. This is the motivation

behind Vinod's Maximum Entropy bootstrap, which he proposes in his paper *Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference* (2004).

Conceptual Framework: Maximum Entropy

The problem has become avoiding the imposition of stationarity for time series analysis. Vinod proposes the use of maximum entropy as a means to create ensembles. Any probability distribution p_i, i=1,…,n of a random variable taking n values provides a measure of uncertainty regarding that random variable. In information theory literature, this measure of uncertainty is called entropy. The origin of the term goes back to thermodynamics. The second law of thermodynamics states that there is an inherent tendency for disorder to increase. A probability distribution gives us a measure of disorder in a sense. Entropy is generally thought of as a measure of expected information, that is, how much information we actually have in the probability distribution p_i. Intuitively, information should be a decreasing function of p_i. The more unlikely an event, the more interesting it is to known that it will actually happen. Shannon (1948) has shown that the ME distribution that maximizes entropy subject just to a normalization condition is the uniform distribution. When additional side conditions are imposed, giving the first two moments of the distribution, the ME distribution is the Normal.

The notion of entropy defines a kind of measure on the space of probability distributions, such as those with high entropy are in some sense favored over others (Jaynes, 1982). The distributions of high entropy represent greater disorder, they are 'smoother,' and they assume less, according to Shannon's interpretation of entropy as an information measure. This paper uses four moments to compare with the Pearson approach, but it is possible to use more moment constraints in deriving ME distributions.

Use of four moment constraints does allow, however, for skewness, kurtosis, and possible modes in the ME distribution.

Shannon (1948) defines entropy (or uncertainty) as: $W=-\int p(x)\log p(x)dx$, where p(x) is a probability density function. Maximizing W subject to various side constraints and conditions is well known in the literature as a method for deriving the forms of minimal information prior distributions. Most often, we want to maximize W subject to:

$$\int p(x)dx=\mu_o'$$
$$\int x\, p(x)dx=\mu_1'$$
$$\int x^2\, p(x)dx=\mu_2'$$
$$\int x^3\, p(x)dx=\mu_3'$$
$$\int x^4\, p(x)dx=\mu_4'$$

where $\mu_o'=1$ and μ_1', μ_2', μ_3', and μ_4' are the given four non-central moments of the distribution. The problem can be expressed as a Lagrangian for which the necessary conditions for a stationary point are: $\log p(x)+1+\lambda_o+\lambda_1 x+ \lambda_2 x^2+ \lambda_3 x^3+ \lambda_4 x^4=0$ plus the five above constraints and the following matrix to be positive semi-definite:

$$\begin{bmatrix} \mu_o' & \mu_1' & \mu_2' \\ \mu_1' & \mu_2' & \mu_3' \\ \mu_2' & \mu_3' & \mu_4' \end{bmatrix}$$

Then our ME distribution takes the following form $p(x|l) = \exp(-[1+l_o+l_1 x+ l_2 x^2+ l_3 x^3+ l_4 x^4])$.

This function can be solved by viewing this as a function of the lambdas, expanding it in a Taylor's series about trivial values for lambdas, dropping the quadratic (in the lambdas) and higher order terms, and then solve iteratively. It should be noted that the first order conditions for the Maximum Likelihood (ML) estimates are mathematically the same as those given above in our ME procedure, with sample

moments replacing the given moments above. A more complete explanation of the iterative process can be found in their paper.

Zellner and Highfield show that the ME approximations seem to approximate the shape of the exact distributions quite well. Differences between the exact and marginal ordinates occur most often at the third and fourth decimal places. Although the largest relative errors occur in the tails, they are still quite small. Note that this is if we are given the first four moments. In Vinod we are not given these first four moments, but this paper provides a useful justification for Maximum Entropy, although surely Maximum Likelihood would also be as useful (Zellner).

A simple choice for this function, already proposed above, is: $H(p)=-\sum p_i \ln(p_i)$. If $p_i=0$ for some i, then the value taken is assumed to be zero. And if $p_i=1/n$ for all I, $H(p)=\ln(n)$ and we have the maximum value of entropy and consequently the least information available from the probability distribution. If we try to find a probability distribution that maximizes entropy, H(p), the optimal solution is of course the uniform distribution ($p^*=i/n$).

In the Bayesian literature, it is common to maximize an entropy measure to find non-informative priors. Jaynes (1957) was the first to consider the problem of finding a prior distribution that maximizes H(p) subject to certain side conditions, which could be given in the form of some side restrictions. Suppose we want to find a least informative probability distribution of a random variable satisfying certain moment restrictions. We can always find some solutions just by satisfying the constraints. However, some maximization makes the resulting probabilities "as smooth as possible" (Bera).

This paper suggests a new method of constructing ensembles by using a maximum entropy density bootstrap from H. D. Vinod. The traditional bootstrap, described in the above section, fails to hold for time series data. In time series analysis, the order in which random realizations occurs is crucial. So theorists work with stochastic processes, defined as a collection of random variables $\{X(w, t), w \in S_{amp}, t \in I\}$, where S_{amp} denotes the sample space containing the set of elementary outcomes and I is the index set. This stochastic process has two explicit arguments, w and t, where t represents the order. However, w_o, the observed time series, represents only one realization of the random variable. A problem that arises is: How can we have a probability theory when there is only one realization? Researchers including Wiener (1930), Kolmogorov (1931), and Khintchine (1934), collectively referred to as WKK for simplicity, developed solutions to this problem. Kolmogorov focused on Markov processes, extrapolations and joint distributions. Wiener constructed stochastic processes as linear combinations of simpler IID processes as building blocks. Khintchine considered stationary processes and a weak law of large numbers. All found that they must regard X_t as a random variable and the entire time series as a random function.

And because f(X) only describes the distribution, we must also describe precisely the dependence and heterogeneity properties of realistic economic time series. If the observations are independent, their joint density is a product of individual densities. If this equality fails to hold, the time series exhibits dependence. Similarly, let $f_{t|past} = f(x_t| x_{t-1},\ldots,x_1)$ denote the conditional density of x_t given x_1 to x_{t-1}. When the densities $f_{t|past}$ are different for two distinct values of t, we have heterogeneity.

The random experiment in traditional theory is motivated by Spanos (1999) with the following three characteristics: 1) all possible distinct outcomes are known in advance; 2) although the outcome is not known for any particular trial, the outcomes exhibit a perceptible regularity of occurrences; and 3) the random experiment can be repeated under identical conditions. In this random experiment, traditional WKK theory envisions distinct sets of $\{x_t\}$ and hence distinct time series w_j for each j=1,… ,J, …, ∞. This infinite set w_j is called the ensemble denoted by Ω and the observed w_o € Ω. The usual lag transformation is defined as: $L^k x_t = x_{t-k}$, where k can be a large positive or negative integer to represent a backwards or forward shift, respectively.

Traditional theory (WKK theory) assumes that these time series data w_o are stationary. (Vinod refers to this assumption as A_{WKK}.) The w_j are imagined to arise from the random experiment described above, where any perceptible regularity is invariant to certain shifts (L^k) and interchange operations on the joint density. The joint density of strictly stationary time series is invariant under the L^k transformation, which shifts each point in the index by a constant k.

However, there are many limitations of the stationary assumption (A_{WKK}) for economic applications. Vinod provides five reasons, along with the formal proofs provided by Lahiri, that the imposition of stationarity is not acceptable for time series analysis.

i. Endowments matter, implying that the stationary assumption can be too strong for time series arising in social sciences. (As an example, initial endowments of wealth)

ii. Stationarity testing is often not definitive. Use of differencing or detrending for creating stationary series requires statistical testing, many of which have low power.

iii. Observed data series in a model have mixtures of orders of integration (mixtures of stationary and non-stationary series). Many economic models have mixtures of I(0) and I(d) variables.

iv. Unrealism of infinite memory.

v. Major historical changes cannot be undone.

A basic description of Vinod's algorithm is described in his paper *Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference* (2004). First, the concept of order statistics is introduced. Reorder the x_t in increasing order of magnitude, which maps the index into a set of ranks. Simply assign tied ranks in the order in which they occur. However, this mapping destroys the dependence and heterogeneity properties of x_t, which is precisely what we are trying to preserve. This is where Vinod's algorithm comes into play.

The traditional bootstrap sample possesses three properties. The traditional bootstrap sample repeats some x_t values while not using exactly as many others. For retaining the perceptible regularity in w_o, the requirement that no resampled values can at all differ from x_t is too restrictive. Also, 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 perceptible regularity in w_o can and should extend beyond the interval $[\min(x_t), \max(x_t)]$. Finally, the traditional bootstrap resample shuffles x_t such that any dependence information in the time series sequence is lost. If we try to

restore the original order to the shuffled resample, we end up with essentially the original set, except that some dropped x_t values are replaced by the repeats of adjacent values. Traditional literature has attempted to remedy this problem by assuming an independent structure and shuffling consecutive blocks of data in the 'moving-blocks bootstrap.' Spanos and Kourtellos (2002) find that these moving blocks bootstraps do not do well in finding replicas of the original data.

Vinod's motivation for this ME bootstrap is that it simultaneously avoids the above three properties without introducing 'unnecessary arbitrariness.' Instead of arbitrarily assuming that the density f(x) is N(0,1), it is more objective to choose the functional form of f(x) which maximizes the entropy, or prior ignorance. The Entropy, *H*, is defined by the mathematical expectation of the Shannon information. Entropy=*H*=E(-log(f(x))). We then select the f(x) which maximizes *H* subject to mass and mean-preserving constraints. Such an f(x) is called the Maximum Entropy (ME) densities.

Intermediate points are the averages of successive order statistics, which help define the ME intervals from pairs of successive z's. Where $z_t= 0.5(x_{(t)}+x_{(t+1)})$. For t=2 to t=T-1 we define the intervals as $I_2=(z_1, z_2),\ldots, \quad I_{T-1}=(z_{T-2}, z_{T-1})$. The ME density also includes $I_1=(-\infty,z_1)$ and $I_T=(z_{T-1},\infty)$. This gives us exactly T intervals, each of which contains exactly one $x_{(t)}$ with probability 1/T. The two constraints are as follows: 1) Mass-preserving: on average, a fraction 1/T of the mass of the probability distribution must lie in each interval, and 2) Mean-preserving: $\sum x_t=\sum x_{(t)}=\sum m_t$ (Where m_t is the mean of f(x) within the interval I_t). The ME density is designed to satisfy mass and mean preserving constraints by 1) preserving the probability mass around each x_t, and 2) preserving E(x bar).

In his paper, Vinod includes seven basic steps to his algorithm. The steps needed to create resamples from the ME density retaining the dependence and heterogeneity of w_o:

1. Define a T×2 sorting matrix called S_1 and place the observed time series x_t in the first column and the index set I={1,2,….T} in the second column.
2. Sort the matrix S_1 with respect to the numbers in its first column. This yields the order statistics $x_{(t)}$ in the first column and a vector I_{ord} if sorted I in the second column. Define z_t and m_t with certain weights on the order statistics $x_{(t)}$ defined earlier. Now construct the intervals I_t.
3. Choose a seed, and create T uniform pseudorandom numbers p_j in the interval [0,1]. Then identify each range wherein each p_j falls, R_t=(t/T, (t+1)/T) for t=0,…,T-1.
4. Match each R_t with each I_t. Use linear interpolation and obtain a set of T values $\{x_{t,t,me}\}$ as the j^{th} resample.
5. Define another T×2 matrix, S_2. Reorder the T members of the set $\{x_{t,t,me}\}$ for the j^{th} resample in an increasing order of magnitude and place them in column one. Place I_{ord} in column two.
6. Sort the S_2 matrix with respect to the second column to restore the order {1, 2, … T}. The jointly sorted column one yields the elements $w_{j,me} = \{x_{j,t}\} \in \Omega$.
7. Repeat steps 1 through 6 a large number of times for j=1, 2, … J.

Vinod maintains that this seven step algorithm gives a constructive approximation to the ensemble, Ω. Given stationary data w_o, traditional theory (WKK assumption revisited) imagines $w_j \in \Omega$ rising from our random experiment, while the perceptible

regularity is invariant with concern to certain shift and interchange operations on the joint density. Under Vinod's assumption (called A_{KEY}), the perceptible regularity, with reference to dependence and heterogeneity, of the observed data w_o of a continuous random variable X, is restored by the reverse map from the ordered set of T numbers $\{x_{t,t,me}\}$. Under A_{KEY}, we can start with non-stationary data w_o, as it is less demanding than the regular assumption, WKK. He proposes a Monte Carlo random experiment based on the seven step ME bootstrap outlined above to approximate the sampling distribution of a statistic b_T which estimates β.

It is now important to revisit the ergodic theorem. Let x-bar$_j$ denote the sample mean of the j^{th} time series $w_{j,me}=\{x_{j,t}\}$ defined by the reverse map over the range t=1,…,T. Let μ denote the population mean, which equals the simple average of the original data, $\{x_t\}=w_o \in \Omega$, under our construction. Because $w_{j,me}$ is one realization of the ensemble, this μ is the ensemble mean for the $w_{j,me}$ time series over all possible j in Ω. In statistics, all ergodic processes permit replacing ensemble averages by their corresponding time averages. Or, symbolically, Lim $_{J\to\infty}$(x-bar$_j$)=μ. (S99 p.424). The Vinod maximum entropy density is designed to satisfy this theorem.

There are three theorems proposed in Vinod's paper. The first is that ergodicity is satisfied through the maximum entropy constructive ensemble. Vinod proposes that $w_{j,me}$ constructed from the ME density with elements $\{x_{j,t}\}$ satisfy the ergodic theorem detailed above. Note again that the maximum entropy density has population mean μ. The algorithm selects J sets of T random numbers p_j from [0,1] and creates $x_{j,t}$ by the inverse map. This random experiment satisfies the mean preserving constraint by construction. Assuming A_{KEY}, let g(Index, w_o) denote the rank correlation coefficient among non-

stationary $\{x_t\}$with mean μ and index of time subscripts. The $w_{j,me}$ created by the inverse mapping from w_o yields J estimates, $b_{T,j}$= g(Index, w_o), which satisfy the ergodic theorem while retaining time dependence of μ to the extent it exists in g. This implies that the satisfaction of the ergodic theorem is unaffected by time-dependent moments in the original data.

The most noted early restriction on the dependence associated with stationary stochastic processes came in the form of ergodicity. This refers to the property of a stationary stochastic process, which enables the modeler to use a single sample path in order to estimate the moments of the distribution underlying the stochastic process. If, for every well-behaved Borel function, the following two conditions are met: 1) $h(\cdot):R \rightarrow R$ such that $E(|h(X_t))|\leq\infty$, and 2)$\lim\sum_{t=1,\ldots,T}h(X_{s,t})=E(h(X_{s,t})):=\int h(x_{t,s})f(x_{t,s})dx_s$, then the stationary process is said to be ergodic. That is, if the limit as $T\rightarrow\infty$ of the time averages of such functions converges to the distribution averages, then the process is ergodic. The time average is over a subset of the index set T, but the distribution average is over s (Spanos 1999).

The second theorem proposed by Vinod is that Doob's Theorem is satisfied. Doob's Theorem states: For any positive integer J, let $\{j_1,\ldots,j_J\}$ be any admissible set of values of J. Then under general conditions, the probabilistic structure of the random process $w_{j,me} = \{x_{j,t}\}$ is completely specified if we are given the joint probability distribution of $\{X_{j1,t},\ldots,X_{jJ,t}\}$ for all values of J and all choices of $\{j_1,\ldots,j_J\}$ (Priestly 1981, p.104). The joint density under traditional WKK theory is created from a product of IID unconditional densities, and is valid under stationarity. Kolmogorov's extension theorem is similar to Doob's theorem and contains links to the conditional density. By

sequentially multiplying the conditional densities, starting with $f_{1|past}$, the resulting density is the same as that of the ME density $f_{me}(x_{(1)},\ldots,x_{(T)})$. That is, the probabilistic foundation of a stochastic process comes in the form of the finite joint distribution of the process. According to this result, the probabilistic structure of a stochastic process can be fully described by a finite dimensional joint distribution. This is commonly referred to as Kolmogorov's extension theorem. Thus, by assuming A_{KEY}, we have used the inverse map to recover the time-subscript-dependence, without hurting the form of the maximum entropy density. Because we have attached the time dimension to the individual segments of the maximum entropy density, we have restored the dependence and heterogeneity properties. So, under the assumption A_{KEY}, the probabilistic structure of the random process $w_{j,me} = \{x_{j,t}\} \in \Omega$ obtained from the inverse mapping from the maximum entropy density is completely satisfied.

The final theorem is completeness. This means that we can ensure uniqueness of any optimal estimator, such as $b_{T,me,j}$ of β. Further assume that $b_{T,me,j}$ is pivotal and asymptotically normal and define its empirical cumulative density function as $G_{me}(z)$. The maximum entropy approximation for G_{me} for pivotal statistics is almost surely G with error of order $O_p(1/T)$ only, and of order $O_p(1/T^{-0.5})$ for non-pivotal statistics.

Completeness revisits our primary goal, which is to find optimal estimators using sufficient statistics. A minimal sufficient statistic cannot guarantee the uniqueness of an estimator. To ensure uniqueness, the modeler needs another property of sufficient statistics called completeness. Intuitively, this amounts to the only unbiased estimator of θ is θ itself. The family of densities Φ is said to complete, if for every function $\tau(X)$, the following relationship holds: $E(\tau(X))=0 \rightarrow \tau(X)=0$ (almost surely) for all $x \in \{x:f_x(x;\theta)>0\}$.

Vinod then goes on in his paper to offer up several numerical examples as proof of his algorithm and its ability to retain perceptible regularity of distribution, dependence, and heterogeneity. It is these examples which will be examined, along with less traditional approaches to testing this new algorithm. In the methodology section, these tests will be explained in further detail.

CHAPTER III

Hypotheses

The hypotheses are based on the experiments proposed in the methodology section, as well as the theoretical framework of Vinod's proposal. The following are the hypotheses:

I. Vinod's new ensemble does not avoid the use of stationarity.

II. Vinod's new ensemble satisfies the mass and mean preserving constraints.

III. Vinod's new ensemble satisfies both the Ergodic and Kolmogorov theorems.

IV. The data generated in experiment one (Normal IID) will be faithfully replicated using Vinod's algorithm. (For a more expansive discussion of the term 'faithfully replicated,' please see the methodology section.)

V. The data generated in experiments two through five will not be faithfully replicated using Vinod's algorithm.

CHPATER IV

Methodology and Data

A conditional distribution $f(Y_t|X_t)$ exists with mean $f(Y_t|X_t=x_t)=B_o+B_1'x_t$ and variance $Var(Y_t|X_t=x_t)=\sigma^2$. If we assume a typical linear regression model, then it is known that $B_1=\sigma_{22}^{-1}\sigma_{12}$, $B_o=\mu_1-B_1'\mu_2$, and $\sigma^2=\sigma_{11}-\sigma_{12}'\sigma_{22}^{-1}\sigma_{12}$. These come from the joint distribution $(Y_t, X_t)'\sim N[(\mu_1, \mu_2)', (\sigma_{11}\ \sigma_{12}, \sigma_{12}\ \sigma_{22})]$. As an example, we can specify that $\mu_1=1$, $\mu_2=2$, $\sigma_{11}=4$, $\sigma_{12}=-1$, $\sigma_{12}=-1$, and $\sigma_{22}=2$, we know the true parameters to be $B_1=-0.5$, $B_o=2$, and $\sigma^2=3.5$. Similarly, the skewness should be equal to zero (a symmetric distribution) while the kurtosis should be equal to three.

I plan to use the simulation experiments in Spanos and Kourtellos (2002) as the basis for testing Vinod's algorithm. All experiments are based on 10,000 bootstrap replications of sample size n=100. Consider the testing of the following hypothesis in the context of the simple Normal model:

$$H_o: \mu=0 \text{ against } H_a: \mu\neq 0,$$

at significance level $\alpha=0.05$. The test-statistic is the well-known t-statistic:

$$\tau_n=\sqrt{(nY_n)}/s_n\sim St(99),$$

with rejection region:

$$C_R=\{|\tau_n|>1.984\}.$$

In order to show how faithfully the Resampling method replicates the probabilistic structure of the observed data, compare the sampling distribution of τ_n under H_o generated by the Monte Carlo simulation with that generated by the Maximum Entropy bootstrap, the technique in question. To assess the reliability of the hypothesis testing, we will be estimating 1) the density of the resampled test statistic, 2) the error

rejection probability (ERP) for H_o, 3) the skewness coefficient, and 4) the kurtosis coefficient. The skewness measure α_3 is a measure of symmetry, while the kurtosis α_4 measures the peakedness. With Normally distributed variables, $\alpha_3=0$ and $\alpha_4=3$. There will be 5 main tests performed. They are discussed below. All testing and programming will be done in GAUSS. Consider the data $\{y_t, t=1,\ldots,100\}$, generated by the simple Normal (IID) model of the form: $y_t= u_t$, $u_t \sim N(0,1)$.

1. **Experiment 1**: The Independent and Identically Distributed (IID): $y_t= u_t$, $u_t \sim N(0,1)$ This is a base test upon which all other testing can be compared. The resulting model should be $Y_t=2-0.5X_t$. The next tests will represent data with "problems." That is, it will no longer be simple Normal data and will truly test whether Vinod's algorithm can accurately replicate the probabilistic structure of the observed data. This original test will not be done on time series data, in order to ensure the soundness of the technique and programming.
2. **Experiment 2**: Heterogeneity in the Mean/Trending mean: Now drop the assumption of identical distribution and consider the data $\{y_t, t=1,\ldots,100\}$, generated by the statistical model of the form: $y_t=0.15t' + u_t$, $u_t \sim N(0,1)$, where t' is a linear trend centered about zero. This represents a trending mean.
3. **Experiment 3**: Heterogeneity in the Variance/Trending variance: Now introduce time heterogeneity in the variance and consider the data $\{y_t, t=1,\ldots,100\}$, generated by a model of the form: $y_t=\sigma_t u_t$, $\sigma_t^2=0.8t$, $u_t \sim N(0,1)$, where t is a linear trend. This represents a trending variance.

4. **Experiment 4**: Heterogeneity in the Mean/Discrete Mean Shift: Now introduce time heterogeneity in the mean as a discrete shift. This differs from Experiment 2 in that the heterogeneity is not a trend, but a jump to another mean. Such a discrete shift could have resulted from a historical change. Again, we expect the slope to remain unchanged as the variance-covariance matrix does not change.
5. **Experiment 5**: Heteroskedasticity: Consider data $\{y_t, t=1,\ldots,100\}$ generated by a heteroskedastic model of the form: $y_t = \sigma_t u_t$, $\sigma_t^2 = (0.75y_{t-1})^2$, $u_t \sim N(0,1)$. This is different than heterogeneity of the variance in that the variance changes with an explanatory variable. In a heterogeneous variance, the variance changes with t. That is, it is index-dependent. Homoskedasticity is not a result of the stationary reduction assumption. It is a direct result of the Normality reduction assumption. However, it is important that if the Maximum Entropy bootstrap can avoid stationarity, it can also handle heteroskedastic data as well.

Each experiment will be examined on the described characteristics of faithful replication of the observed data and reliability of inference. The data will be generated as described above using GAUSS, and then the data with "problems" will be run through Vinod's algorithm in his provided GAUSS programs, collectively referred to as meboot.g.

Revisiting the seven steps of Dr. Vinod's algorithm:

1. Define a T x 2 sorting matrix (S_1) and place the observed time series in the first column, and the index set in the second column.

2. Sort the numbers in S_1 with respect to the first column. Now the first column contains order statistics and a vector of sorted indices in the second. Define z_t and m_t with certain weights on the order statistics. Construct the intervals, I_t.
3. Choose a seed, and create T uniform pseudorandom numbers p_j in the interval [0,1]. Then identify the range $R_t=(t/T, (t+1)/T)$ for t=0,…,T-1.
4. Match each R_t with I_t. Obtain a set of T values $\{x_{j,t,ME}\}$ as the j-th resample. This is where the mean preserving constraint comes in.
5. Define another T x 2 sorting matrix (S_2). Reorder the T members of the set $\{x_{j,t,ME}\}$ for the j-th resample in an increasing order of magnitude and place them in column one. Also, place the sorted indices from step 2 in column 2 of S_2.
6. Sort S_2 with respect to the second column to restore the order. The jointly sorted column 1 yields the elements in w_j, a resample endowed with perceptible regularity.
7. Repeat the above steps a large number of times.

Dr. Vinod has already coded his bootstrap technique in GAUSS. His zip file contains many individual files; a brief summary of the most important procedures, denoted in bold, follows.

meboot8 (This is step one and step two in Vinod's seven-step algorithm.) Creates shuffled values by the ME method. Steps 1 and 2 create a sorting (S_1) matrix with second column having 1,2,...T. It is called xseq and is used to sort on the first column to get order statistics. Then we go on to get the intermediate points from order statistics of the data. X denotes the sorted order stats. Since the tails are also uniformly distributed, we

need to find the lower limit for the left tail and upper limit for the right tail. This is done by computing the 10% trimmed mean of deviations among all consecutive observations. Z becomes the place where we store the intermediate points. We also sort probabilities from the uniform random generation (**rndus**)- sorted p-values. This then goes on to define the interval means:

desintxb[1]=0.75*x[1]+0.25*x[2]; (this is denoted m_1 in the paper)

desintxb[n]=0.25*x[n-1]+0.75*x[n]; (this is m_T in the paper).

These are then sorted into quartiles. However, we still need to give them time series dependence and heterogeneity, by sorting from one to T.

The procedure **mebootx1** creates shuffled values via the maximum entropy method. It requires **intpola2.g** to work, a procedure that interpolates an answer from a table. x will denote order stats for convenience in finding desired means. The mean preserving constraints are as follows:

desired mean= 0.75x1 + 0.25 x2 in ending left side interval

desired means =0.25 x1 + 0.5 x2 + 0.25 x3 for intermediate intervals

desired mean= 0.75x(n-1) + 0.25 x(n) in ending right side interval

Next, input the items to be shuffled, denoted by xx, which is an (n x 1) vector and we input random numbers between 0 and 1, denoted by p, which is also an (n x 1) vector. Then the procedure goes on to created a sorted matrix- first column of order statistics, the second column of 1, 2, … , T. Then compute the 10% trimmed mean of deviations among all observations. This is done b/c the tails are uniformly distributed and we need to find lower and upper limits for the tails. Store the absolute deviations in Z, then get z_t values, the intermediate points from data. The program then defines and array of uniform random

numbers from smallest to largest. Using this, we then define and save the desired interval means. If the mean falls in the left tail, or right tail, it is stored after sorting. If the mean falls in the middle, the procedure must use interpolation. Then the values are sorted on the second column (1,2,…,T) to bring them in order 1 to T.

The procedure **cenmomsa** (Central Moment Sample) computes Vinod's numerically accurate sample, dividing by (n-1). This requires **mean2c.g**, an averaging routine. This then finds the second, third and fourth moments measured from the mean (from an n by p matrix, denoted by x) and save each in a (p by 1) vector. Finally, the procedure then finds column means.

ensem checks to ensure the Ergodic theorem is met. Recall we want the values to be robust, that the trimmed means must be close. From the original data set, define:

```
dsori[1]=bigt;
dsori[2]=meanc(xx);
 dsori[3]=stdc(xx);
 dsori[4]=minc(xx);
  dsori[5]=maxc(xx);
dsori[6]=trimmed3(xx, thalf);
```

These are the trimmed means, so ideally we want abs(dsori[6]-dsens[6]) to be close, where dsens[6] is the trimmed mean of the ensemble. Initial value is set at the value in the data. The **ensem** procedure uses **acf.g** and **pacf.g** which compute sample autocorrelations and sample partial autocorrelations respectively.

Another procedure, **ensemx** (this version calls mebootx1) creates an ensemble of J values for each series in the columns of xx matrix. It is important to define the variables in Vinod's programs:

Input:

xx=T by 1 matrix of data

bigJ= size of the bootstrap

seed= seed for the uniform random variable

n5=number of ensembles, say 5, wanted by the user for further analysis

nacf=number of autocorr and pacf wanted by the user.

Output:

collect= the T by bigJ matrix of the ensemble

dsori=descriptive stats for original data

dsens=descriptive stats for ensemble excluding original data

yacf=n5 by nacf matrix of acf, first row is for original data

ypacf=n5 by nacf matrix of pacf, first row is for original data

Then the procedure goes on to use **mebootx1** to do the Maximum Entropy bootstrap.

CHAPTER V

Results

Table 1 summarizes the results of the Monte Carlo experiments described above. The parameter estimates and standard errors are reported, along with the skewness and kurtosis values. Recall that the standard t-statistic for the significance of a parameter is the estimate divided by the standard error. The final row in Table 1 is a summary of the Bera Jarque test for Normality. The null hypothesis for this test is skewness=0 and kurtosis=3 (Normally distributed data). The alternative rejects Normality in favor of some alternate distribution. In addition to these estimates, the ensembles are graphed as t plots and a histogram of the residuals in included for each experiment as well.

Table 1: Summary of Statistics from Monte Carlo Experiments

	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
B_o	1.8129713	1.3141665	1.6429649	3.7004458	1.6315462
$Se(B_o)$	0.41442674	0.27962299	0.26331258	0.54447424	1.016940
B_1	-0.46442674	-0.20399104	-0.29254525	-0.37548285	0.27725306
$Se(B_1)$	0.16566636	0.12517907	0.10836661	0.13124747	0.041814462
Skewness	-0.12982185	-0.02661555	0.21456481	0.065031596	0.000386026
kurtosis	2.8065105	2.6095774	2.6942313	2.7715708	2.7913738
Normality	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Fail to reject

The baseline simulation, that of Normal IID data, results in a true equation of y=2-0.5x+u. The experiment yields y=1.8129713-0.46442674x. The coefficients are individually significant at the 0.05% level, and both fail to reject the null that their values are the same as hypothesized (H_o: B_o=2 against H_a: $B_o \neq 2$). The skewness and kurtosis

values, along with the Bera Jarque test, indicate that the Maximum Entropy bootstrap yields Normally distributed sampling densities.

Now the data becomes time series, with the x corresponding to the lagged values of y. Upon relaxing the assumption of identical distribution, a time trend has been introduced to the mean. The slope coefficient should still equal -0.5, as the variance covariance matrix remains unchanged. However, our intercept (B_o) becomes (1+0.15t)-0.5(2+0.15t)=0+0.075t. There should now be three parameter estimates now where B_o=0, B_1=0.075, and B_2=-0.5. However, the ME bootstrap fails to introduce this new parameter and returns the equation y=1.3141665-0.20399104x. The skewness indicates a Normal distribution, but kurtosis is more peaked than a Normal distribution. However, the Bera Jarque test indicated Normality. In addition, the trend is not captured. Both parameter estimates are indicated as statistically significant (different than the null of zero), but do reject the null of their hypothesized values, which should be expected. It should be mentioned that in the presence of misspecification errors (an incorrectly specified model), the actual size of these tests would not be 5%, but rather much higher. When assessing a misspecified model, these t-tests at the 5% level are essentially meaningless. However, they have been included as traditional literature often overlooks this shortcoming and proceeds with analysis and inference anyway.

In experiment three, with variance heterogeneity, the true slope coefficient becomes –0.4t, while the intercept equals 1-0.8t. Our new regressor should not be simply x, but x*t. The true regression equation is y=1 -0.8t -0.4x*t +u. Clearly the maximum entropy bootstrap fails to capture this variance heterogeneity, while indicating that the coefficients are both individually significant and close to their hypothesized values of 2

and –0.5. The regression equation estimated is y=1.6429649-0.29254525x. The slope coefficient is marginally significant, while the intercept remains significant. However, both the intercept and slope should change with a heterogeneous variance, which introduces a time trend. The skewness indicates a positively skewed distribution while the kurtosis fluctuates below three, but both again fail to reject the null of Normality.

Another heterogeneous mean is modeled in experiment four. This time, the heterogeneity comes in the form of a discrete shift at time t=50, rather than a linear trend. The true equation is y=2+2D-0.5x+u, where D is a dummy variable taking values 0 for t=1,2,…50 and taking value 1 for t=51, 52,…100. The bootstrap yields an estimated equation of y=3.7004458-0.37548285x**.** The skewness and kurtosis indicates Normality. However, for the discrete mean shift, a dummy variable should be included in the regression to differentiate the means before time t=50 and after. The maximum entropy bootstrap has treated the mean shift as if it were time t=1 instead and therefore lacks the inherent heterogeneity.

The final experiment tests the bootstrap on heteroskedastic data. This differs from heterogeneous variance in that the former changes with the regressor while the latter changes with time. With heteroskedasticity, the skedastic function depends on values of the conditioning variable. Formally, $Var(Y \mid X=x)=h(x)$ is heteroskedastic, while $Var(Y \mid X=x)=c_o$ is homoskedastic, with c_o being a constant. The AR(1) and bivariate Normal models are both homoskedastic, while the bivariate student's t is one example of a heteroskedastic distribution. The homoskedastic model assumption is a direct result of the Normality reduction assumption for both the AR(1) and NLRM. In the presence of heteroskedasticity, the reduction assumption of Normality must be tested for

misspecification errors. In the presence of heteroskedasticity, the ME bootstrap indicates that the data are symmetric (skewness≈0), but less peaked than a Normal distribution (kurtosis fluctuates below three). The Bera Jarque Normality test does indicate Normality. This is a potential problem, as homoskedasticity of the model is a direct result of the Normality reduction assumption as previously noted. The modeler cannot retain one while rejecting the other. Looking at the histogram of the residuals, along with the skewness and kurtosis coefficients can help shed some light on the true underlying distribution.

Another assessment of how faithfully the ME bootstrap replicates the observed data is to look at graphs of the ensembles. Figure one represents the plots of the bootstrapped sample for simplest case, experiment one. The t plots appear to represent normal, independent, and identically distributed random variables. This is exactly what was specified in the data generation and the maximum entropy bootstrap has captured the NIID structure of the sample. Further misspecification testing should be done in order to ensure the soundness of these assumptions.

In figure two, from experiment two with a trending mean, the plots are less indicative of the underlying data structure. The t plots of the explanatory and dependent variables do not reflect the trending mean that the data were specified to have. This is because the model was not correctly specified when the bootstrap was run, rendering our inference unreliable. If the data had been assessed a priori and a correctly specified model were chosen, then the bootstrap may have been able to capture the structure of the observed process. A linear trend needs to be added as a regressor in this experiment.

The data for experiment three was prespecified to have a trending variance. Upon looking at figure three, the t plot of the residuals indicates that this trend is not captured by the ME bootstrap. In addition, the t-plot of Y seems to show a decreasing variance of Y over time while the t-plot of X (the lagged values of Y) indicates the variance of X increasing over time.

In experiment four, the heterogeneity introduced is in the form of a discrete mean shift, which could occur as a result of a historical change. The t plots do not reflect this discrete jump in the means of the dependent or explanatory variable. Instead, they seem to show that the discrete shift occurred at time=1 rather than time t=50. Including a dummy variable as a regressor would allow the ensembles to reflect two different time periods within the same analysis.

The final experiment does not test stationarity, but rather heteroskedasticity. This is a direct result of the Normality reduction assumption, yielding the model assumption of Homoskedasticity. The simulated data for this experiment contain a heteroskedastic variance, but this is not reflected in the residual t-plots or the histogram. The skewness and kurtosis coefficients indicate data that are Normal as well, so the ME bootstrap fails for this model assumption.

It must be mentioned that unfortunately, the Bera Jarque Normality test is the least powerful of all the tests for Normality. For the first model, the test indicates Normality, however a skewness coefficient of -0.12 indicates a negatively skewed distribution. This suggests that even for the NIID case the ME bootstrap may not work as well as Efron's original bootstrap. This also explains the discrepancy between the heteroskedastic model and the passing of the Normality misspecification test. As

mentioned previously, the Normality reduction assumption leads directly to a model assumption of homoskedasticity.

What then should be done in order to capture the non-stationary data? Finding a correctly specified model is of the utmost importance. The data need to be assessed a priori in order to determine if, for the AR(1), the reduction assumptions of Normal, Markov and Stationary are appropriate for the model. This can be done through looking at plots of the *original* sample such as t plots of the variables and the residuals and histograms of the residuals.

CHAPTER IV

Conclusions

The assumption of stationarity refers to the time-homogeneity of the model parameters. Often, time series data does not exhibit stationarity which poses many problems. The traditional bootstrap fails to capture this non-stationarity and many researchers turn to differencing or detrending to result in a stationary series. Again, there are problems with these processes which can destroy the economic theory behind the original time series. The Maximum Entropy bootstrap proposed by Dr. H. D. Vinod in *Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference* (2004) is an attempt to deal with non-stationary data.

Through Monte Carlo simulations, the reliability of Vinod's bootstrap was tested. The first experiment was done on Normal, Independent, and Identically Distributed (NIIID) data. The ME bootstrap performed well for this sample. The second, third, and fourth experiments introduced non-stationarity by dropping the ID assumption. The second experiment introduced a linear trend, the third a trending variance, and the fourth a discrete mean shift. The ME bootstrap failed to faithfully replicate the prespecified properties of the data. The final experiment was done with heteroskedastic data, which is not a form of non-stationarity, but is important to examine. The ME bootstrap also failed for this experiment.

The importance of a priori analysis becomes very clear when specifying models and using bootstraps. The appropriate statistical model must be chosen to account for all the systematic information in the observed data series. This statistical systematic information comes in the form of chance regularity patterns exhibited by the time series

data. The success of modeling, bootstrapping, and inference depends upon recognizing the regularity patterns and choosing the appropriate statistical concepts. These concepts are chosen in the form of reduction assumptions, which then translate into model assumptions.

In order for the ME bootstrap to become successful, the model must be assessed a priori. That is, the t plots of the variables and residuals should be examined. For example, this would help illuminate the need for a linear trend as in experiment two. As it is, the ME bootstrap fails to recognize this trend and thus any inference would be unreliable. The remainder of the experiments indicate that the ME bootstrap must allow for trends and heteroskedasticity in order to faithfully replicate the data.

Appendix

Figure One: NIID data

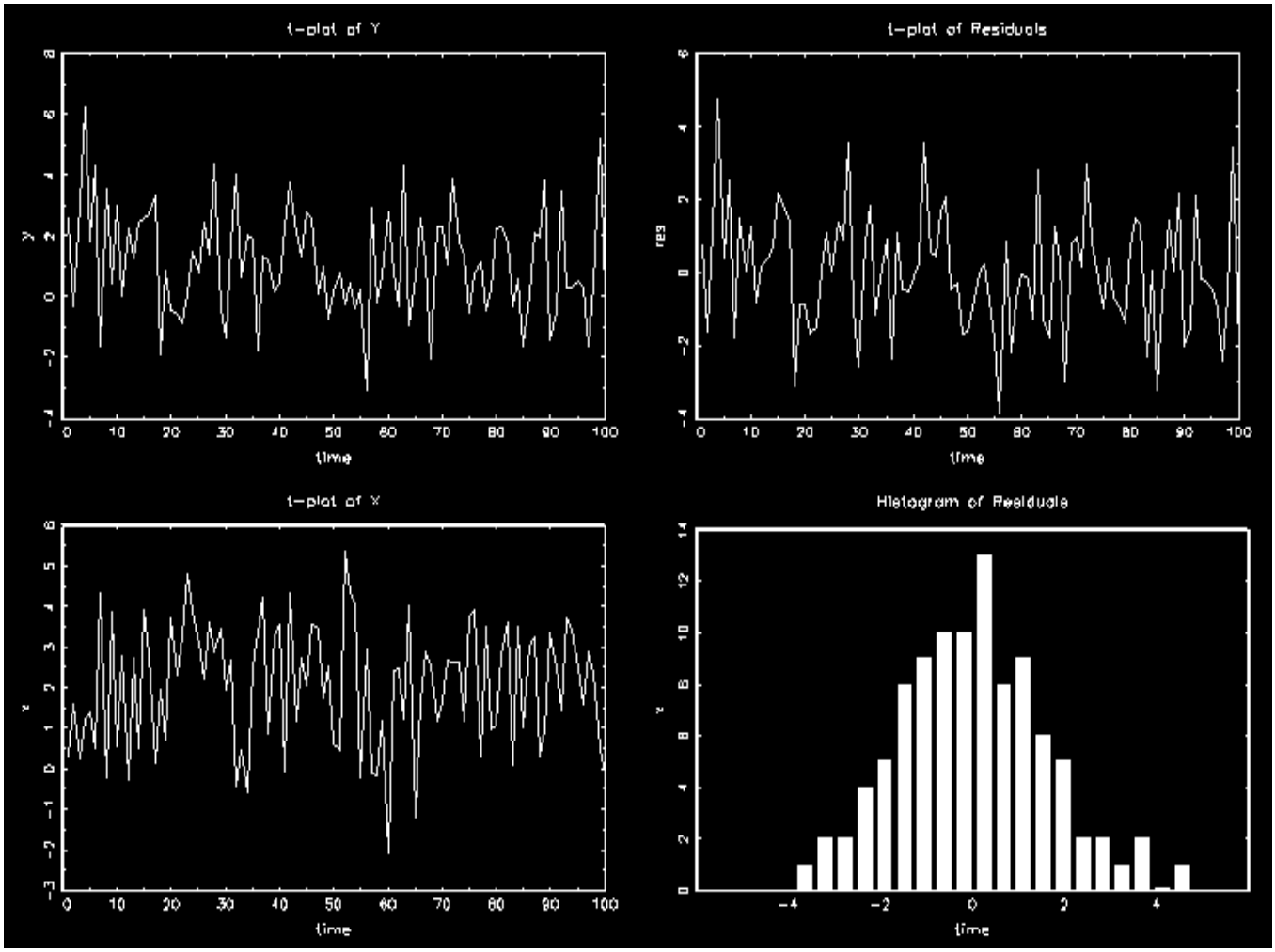


Figure Two: Linear Trend in Mean

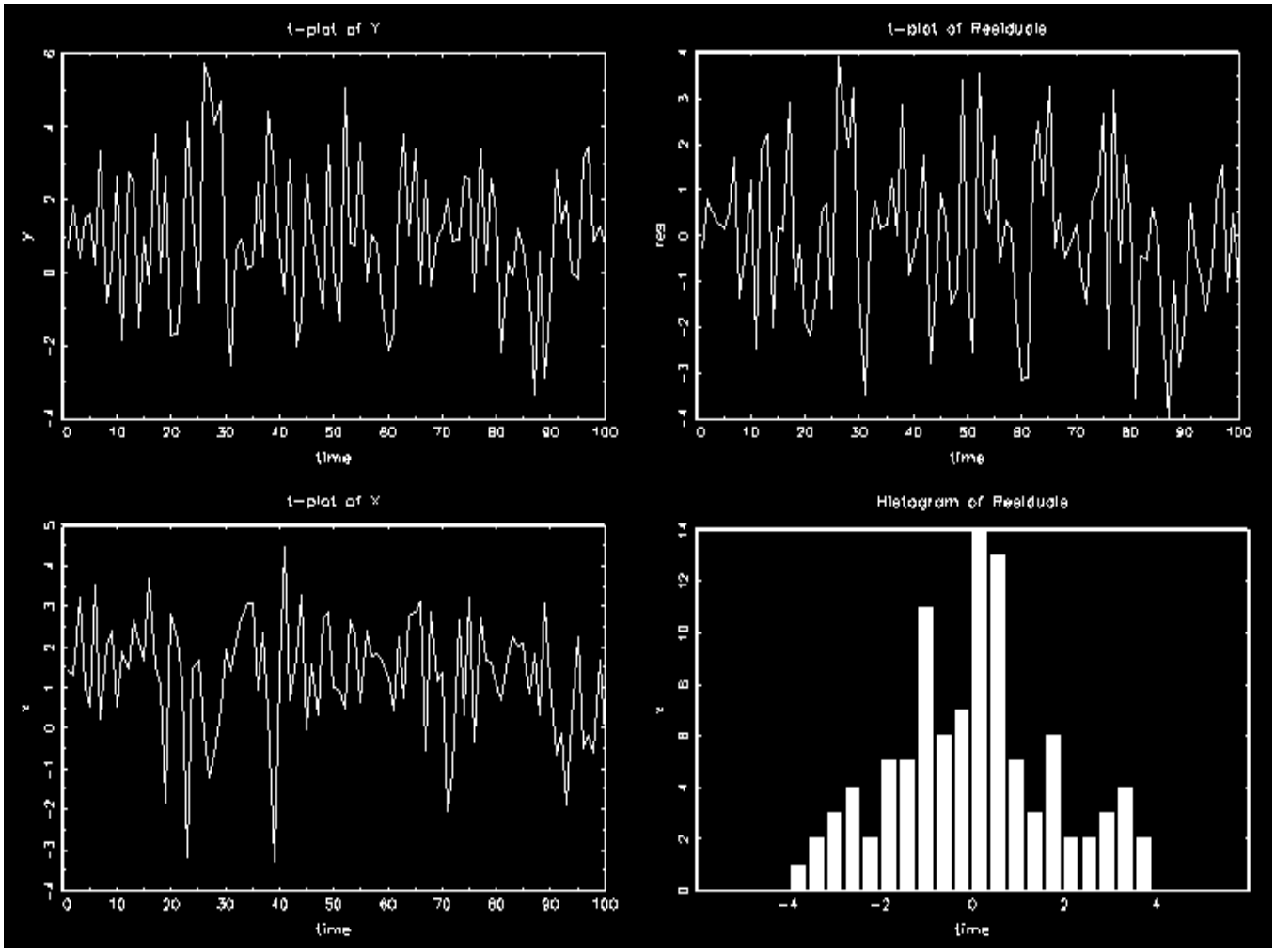


Figure Three: Trending Variance

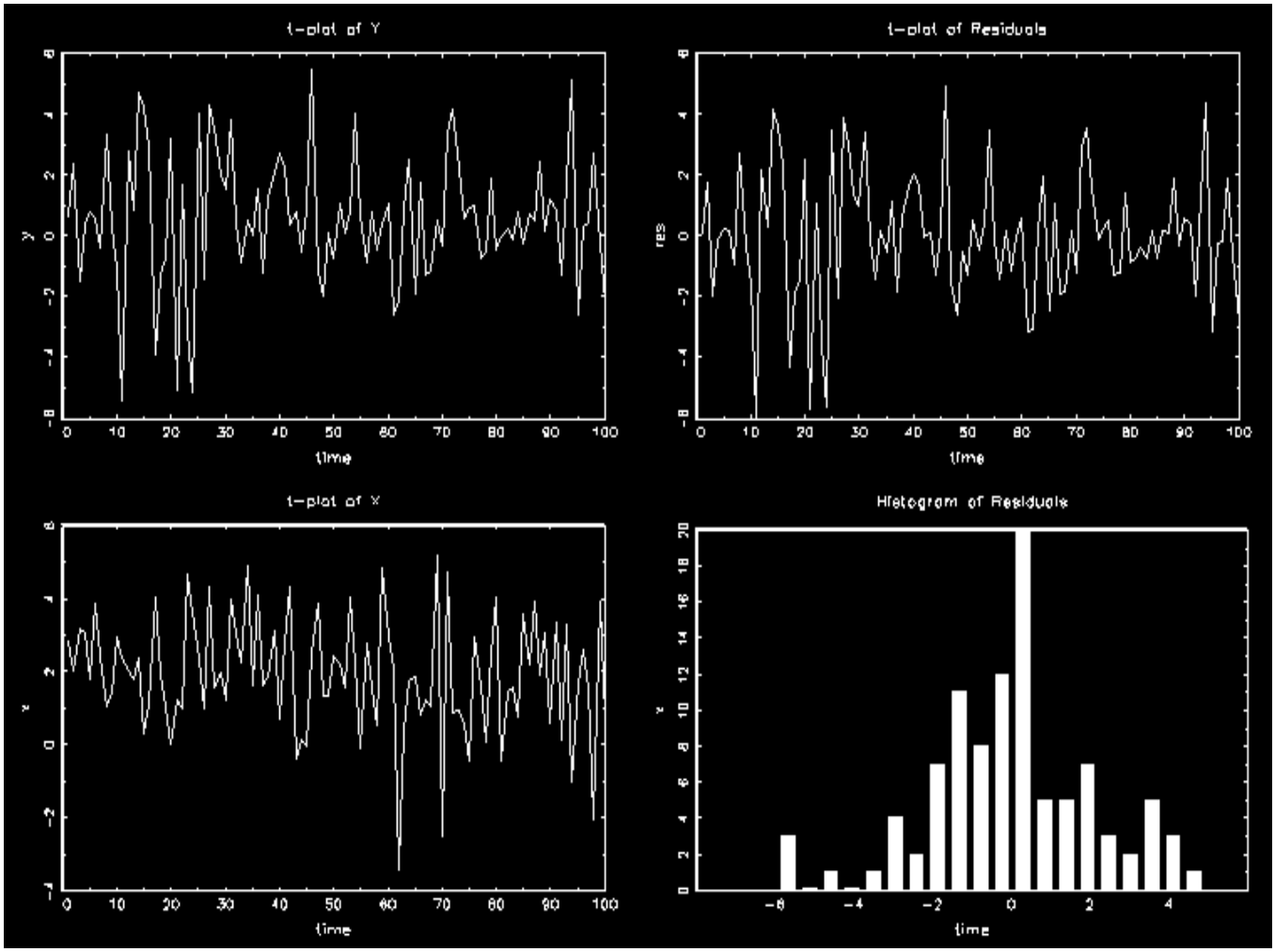


Figure Four: Discrete Mean Shift

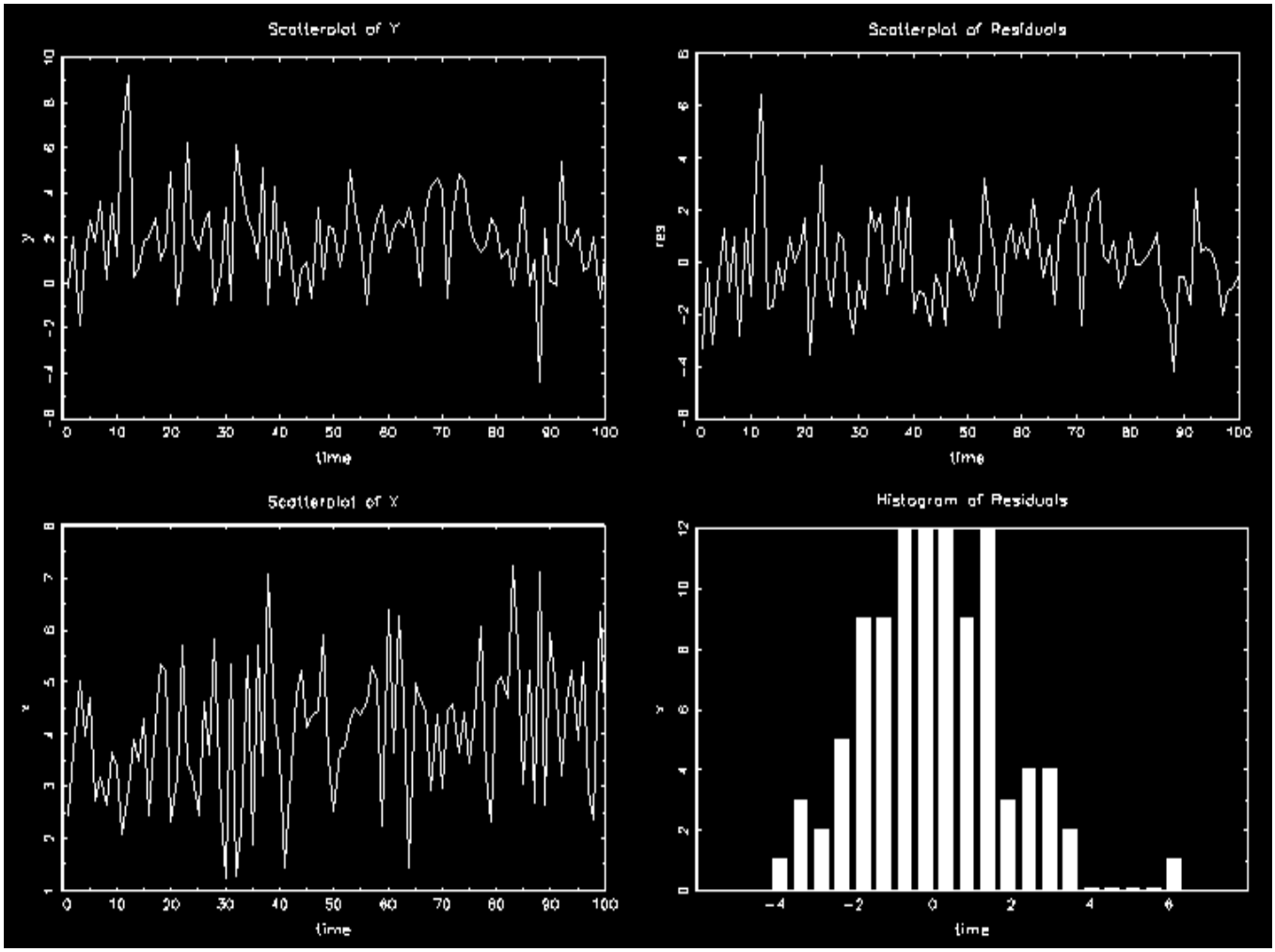
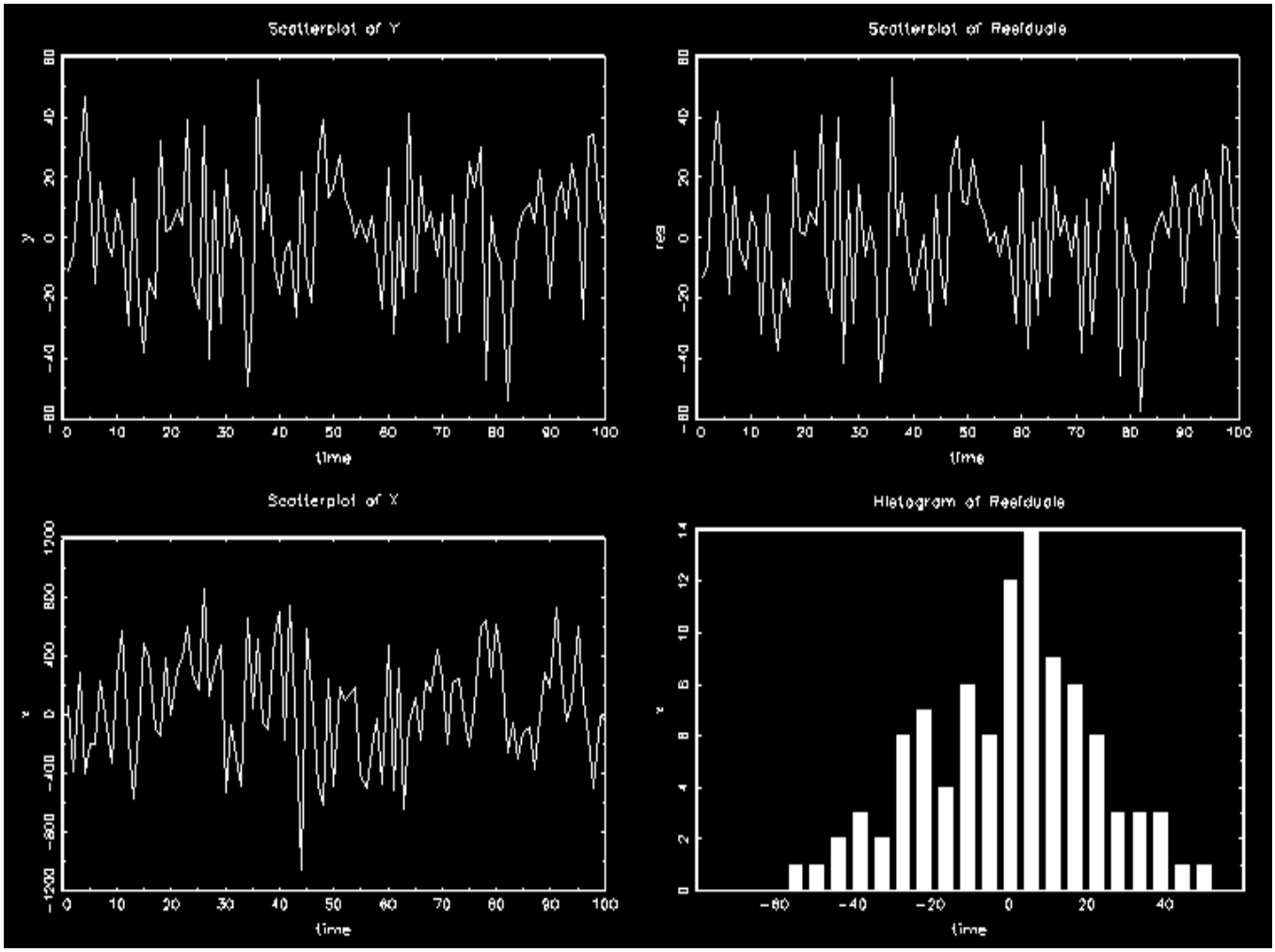


Figure Five: Heteroskedasticity



Works Cited- For Dr. Vinod's Maximum Entropy Bootstrap

Bera, Anil and Yannis Bilias. *The MM, ME, ML, EL, EF, and GMM Approaches to Estimation: A Synthesis*. Journal of Econometrics. Volume 107, Issues 1-2, March 2002, pages 51-86.

Berkowitz, Jeremy and Lutz Kilian. *Recent Developments in Bootstrapping Time Series*. Econometric Reviews, 19(1), 1-48 (2000).

Brillinger, David R.. *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart, and Winston, INC. (1975).

Chatfield, C.. *The Analysis of Time series: An Introduction*. London: Chapman and Hall (1980).

Efron, Bradley and Robert Tibshirani. *An Introduction to the Bootstrap*. New York: Chapman and Hall (1993).

Hilmer, Christiana E.. *A Comparison of Resampling Techniques when Parameters are on a Boundary: The Bootstrap, Subsample Bootstrap, and Subsample Jackknife*. Unpublished Technical Report (2004).

Lahiri, S. N.. *Resampling Methods for Dependent Data*. New York: Springer (2003).

Spanos, Aris. *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge University Press: Cambridge (1999).

Spanos, Aris. *Statistical Foundations of Econometric Modeling*. Cambridge University Press: Cambridge (1986).

Spanos, Aris and Andros Kourtellos. *Model Validation and Resampling*. Lecture notes/Technical report (July 2002).

Vinod, H. D.. *Maximum Entropy Constructive Ensembles for Time Series Analysis and Inference*. Unpublished Technical Report (2004).

Zellner, Arnold and Richard A. Highfield. *Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions*. Journal of Econometrics 37 (1988) 195-209.