Testing for Structural Breaks and other forms of Non-stationarity: a Misspecification Perspective*

Maria S. Heracleous† **Andreas Koutris**‡ **Aris Spanos**§

Abstract

In the 1980s and 1990s the issue of non-stationarity in economic time series has been discussed in the context of unit roots vs. mean trends in AR(p) models. More recently this perspective has been extended to include structural breaks. In this paper we take a much broader perspective by viewing the problem of changing parameters as one of misspecification testing due to the non-stationarity of the underlying process. The proposed misspecification testing procedure relies on resampling techniques to enhance the informational content of the observed data in an attempt to capture heterogeneity 'locally' using rolling window estimators of the primary moments of the stochastic process. The effectiveness of the testing procedure is assessed using extensive Monte Carlo simulations.

Keywords: Maximum Entropy Bootstrap, Non-Stationarity, Parameter/Structural Stability, Time-varying parameters, Rolling estimates

JEL classifications: C22

*Thanks are due to Amos Golan, Essie Maasoumi and two anonymous referees for valuable comments/suggestions.

†Department of Economics, American University, Washington DC, 20016, USA. <*email*: heracleo@american.edu>

‡Department of Economics, Virginia Tech (0316), Blacksburg, VA 24061, USA. <*email*: akoutris@vt.edu>

§Department of Economics, Virginia Tech (0316), Blacksburg, VA 24061, USA. <*email*: aris@vt.edu>

1 Introduction

One of the most crucial assumptions underlying empirical modeling in econometrics is that of the constancy (homogeneity, invariance) of model parameters. For example, in the simple AR(1) model, $y_t = a_0 + a_1 y_{t-1} + u_t$, $t \in \mathbb{T}:=\{0, 1, ...\}$, the parameters (a_0, a_1, σ^2) are assumed to be constant ($t-$invariant). This assumption lies at the heart of empirical modeling because without it no reliable statistical modeling and inference is possible. Indeed, the primary objective of empirical modeling is to capture the invariant features of the phenomenon of interest in the form of such constant parameters. Despite its importance, this assumption is rarely tested in practice and the current tools for detecting essential forms of departures are not adequately effective.

It is widely appreciated that economic data usually exhibit heterogeneity over time (times series) as well as over individuals (cross-section). In time series modeling this heterogeneity occurs frequently because, invariably, economies grow and change over time. The main reason why econometricians were very slow to adopt the ARMA(p,q) model was the fact that this model assumes stationarity. This form of modeling became popular after Box and Jenkins (1970) proposed a way to address the presence of heterogeneity in economic time series using differencing. That led to a revival of time series modeling in econometrics, but raised the question of the appropriateness of differencing as a general way of addressing non-stationarity. The 'unit root' literature, led by Dickey and Fuller (1979) and Phillips (1986, 1987), provided certain partial answers to this question, but it also gave the misleading impression that unit roots (UR) and cointegration (Engle and Granger, 1987, Johansen, 1991) provide general ways to capture the time heterogeneity in general. In the case of the UR(1) model, which corresponds to the AR(1) model with $a_1 = 1$, the most general form of heterogeneity implicitly imposed is that $E(y_t) = \mu_0 + \mu_1 t$, and $Cov(y_t, y_{t-k}) = \sigma(0)\cdot(t-k)$, where $k = 0, 1, ..., t$. Perron (1989) raised the issue of another form of heterogeneity, structural breaks, that economic time series often exhibit, but cannot be captured by unit root modeling, advancing the possibility that a_0 or/and a_1 might change abruptly at specific points t_0. More generally, one can make a strong case that the use of differencing and time trends in the mean in the context of AR(p) models accounts for only a small fraction of heterogeneity structures one can expect in time series modeling; see Spanos (1999), ch. 8.

Testing and estimation of parameters in the context of statistical models that are subject to t-heterogeneity has been the subject of considerable research. One of the pioneering studies in this area is that of Chow (1960), who proposed an F-test for a single structural break in a linear regression model. However an important limitation of this test is that the date of the break must be known. To overcome this problem researchers have developed testing procedures which do not presuppose knowledge of the break point(s). Quandt (1960) proposed choosing the largest Chow statistic over all possible break points. Brown, Durbin and Evans (1975) developed an alternative procedure based on recursive residuals, by proposing the cumulative sum (CUSUM) and CUSUM squared tests to deal with cases where the break point is unknown. Recent work has extended these tests in several directions to allow for multiple breaks, unit root dynamics and heteroskedasticity. Some important contributions include Nyblom's (1989) test for martingale parameter variation, Andrews's (1993) asymptotic theory for Quandt's (1960) test, and the exponentially weighted tests of Andrews and Ploberger (1994). Also Ploberger, Kramer and Kontrus (1989), Hansen (1992), Andrews, Lee and Ploberger (1996), and Bai and Perron (1998, 2003) develop tests for consistently estimating the size and timing of the breaks. For a recent survey of the structural break literature see Perron (2005). Most of the tests developed in the structural break literature are designed to detect discrete shifts in the model parameters.

In this paper we develop an alternative approach to testing for t-invariance of the model parameters, based on the stationarity of the primary sample moments; means, variances and covariances. The proposed procedure differs from the existing literature in two important ways. *First*, it focuses on detecting more general forms of non-stationarity rather than just abrupt changes. *Second,* it is based on rolling window estimates of the primary moments (mean, variance and covariance) of the variables, rather than the model parameters**.** The rationale is that the model parameters $\boldsymbol{\theta}$ are functions ($\boldsymbol{\theta} = \mathbf{H}(\boldsymbol{\varphi})$) of the primary moments $\boldsymbol{\varphi}$ of the underlying stochastic process and any t-heterogeneity in the latter ($\boldsymbol{\varphi}(t)$) is almost certain to be reflected in $\boldsymbol{\theta}(t)$. Focusing on the first two moments, mean, variance and covariance is motivated by the fact that higher central moments, in terms of which the model parameters are specified, are functions of the first two moments. Using a single realization of a non-stationary and highly dependent process is often inadequate for a thorough probing for departures from the parameter t-invariance assumption. Hence, to implement this procedure we use the Maximum Entropy (ME) density bootstrap of Vinod (2004) in order to enhance the available data information by generating several faithful replicas of the original data. We carry out a number of Monte Carlo experiments to demonstrate and evaluate the performance of the proposed testing procedure. The simulation results indicate that the testing procedure has sufficient power to detect non-stationarity even for small sample sizes, as well as the capacity to distinguish whether the t-heterogeneity arises from the mean or the variance of the process.

The remainder of the paper is organized as follows. Section 2 motivates the idea behind testing the primary moments for t-invariance using heterogenous variants of the Normal, Autoregressive model. The need for alternative tests is stimulated in section 3 and in section 4 we provide a description of the suggested testing procedure which is based on the ME bootstrap and the idea of a rolling window estimator (RWE). The simulation design and results are presented in section 5 and in section 6 we apply the testing procedure to a number of macroeconomic series in order to assess its ability to detect a variety of forms of non-stationarity. We conclude by summarizing the main points, and indicating possible refinements and extensions.

2 Motivation

In this section we propose an alternative way of testing for t-invariance, that is based on the primary moments (marginal and joint moments) of the series involved. The rationale is that the model parameters are functions of the primary moments and departures from t-homogeneity in the latter are likely to be imparted onto the former. To see the relationship between the primary moments and the model parameters, consider the underlying parametrization of the Normal Autoregressive Model.

[A] **The Normal Autoregressive (AR(1)) Model**, takes the form**:**

$$y_t = a_0 + a_1 y_{t-1} + u_t, \qquad (u_t|\mathfrak{F}_{t-1}) \sim \mathsf{N}(0, \sigma^2), \tag{1}$$

where $\mathfrak{F}_{t-1} = \sigma(\mathbf{Y}^0_{t-1})$ is sigma field generated by the past history of y_t, $\mathbf{Y}^0_{t-1} := (y_{t-1}, y_{t-2}, ..., y_1)$. In this case the relevant reduction assumptions on the process $\{y_t, t \in \mathbb{T}\}$ that would give rise to model (1) are: (i) (D) Normal, (ii) (M) Markov and (iii) (H) Stationarity (see Spanos, 2001), where:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_0 & \sigma_1 \\ \sigma_1 & \sigma_0 \end{pmatrix}\right), \; t \in \mathbb{T}.$$

The relationship between the model parameters $\boldsymbol{\phi} := (a_0, a_1, \sigma^2)$ and the primary parameters $\boldsymbol{\psi} := (\mu, \sigma_0, \sigma_1)$ is:

$$a_0 = (1 - a_1)\,\mu_y, \quad a_1 = \tfrac{\sigma_1}{\sigma_0}, \quad \sigma^2 = (\sigma_0 - \tfrac{\sigma_1^2}{\sigma_0}) = \sigma_0(1 - a_1^2).$$

The complete specification for the AR(1) model is given in table 1.

Table 1: **Normal Autoregressive (AR(1)) model**

	Statistical GM:	$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \; t \in \mathbb{T}$,
[1]	Normality:	$f(y_t \mid \mathbf{Y}^0_{t-1}; \boldsymbol{\theta})$, for $\mathbf{Y}^0_{t-1} := (y_{t-1}, ..., y_1)$,
[2]	Linearity:	$E(y_t \mid \sigma(\mathbf{Y}^0_{t-1})) = \alpha_0 + \alpha_1 y_{t-1}$,
[3]	Homosked.:	$Var(y_t \mid \sigma(\mathbf{Y}^0_{t-1})) = \sigma_0$, free of $\mathbf{Y}^0_{t-1}$,
[4]	Markoveness:	$\{(y_t \mid \mathbf{Y}^0_{t-1}), \; t \in \mathbb{T}\}$ is a Markov process,
[5]	t-homogeneity:	$\boldsymbol{\theta} := (\alpha_0, \alpha_1, \sigma_0)$ are t-invariant $\forall t \in \mathbb{T}$.

A crucial distinction made in this specification is that of *heteroskedasticity* vs. *t-heterogeneity*; see Spanos (1986). Heteroskedasticity denotes the functional dependence of the conditional variance on the conditioning variables $\mathbf{Y}^0_{t-1}$, i.e. $Var(y_t \mid \sigma(\mathbf{Y}^0_{t-1})) = h(\mathbf{Y}^0_{t-1})$, as opposed to t-heterogeneity which refers to the functional dependence on the index t. i.e. $Var(y_t \mid \sigma(\mathbf{Y}^0_{t-1}))=\sigma(t)$. The distinction is important because the source of the departure in the two cases is different; the former can arise because of departures from the Normality assumption, but the latter arises when the process $\{y_t,\, t \in \mathbb{T}\}$ is non-stationary. Although, it is possible that $\boldsymbol{\psi}$ are time heterogeneous but $\boldsymbol{\theta}$ are t-invariant, it can only happen in very restrictive circumstances, as demonstrated below. Hence, detecting heterogeneity in the primary moments $\boldsymbol{\psi}$ will usually indicate departures from parameter t-invariance.

In order to illustrate how any form of heterogeneity in the primary moments is likely to give rise to departures from the t-invariance of the model parameters let us consider the following Heterogeneous Normal Autoregressive Model (see Spanos, 2001).

[B] The Non-Stationary, Normal Autoregressive Model:

$$y_t = a_0(t) + a_1(t)y_{t-1} + u_t, \qquad \left(u_t|\ \mathbf{Y}^0_{t-1}\right) \sim \mathsf{N}\left(0, \sigma^2(t)\right). \tag{2}$$

In this case the relevant reduction assumptions on the process $\{y_t,\, t \in \mathbb{T}\}$ that would give rise to (2) are: (D) Normal, and(M) Markov (see Spanos, 2001):

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \mu(t) \\ \mu(t{-}1) \end{pmatrix}, \begin{pmatrix} \sigma(t,t) & \sigma(t,t{-}1) \\ \sigma(t{-}1,t) & \sigma(t{-}1,t{-}1) \end{pmatrix}\right),\ t \in \mathbb{T}, \tag{3}$$

where the primary moments $\boldsymbol{\psi}(t) := (\mu(t), \sigma(t,t), \sigma(t,t{-}1))$ are allowed to be *arbitrary functions* of t. The relationship between the model parameters $\boldsymbol{\phi}(t) := (a_0(t), a_1(t), \sigma^2(t))$ and the primary parameters $\boldsymbol{\psi}(t)$ is:

$$a_0(t){=}\left(1{-}a_1(t)\right)\mu(t), \quad a_1(t){=}\tfrac{\sigma(t,t-1)}{\sigma(t-1,t-1)}, \quad \sigma^2(t){=}\sigma(t,t){-}\tfrac{[\sigma(t,t-1)]^2}{\sigma(t-1,t-1)}{=}\sigma(t,t)(1{-}a_1^2(t)).$$

These parameterizations indicate most clearly that any heterogeneity, either in the mean or the variance or both, is likely to give rise to t-heterogeneity in the parameters. To see this consider a very special case of (3) where $\mu(t)$ and $\sigma(t,s)$ take the simplest forms:

$$\mu(t) = \mu_0 \cdot t, \quad \sigma(t,s) = \sigma(|t-s|) \cdot \min(t,s),$$

known as separable heterogeneity; see Spanos (1999). In this case (3) reduces to:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \mu \cdot t \\ \mu \cdot (t{-}1) \end{pmatrix}, \begin{pmatrix} \sigma(0)\cdot t & \sigma(1)\cdot(t-1) \\ \sigma(1)\cdot(t-1) & \sigma(0)\cdot(t-1) \end{pmatrix}\right),\ t \in \mathbb{T}, \tag{4}$$

giving rise to an AR(1) model with t-heterogeneous parameters $a_0(t)$ and $\sigma^2(t)$:

$$a_0(t){=}\mu a_1 + \mu(1-a_1)t, \quad a_1{=}\tfrac{\sigma(1)}{\sigma(0)}, \quad \sigma^2(t) = \sigma(0)\left[a_1^2 + (1-a_1^2)t\right];$$

see Spanos (1990). The t-heterogeneity disappears only when $\sigma(0) = \sigma(1)$, which is the well-known unit root case ($a_1 = 1$), i.e. $\{y_t,\, t \in \mathbb{T}\}$ is a Wiener process. This example demonstrates clearly that unit root heterogeneity is only a very special form of moment heterogeneity. It's clear, however, that almost every other form of moment heterogeneity one can think of will give rise to model parameter t-heterogeneity. It can be shown that heterogeneous primary moments will give rise to t-homogenous model parameters only on a set of measure zero; see Maasoumi (2001).

The generic null and alternative hypotheses will be of the form:

$$H_0 : \boldsymbol{\kappa}(t,s) = \boldsymbol{\kappa}(|t-s|), \quad \text{vs.} \quad H_1 : \boldsymbol{\kappa}(t,s) \neq \boldsymbol{\kappa}(|t-s|),\ t,s \in \mathbb{T}, \tag{5}$$

where $\boldsymbol{\kappa}(t,s)$ denotes the first two moments of the model distribution, say $\boldsymbol{\kappa}(t,s) := (\mu(t), \sigma(t,t), \sigma(t,s)),\ t \neq s$. Confining the discussion to the first two moments is motivated partly by the fact that the higher moments and cumulants (in terms of which the model parameters are defined) are functions of the first two moments, and partly to keep the discussion of the different scenarios manageable. Extending the results to include higher moments is straightforward. Another crucial issue raised by hypotheses of the form (5) is that of the incidental parameter problem since evaluation under the alternative requires one to estimate parameters that change with t. The way to deal with this problem is to use a bootstrapping algorithm that enhances the sample information in appropriate ways that addresses the incidental parameter problem; the technique adapted for this purpose is the Maximum Entropy bootstrap proposed by Vinod (2004).

The perspective adopted in this paper is one of *Mis-Specification (M-S) testing* because the issue of whether the model parameters change with the index t, concerns probing outside the boundaries of the model. As such specifying particular forms of heterogeneity in H_1 is not particularly helpful because the rejection of the null in this context does *not* entitle one to deduce the alternative is valid; see Spanos (2000). This is a classic case of the *fallacy of rejection*: evidence against H_0 is (mis)interpreted as evidence for the alternative - see Mayo and Spanos (2004). Adoption of the alternative should be justified on its own merit on the basis that it gives rise to a statistically adequate model; see Spanos (2000). That is, once some form of non-stationarity is detected one can proceed to determine its form and nature as well as provide a structural interpretation when appropriate.

2.1 Andrews & Ploberger tests

In an attempt to motive the need for alternative testing procedures which can detect smooth changing t-heterogeneity in the parameters we assess the capacity of the Andrews & Ploberger (1994) tests to detect smoothly changing mean and variance trends using Monte Carlo experiments. While these tests were originally designed for abrupt model parameter shifts, the lack of alternative testing procedures for other forms of heterogeneity, made this popular for testing any form of parameter invariance. We will assess the ability of these tests to detect departures from homogeneity using several functional forms for $\mu(t)$ and $\sigma^2(t)$ and we report the actual rejection rates at significance level $\alpha = 5\%$. For each scenario we use a sample of size $n = 100$ and N=$10,000$ replications; the results concerning the actual power of these tests are summarized in tables 3-4. For the details on these choices and additional scenarios, see Koutris (2005)

The Andrews & Ploberger (1994) statistics are based on Quandt's idea which evaluates the Chow statistic at every possible breakpoint. This is equivalent to the statistic SupF $= sup_t F_t$, where the supremum of the Chow statistic is taken over the time t. They developed the exponentially weighted Wald statistic ExpF=ln $\int \exp\left(\frac{F_t}{2}\right) dw(t)$ and the average F test AveF=$\int_t F_t dw(t)$ where w is a measuring putting weight $\frac{1}{t_2-t_1}$ on each integer t in the interval $[t_1, t_2]$, and showed that these are optimal against distant and very local alternatives, respectively. The simulation results in Table 2a (see also Table 2b in Appendix A) indicate that the p-value approximations proposed by Hansen (1997) show a systematic upward discrepancy from the nominal type I error. Therefore, tables 3 and 4 report the size-corrected empirical power based on the percentiles of the empirical distribution of the A&P statistics, evaluated for n=100 observations under the null.

Table 2a: Empirical type I error (α=5%) based on R=10,000

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	8.38	14.95	14.04	13.02
ExpF	2.61	14.64	12.01	11.29
AveF	2.01	10.71	10.49	12.75

It reports the percentage of false rejections in R=10,000 experimental trials.

Ideally this should be equal to a=5%

Table 3: Empirical, size-corrected Power of A&P tests under mean trend; α=5%

Trend Function	SupF	ExpF	AveF
$\mu(t) = \mu + 0.02t$	21.82	24.94	23.14
$\mu(t) = \mu + 0.001t + 5(10^{-4}t^2)$	7.61	9.61	9.29
$\mu(t) = \exp(0.01t) + \mu$	18.3	18.98	18.27
$\mu(t) = \left(\frac{5}{1+\exp\left(\frac{-t}{4}\right)}\right) + \mu$	7.45	8.21	6.98

It reports the percentage of correct rejections in R=10,000 experimental trials.

Ideally this should be as close to 100% as possible

The results in tables 2 - 4 (see also Koutris, 2005) indicate that the A&P statistics have very low power in detecting smoothly trending t-heterogeneity in the parameters. The results are not surprising because these tests were designed to detect abrupt parameter shifts. This, however, raises the need for more effective testing procedures under more general forms of heterogeneity.

Table 4: Empirical, size-corrected Power of A&P tests under variance trend; α=5%			
Trend Function	SupF	ExpF	AveF
$\sigma^2(t)=\sigma^2+0.05\cdot t$	10.78	3.94	9.95
$\sigma^2(t)=\sigma^2+0.03\cdot t+0.01\cdot t^2$	21.53	4.65	19.58
$\sigma^2(t)=\sigma^2+\exp(0.02\cdot t)$	14.87	5.94	14.1

It reports the proportion of correct rejections in R=10,000 experimental trials.

Ideally this should be as close to 100% as possible

3 Testing for non-stationarity using resampling

In this section we investigate the t-invariance of the primary moments by using the idea of rolling window estimator and applying the Maximum Entropy bootstrap by Vinod (2004). The choice of the rolling window (overlapping and non-overlapping) estimator is motivated by the fact that stationarity implies constancy across window estimates. Hence, by testing the non-constancy of such estimators we propose tests for stationarity.

3.1 Rolling Window Estimator

According to Banerjee, Lumsdaine and Stock (1992) the term 'recursive estimator' is due to Brown et al. (1975). The notion of a rolling or fixed-window (see Spanos, 1986, p. 562) estimator dates back to early statistical quality control literature; see Shewhart (1939).

Definition 1 *Let $\{R_t\}_{t=1,..,n}$ be a random process, and θ be the unknown parameter to be estimated and $\hat{\theta} = g(\mathbf{R})$ be an estimator based on the process. Furthermore, let $P_R = \{P_{R_i}\}_{i\in I}$ be a partition of the process, such that:*

$$P_{R_{t_i}}=\{R_t{:}t\in[t_i,t_i-1+l]\},\ t_i=1,2,...,n-(l-1), \tag{6}$$

where l is the fixed window size. The rolling estimator $\hat{\theta}_{r_{t_i}}$ of the unknown parameter θ is defined as:

$$\hat{\theta}_{rti} = g(P_{R_{t_i}}) \text{ for } t_i = 1,2,...,n-(l-1). \tag{7}$$

The rolling (window) estimators are based on a changing subsample of fixed length l that moves sequentially through the sample, giving rise to a series of estimates for θ.

The first weakness of the fixed window estimators, is that they require a large sample size. The second problem when using a rolling estimator is the trade off between the window size l and the number of rolling window estimates. Even though a large window size could yield a more precise estimate of the unknown population parameter θ, this may not be a representative of the θ especially in the presence of heterogeneity. In this paper we use a rolling window of small size and we apply resampling techniques to each window so as to estimate θ with higher precision. There are still two problems with this strategy. First, traditional resampling techniques require large sample sizes. Secondly, departures from the IID assumption affect the performance of the bootstrap methods (see Spanos and Kourtellos, 2002). In an effort to overcome some of these problems we apply the Maximum Entropy bootstrap of Vinod (2004) which is reliable for small sample sizes and is designed to be robust to deviations from the IID assumption.

3.2 Maximum Entropy Bootstrap

The Maximum Entropy (ME) bootstrapping procedure proposed by Vinod (2004) is an essential component of our procedure. It provides a reliable resampling algorithm for short non-stationary time series. The ME bootstrap is similar to Efron's traditional bootstrap but avoids the three restrictions which make the traditional bootstrap unsuitable for economic and financial time series data. To explain these three restrictions consider a time series x_t over the range $t = 1, \ldots, T$. The traditional bootstrap sample repeats some x_t values and requires that none of the resampled values can differ from the observed ones. It also requires the bootstrap resamples to lie in the interval $[\min(x_t), \max(x_t)]$. These two conditions are quite restrictive in practice. The third restriction arises because the bootstrap resample shuffles x_t in such a way that all dependence and heterogeneity information in the time series sequence is lost. This is particularly crucial for testing heterogeneity because any reordering will distort the information contained in the data. To address these issues the traditional literature has made attempts to remove one or two of these restrictions but not all three. For example the 'smooth bootstrap is supposed to be able to avoid the second restriction while 'block resampling' is designed to avoid destroying the dependence information (see Berkowitz and Killian, 2000).

The ME bootstrap is more appealing because it simultaneously avoids all three problems. Moreover, the bootstrap algorithm is based on the ME density and satisfies the ergodic theorem, Doob's theorem and almost sure convergence of sampling distributions of pivotal statistics without assuming stationarity. In particular, the ME density $f(x)$ is chosen so as to maximize $H = E(-\log f(x))$ (Shannon's information), subject to certain *mass-preserving* and *mean preserving* constraints; see Vinod (2004) for the details.

Using the idea of maximum entropy density, the ME algorithm to generate multiple ensembles of stochastic process realization is specified in the following steps:

Step 1: Define a $T \times 2$ sorting matrix called S_1. In the first column place the observed time series x_t while in the second column place the index set $I_{ndx} = \{1, 2, ..., T\}$.

Step 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 I_{ord} of sorted I_{ndx} in the second column to be used later. Then compute 'intermediate points' z_t as averages of successive order statistics as follows:

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

and construct the intervals I_t defined on z_t and m_t with specific weights on the order statistics $x_{(t)}$ defined in the equations shown below:

$$a.\quad f(x) = \frac{1}{m_1}\exp\left(\frac{[x-z_1]}{m_1}\right), x \in I_1,\ m_1{=}\frac{3x_{(1)}}{4} + \frac{x_{(2)}}{4}.$$

$$b.\quad f(x) = \frac{1}{z_k-z_{k-1}}, x \in (z_k, z_{k+1}],\ \text{with mean } m_k :$$
$$m_k = \frac{x_{(k-1)}}{4} + \frac{x_{(k)}}{2} + \frac{x_{k+1}}{4} \text{ for } k = 1, 2, .., T-1.$$

$$c.\quad f(x) = \frac{1}{m_T}\exp\left(\frac{[z_{T-1}-x]}{m_T}\right),\ x \in I_T,\ m_T{=}\frac{x_{(T-1)}}{4}{+}\frac{3x_T}{4}.$$

Step 3: Choose a seed, create T uniform pseudorandom numbers p_j in the $[0, 1]$ interval, and identify the range $R_t = \left(\frac{t}{T}, \frac{t+1}{T}\right]$ for $t = 0, \ldots T-1$ wherein each p_j falls.

Step 4: Match each R_t with I_t by using the following equations:

$$x_{j,t,me} = z_{T-1} - |\theta| \ln(1 - p_j) \text{ if } p_j \in R_0\ ,$$

$$x_{j,t,me} = z_1 - |\theta| \left|\ln(1 - p_j)\right| \text{ if } p_j \in R_{T-1}$$

or as a linear interpolation and obtain a set of T values $\{x_{j,t}\}$ as the j-th resample. Here θ is the mean of the standard exponential distribution. Make sure that the mean of the uniform for each interval equals the correct mean m_t by using add factors (see Vinod, 2004 Remark 4, for more details).

Step 5: Define another $T \times 2$ sorting matrix S_2. Reorder the T members of the set $\{x_{j,t}\}$ 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 I_{ord} of step 2 in column 2 of S_2.

Step 6: Sort S_2 matrix with respect to the second column to restore the order $\{1, 2, .., T\}$ there. The jointly sorted column 1 of elements is denoted by $\{x_{s,j,t}\}$, where s reminds us of the sorting step.

Step 7: Repeat steps 1 to 6 a large number of times for $j = 1, 2, ..., J$.[1]

3.3 Description of the Testing Procedure

The primary objective of this paper is to develop a procedure for detecting departures from the t-invariance of the first two moments of the stochastic vector $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$.

[1]For our procedure J = 100

Second order stationarity allows us to infer that the model parameters based on these moments will also be t-invariant. This procedure is based on a rolling window estimator. For each individual window we then use the ME bootstrap method to replicate the series and create an ensemble in order to extract all the systematic statistical information present in the data. This method allows us to efficiently estimate the moments of the series based on multiple sets of realizations. Note that by choosing a sufficiently small time window (i.e. fewer than 10 observations) and by focusing on smooth functions of the time trend we can safely assume local homogeneity. In this way we create a sequence of estimates for the first two moments of each random variable.

Using the sequence of ME resampled replicas we formulate an F-type test for the hypothesis of constant moments over time. This F-statistic is based on the residuals from a restricted model and an unrestricted model for each sequence of estimates. The restricted model assumes moment constancy over time. It is formed by using the AR(1) specification for the estimated mean or variance since by construction these sequences form a Markov(1) process. In the unrestricted model we allow for time heterogeneity of a general form by adding a Bernstein polynomial of a specific degree to the AR(1) model. The functional form of the Bernstein polynomial is:

$$B_{k,t_i} = \sum_{j=0}^{k} \beta_j \left(\begin{array}{c} k \\ j \end{array} \right) t_i^j (1-t_i)^{k-j},\ k \in \mathbb{N},\ 0 < t_i < 1, \tag{8}$$

where $\mathbb{N} = \{1, 2, .., \}$, and $\{\beta_j\}_{j=1,2,..,k}$ denote unknown constants. The Bernstein polynomials 8 form an orthogonal basis for the power polynomials of degree less than of equal to w for any $w \geq 1$; Lorentz (1986). The orthogonality of these polynomials also has practical implications - it allows us to use a high degree polynomial without the problem of near-multicolinearity. Another important attribute of these polynomials is that they also provide good approximations for a variety of trend functions that are present in real economic series.

The F-type test implemented leads to inference about the presence or not of time trend in the moments of the processes. If we fail to reject the hypothesis of time invariance, the sufficiency of this assumption allows us to conclude that the model parameters based on these variable will also be t-invariant.

The proposed testing procedure can be described in the following 7 steps:

1. We begin by investigating each individual variable for time invariance. We first determine the appropriate window size l. Based on our simulations we find that a rule of thumb to chose the window size is: $l = \left[\frac{n}{10}\right] - 2$, for sample sizes $n \leq 150$. Note that for larger sample sizes one should consider non-overlapping rolling window estimates; see Koutris (2005). It is important to stress that in practice one should vary the window size in an attempt to establish the robustness of the results with respect to the choice of the window size. In the simulations that follow the robustness to such small changes was confirmed.

2. For each window of size l we generate an additional number of Vinod bootstrap ($VB = 100$) samples denoted by VB.

3. Using the total number of observations available to us after bootstrapping, which amount to $(TVB) = l \times (VB + 1)$, we estimate the sample mean and variance for each window. This gives rise to a sequence of $T = n-(l-1)$ sample means, $\hat{\mu}(t_i)$ and variance estimates, $\hat{\sigma}^2(t_i)$.

4. The assumption of time invariance of the moments, implies that these sequences should have a constant mean and variance over time. We first check for time invariance of the mean. The null hypothesis of our test in this case is: $H_0 : \mu(t_i) = \mu$ for $t_i = 1 \ldots n-(l-1)$. Since we have overlapping windows the constructed sequences exhibit first order Markov dependence and we use an AR(1) specification for the *restricted formulation*:

$$\hat{\mu}(t_i) = a_o + a_1 \cdot \hat{\mu}(t_i-1) + u_{r\mu}(t_i), \tag{9}$$

where α_0, a_1, are the unknown model parameters to be estimated, and $u_{r\mu}$ are NIID white noise errors. From this model we estimate the Restricted Sum of Squared Residuals (RSSR) to be used in formulating the F-statistic.

5. In order to test for time trend in the moments of the series, we extend the above AR(1) specification to incorporate a time trend of polynomial form. We use Bernstein orthogonal polynomials of sufficiently high degree, so that we can approximate various smooth trend functions. The alternative hypothesis in this case is: $H_1 : \mu(t_i) \neq \mu$ for any $t_i = 1 \ldots n-(l-1)$. We thus estimate the *unrestricted formulation* for the mean:

$$\hat{\mu}(t_i) = a_o' + a_1' \cdot \hat{\mu}(t_i-1) + B_{k,t_i} + u_{u\mu}(t_i), \tag{10}$$

and evaluate the Unrestricted Sum of Squared Residuals (USSR) to be used in the F-test statistic, where $B_{k,t}$ is the $k-$th degree Bernstein Orthogonal polynomial at time t and $u_{u\mu}$ is NIID errors.

6. We then calculate the F-statistic based on the RSSR_μ and the USSR_μ and adjusted for the appropriate degrees of freedom $(T-(k+2), k)$.

In a similar way we postulate, respectively, the restricted and unrestricted formulations for the variance:

$$\hat{\sigma}^2(t_i) = c_0 + c_1 \cdot \hat{\sigma}^2(t_i-1) + u_{r\sigma^2}(t_i), \tag{11}$$

$$\hat{\sigma}^2(t_i) = c_0' + c_1' \cdot \hat{\sigma}^2(t_i-1) + B_{k,t_i}' + u_{u\sigma^2}(t_i), \tag{12}$$

where $u_{r\sigma^2}$, $u_{u\sigma^2}$ are NIID errors. The estimation of (11)-(12) gives rise to the RSSR_{σ^2} and the USSR_{σ^2}, respectively, which form the F-statistic for testing $H_0 : \hat{\sigma}^2(t_i) = \hat{\sigma}^2$ for $t_i = 1 \ldots n - (l - 1)$ against $H_1 : \hat{\sigma}^2(t_i) \neq \hat{\sigma}^2$ for any $t_i = 1 \ldots n - (l - 1)$.

7. We repeat the same procedure for all the relevant variables in our model.

The absence of t-heterogeneity in the moments, after thorough probing, is interpreted as evidence of its absence which, in turn, provides support for the t-invariance of the model parameters. On the other hand, the presence of t-heterogeneity in the moments calls for further testing and respecification of the statistical model. Respecification is a different aspect of empirical modeling which we do not discuss explicitly in this paper; see Spanos (2000, 2001).

4 Simulation design and results

To evaluate the proposed testing procedure we perform a number of Monte Carlo experiments on the basis of which we assess both their size and local power. In these experiments we simulate a variety of departures from the assumption of stationarity of the moments. All experimental results reported are based on 10,000 replications of sample sizes $n = 60$, $n = 80$ and $n = 100$. We have chosen these sample sizes to illustrate the fact that the proposed procedure performs reasonably well even for small sample sizes. Furthermore we report the percentage of rejection for three different levels of significance $(0.01, 0.05$ and $0.10)$.

To ensure the correct actual size of the proposed testing procedure we relate the choice of the appropriate window length to the .01, .05 and .10 quantiles of the empirical distribution. In Table 5 (see Appendix A) we report simulation results concerning the appropriate window length for different sample sizes; see Koutris (2005) for further details. The 'appropriate' window length for sample of $n = 60$ appears to be $l = 5$, for sample size $n = 80$ it is $l = 6$ and for sample size of $n = 100$ it is $l = 8$. For these window sizes the estimated actual type I error seems to be reasonably close to the nominal for the three different levels of significance considered.

4.1 Monte Carlo Experiments

In this section we describe the simulation design for our experiments. First we generate $R = 10{,}000$ samples of size n of the process $\{u_{t_i}, t_i = 1, 2, .., n\}$:

$$(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, .., \mathbf{u}^{(R)}),$$

where each $\mathbf{u}^{(r)}$, $r = 1, 2, ..., R$ represents a vector of n pseudo-random numbers from $\mathsf{N}(0, 1)$. By 'feeding' sequentially each $\mathbf{u}^{(r)}$ into the statistical generating mechanism:

$$y_{t_i} = \mu(t_i) + \sigma(t_i)u_{t_i}$$

we simulate the artificial data realizations. We then introduce a number of different departures from the assumption of moment time homogeneity by considering different functional forms of $\mu(t_i)$ and $\sigma(t_i)$. The simulations and empirical analysis are performed using the GAUSS programming language.

4.1.1 Experiment 1: Smooth Mean Trend

In experiment 1 we generate data with four different time trends in the mean of the series. The functional forms of these trends are shown below:

1. Linear: $\mu(t_i) = \mu + 0.02t_i, \quad \sigma^2(t_i) = 1$
2. Quadratic: $\mu(t_i) = \mu + 10^{-3}t_i + 5 \cdot 10^{-4}t_i^2, \sigma^2(t_i) = 1$
3. Exponential: $\mu(t_i) = \exp(10^{-2}t_i) + \mu,\ \sigma^2(t_i) = 1$
4. Logistic: $\mu(t_i) = (\frac{5}{1+\exp(\frac{-t_i}{4})}) + \mu,\ \sigma^2(t_i) = 1$

4.1.2 Experiment 2: Smooth Variance Trend

Experiment 2 is designed to generate data series that exhibit three forms of variance trends shown below:

1. Linear: $\mu(t_i) = 0, \quad \sigma^2(t_i) = \sigma^2 + 0.05t_i$
2. Quadratic: $\mu(t_i) = 0,\ \sigma^2(t_i) = \sigma^2 + .03t_i + .01t_i^2$
3. Exponential: $\mu(t_i) = 0,\ \sigma^2(t_i) = \sigma^2 + \exp(0.02 \cdot t_i)$

4.1.3 Experiment 3: Single Mean Break

The main purpose of experiment 3 is to examine the extent to which our testing procedure can detect single breaks − mean shifts− introduced at three different locations of the sample. We introduce a single mean shift of size two standard deviations at the first (Q_1), second (Q_2) and third (Q_3) quarter of the sample.

1. Single Mean Break at Q_1: $\mu(t_i) = \mu + 2\sigma(t_i) I_{\{t_i \geq \frac{n}{4}\}}, \quad \sigma^2(t_i) = 1.$
2. Single Mean Break at Q_2: $\mu(t_i) = \mu + 2\sigma(t_i) I_{\{t_i \geq \frac{n}{2}\}}, \quad \sigma^2(t_i) = 1.$
3. Single Mean Break at Q_3: $\mu(t_i) = \mu + 2\sigma(t_i) I_{\{t_i \geq \frac{3n}{4}\}}, \quad \sigma^2(t_i) = 1.$

4.1.4 Experiment 4: Single Variance Break

Experiment 4 is similar to previous one in that it introduces single breaks at the first (Q_1), second (Q_2) and third (Q_3) quarter of the sample. In this experiment we introduce a variance shift of two standard deviations.

1. Single variance Break at Q_1: $\mu(t_i) = \mu, \quad \sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{n}{4}\}}$
2. Single variance Break at Q_2: $\mu(t_i) = \mu, \quad \sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{n}{2}\}}$
3. Single variance Break at Q_3: $\mu(t_i) = \mu, \quad \sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{3n}{4}\}}$

4.2 Monte Carlo Results

Tables 6 through 9 present the results of the Monte Carlo simulations. In Table 6 we present the rejection frequencies of the test when the mean has four different types of heterogeneity: linear, quadratic, exponential and logistic function of time t. The power of the test is reasonably high for all scenarios, and it increases with sample size. Moreover, the actual size of the test for σ^2 constant is close to the nominal. In addition the test seems to have remarkably high power against the alternative of a linear or a quadratic time-trend even in a sample of $n = 60$ observations. The simulated power for the quadratic trend is 66.5%, 99.50% and 100% when $n = 60, 80, 100$, respectively. We conclude that the closer is the functional form of the time trend to the polynomial family, the better is the performance of the test. On the other hand, the exponential function is less detectable and requires a sample size of n equal to 80 or greater in order to have satisfactory power. Overall we see from Table 5 that even for weak mean time trends the performance of the test is promising.

Table 6: Trending Mean

Trend Function	$\alpha\%$	$H_0 : \mu$ constant			$H_0 : \sigma^2$ constant		
		n=60	n=80	n=100	n=60	n=80	n=100
Linear trend $\mu(t_i)=\mu+0.02\cdot t_i$	1	11.28	30.85	60.47	2.13	1.69	2.23
	5	27.89	57.99	83.86	6.06	5.13	6.35
	10	40.22	72.03	91.86	10.05	8.98	10.86
Quadratic trend $\mu(t_i)=\mu+10^{-3}\cdot t_i+5\cdot 10^{-4}\cdot t_i^2$	1	18.93	81.41	99.86	1.36	1.75	2.34
	5	48.15	97.35	100	4.67	5.20	6.46
	10	66.52	99.50	100	7.87	8.97	10.55
Exponential trend $\mu(t_i)=\exp(10^{-2}t_i)+\mu$	1	5.06	15.18	40.63	2.17	1.69	2.24
	5	15.05	35.31	67.28	6.30	5.11	6.44
	10	24.15	49.75	80.30	10.47	8.96	10.86
Logistic trend $\mu(t_i)=\left(\frac{5}{1+\exp\left(\frac{-t_i}{4}\right)}\right)+\mu$	1	34.50	31.42	54.05	2.39	2.01	3.03
	5	56.52	50.25	61.89	7.02	6.43	7.42
	10	65.06	59.47	66.98	12.15	10.41	13.53

It reports the rejection percentage for the null hypothesis in R=10,000 experimental trials, for each scenario.
Ideally for "μ constant" should be as close to 100 as possible and for "σ^2 constant" close to a%

The results of the proposed testing procedure in the presence of variance heterogeneity are summarized in Table 7. The power is again reasonably high for the polynomial time trends but for the exponential trend the testing procedure requires larger sample sizes to perform well. Another interesting point to note is the ability of the procedure to correctly distinguish between a smooth mean trend and a trend in the variance. It is noticeable that there are some size distortions in testing for μ constant.

Table 7: Trending Variance							
Trend Function	$a\%$	$H_0 : \mu$ constant			$H_0 : \sigma^2$ constant		
		n=60	n=80	n=100	n=60	n=80	n=100
Linear trend $\sigma^2(t_i){=}\sigma^2{+}0.05{\cdot}t_i$	1	1.79	2.91	4.99	13.57	24.83	36.62
	5	7.29	10.29	14.12	30.02	48.70	62.43
	10	13.19	16.77	22.43	41.87	62.49	75.27
Quadratic trend $\sigma^2(t_i){=}\sigma^2{+}0.03{\cdot}t_i{+}0.01{\cdot}t_i^2$	1	8.53	14.72	20.36	44.68	52.28	60.40
	5	28.37	30.45	38.81	67.35	73.38	79.79
	10	40.14	40.47	49.38	76.92	81.97	87.33
Exponential trend $\sigma^2(t_i){=}\sigma^2{+}\exp(0.02{\cdot}t_i)$	1	2.53	3.31	7.34	10.41	24.37	46.35
	5	9.29	10.54	19.07	23.81	45.86	69.07
	10	15.42	17.07	28.14	33.14	57.59	78.80
It reports the rejection percentage for the null hypothesis in R=10,000 experimental trials, for each scenario. Ideally for "μ constant" should be close to a% and for "σ^2 constant" as close to 100% as possible							

Finally we report Monte Carlo results which illustrate the ability of our test to detect single, discrete breaks in the mean or variance. The simulations are performed by considering a break of two standard deviations introduced at the first, second and third quarter of the sample.

Table 8: Single Mean Break							
Mean Break	$\alpha\%$	$H_0 : \mu$ constant			$H_0 : \sigma^2$ constant		
		n=60	n=80	n=100	n=60	n=80	n=100
$\mu(t_i){=}\mu{+}2\sigma{\cdot}I_{\left\{t_i\geq\frac{n}{4}\right\}}$	1	79.90	90.40	96.42	1.30	1.42	1.49
	5	93.91	98.74	99.35	5.10	4.05	4.10
	10	97.00	99.72	99.89	7.80	6.68	7.21
$\mu(t_i){=}\mu{+}2\sigma{\cdot}I_{\left\{t_i\geq\frac{n}{2}\right\}}$	1	16.60	26.89	34.14	0.79	0.83	1.02
	5	50.11	63.61	72.17	3.09	3.46	3.46
	10	71.41	82.65	88.17	5.71	5.73	6.54
$\mu(t_i){=}\mu{+}2\sigma{\cdot}I_{\left\{t_i\geq\frac{3n}{4}\right\}}$	1	6.33	4.99	5.48	1.24	1.12	1.56
	5	23.11	21.62	21.77	4.45	3.79	4.00
	10	39.76	39.50	40.47	7.65	6.48	7.16
It reports the rejection percentage for the null hypothesis in R=10,000 experimental trials, for each scenario. Ideally for "μ constant" should be as close to 100 as possible and for "σ^2 constant" close to a%							

The results reported in Tables 8 and 9 suggest that even though the test is designed to detect smooth trends, it is also effective in detecting mean and/or variance shifts in the series. As before the power of the test increases with the sample size n; the test has higher power when the break is introduced earlier in the sample.

Table 9: Single Variance Break							
Variance Break	$a\%$	$H_0 : \mu$ constant			$H_0 : \sigma^2$ constant		
		n=60	n=80	n=100	n=60	n=80	n=100
$\sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{n}{4}\}}$	1	1.78	1.10	2.13	4.81	6.88	15.35
	5	7.49	5.61	7.74	17.83	25.09	43.27
	10	14.31	10.21	14.37	31.31	41.66	61.82
$\sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{n}{2}\}}$	1	3.42	2.78	4.10	8.10	9.24	11.78
	5	11.66	9.52	12.24	24.03	28.18	33.50
	10	19.76	15.31	19.54	37.33	43.56	51.85
$\sigma^2(t_i) = \sigma^2 + 2\sigma^2 I_{\{t_i \geq \frac{3n}{4}\}}$	1	8.66	6.84	9.50	23.30	23.27	23.54
	5	20.49	17.02	21.56	36.82	36.68	37.34
	10	28.85	24.56	29.81	44.22	44.97	46.57

It reports the rejection percentage for the null hypothesis in R=10,000 experimental trials, for each scenario.
Ideally for "μ constant" should be close to a% and for "σ^2 constant" as close to 100% as possible

5 Empirical Illustration

The proposed resampling test procedure is applied to a number of macroeconomic series in order to assess its ability to detect a variety of forms of non-stationarity. The empirical analysis intents to complement the simulation evidence in assessing the effectiveness of the proposed testing procedure.

The empirical investigation of these macro-series is based on the AR(p) specification which is a simple extension of the AR(1) model given in table 1, i.e. $E(y_t \mid \sigma(\mathbf{Y}_{t-1}^0)) = \alpha_0 + \sum_{k=1}^{p} \alpha_k y_{t-k}$. The focus of the empirical modeling is the stationarity of the first two moments of the process $\{y_t,\ t \in \mathbb{T}\}$. The proposed testing procedure is applied to the 'de-memorized' series which result from estimating an adequate AR(p) model and taking the residuals. The choice of p is based exclusively on statistical adequacy grounds, ensuring that the residuals from the AR(p) do not have any 'lingering' temporal dependence. We do not use Akaike-type information criteria for the choice of p because when the estimated model is misspecified such procedures can give rise to very misleading inferences; see Andreou and Spanos (2003). For comparison purposes we apply a variety of other structural change tests proposed in the literature. In particular we apply the SupF, ExpF, AveF test statistics proposed by Andrews and Ploberger (1994) and the approximations of these tests proposed by Hansen (2001).

5.1 Data

We consider four macroeconomic series exhibiting a variety of forms of non-stationarity in the mean and/or the variance:

(i) Quarterly Yen/US dollar Exchange Rate[1] (see fig. 1(a)) over the period 1982Q2-2005Q2,

(ii) Monthly European Total Turnover Index[2] (see fig. 2(a)) over the period 01/1995 - 08/2004,

(iii) Annual US Industrial Production Index[1] (see fig. 3(a)) over the period 1921-2004,

(iv) Quarterly US Investment[3] (see fig. 4(a)) over the period 1963Q2-1982Q4.

For the de-memorized US/Japan Exchange Rate series in fig. 1(b) we can clearly discern some variance heterogeneity; the variance of the series appears to be decreasing over time, while its mean seems to be stable over time. On the other hand the de-memorized EU Total Turnover Index series (see fig. 2(b)) seems variance stationary around a trending mean. The Industrial Production series (see fig. 3(b)) seems to have constant mean but a non-constant variance. Finally, the US Investment series (see fig. 4(b)) seems to exhibit both mean and variance heterogeneity. Note that all the above are educated conjectures based on eyeballing the time plots of the series. In order to assess the heterogeneity characteristics of the series we need to formally test these conjectures.

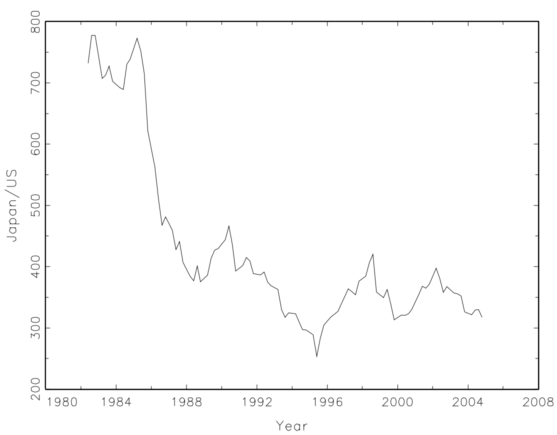

Fig. 1(a): US/Japan Exchange

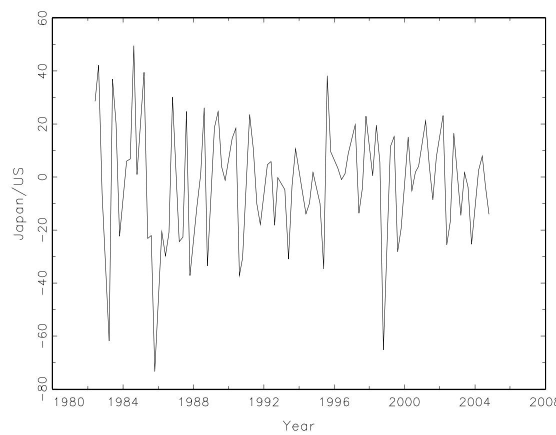

Fig. 1(b): Dem US/Jp Exchange

---

[1] Obtained from the St. Louis Reserve Federal Bank database

[2] Obtained from the Monthly Bulletin of the European Central Bank.

[3] Obtained from the Bureau of Economic Analysis.

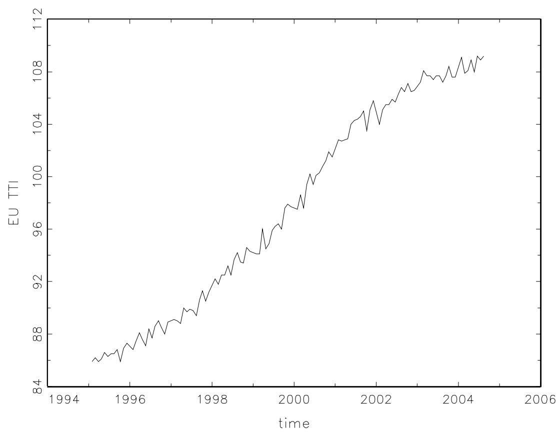

Fig. 2(a): EU TT Index

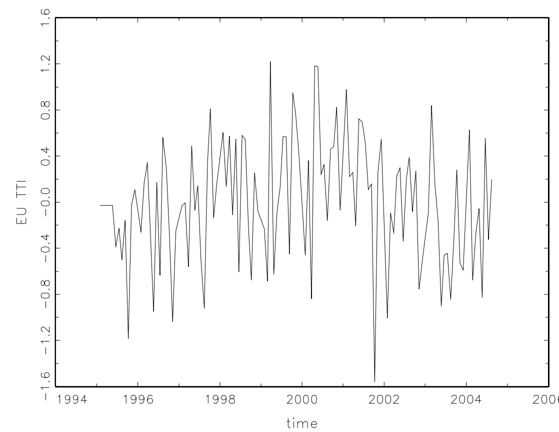

Fig. 2(b): Dem EU TTI

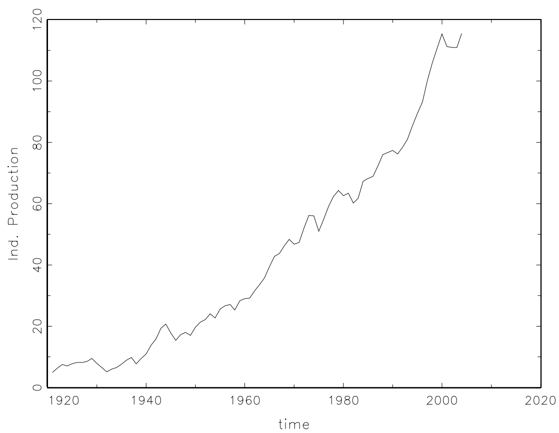

Fig. 3(a): US Ind Prod.

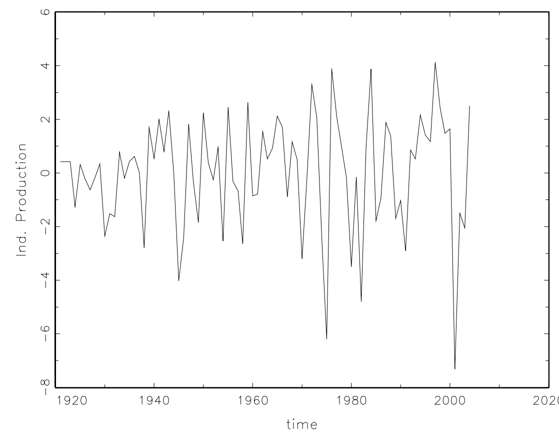

Fig. 3(b): Demem Ind Prod.

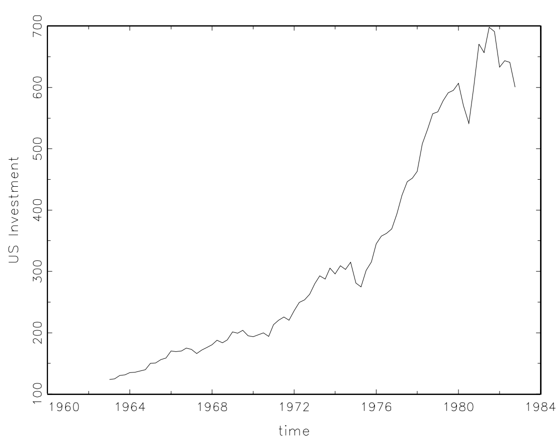

Fig. 4(a): US Invest.

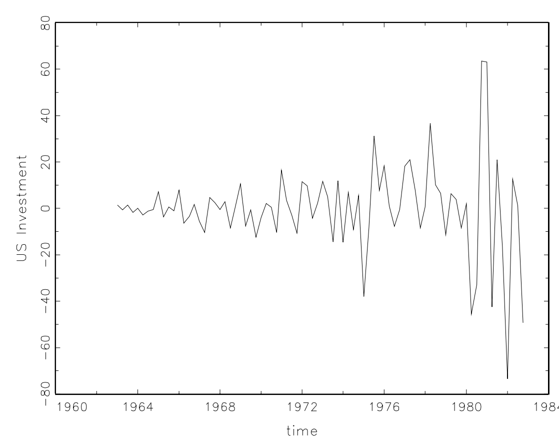

Fig. 4(b): Dem Invest

5.2 Empirical Results

Using statistical adequacy in choosing the appropriate lag length, for the US/Japan Exchange Rate, the US Industrial Production and the US Investment series $p = 2$, whereas for the EU/TTI series $p = 3$. On the de-memorized series we apply the proposed testing procedure based on a Rolling Overlapping Window Estimator (ROWE); see table 10.

Table 10: Empirical Results

Variable	Sample Size	l	$H_0 : \mu$ constant	$H_0 : \sigma^2$ constant
US/Japan Exchange	91	7	0.403 (0.806)	3.495 (0.011)**
EU Total Turnover	116	10	2.190 (0.075)*	0.499 (0.737)
US Ind Production	84	7	0.114 (0.977)	2.716 (0.036)**
US Investment	80	6	3.698 (0.009)***	1.864 (0.127)

Notes:
1. Entries are test statistics with p-values in parentheses
2. l is the rolling window length
3. (*), (**) , (***) refer to the rejection of the null hypothesis at 10%, 5% and 1% level of significance, respectively.

The testing results given in Table 8, indicate that the US/Japan Exchange Rate series is variance heterogeneous, while we have no evidence against the mean homogeneity assumption. For the EU TTI series our test indicates the presence of some mean heterogeneity. For the US Industrial Production series the test indicates no evidence against mean homogeneity, but it rejects variance homogeneity. Finally, the testing procedure clearly rejects the mean homogeneity assumption for the US Investment series while it doesn't provide much of support for the variance homogeneity.

For comparison purposes we apply the Andrews and Ploberger (A&P) (1994) statistics and their approximations provided by Hansen (2000) to the same AR(p) models estimated above. The results from the Andrews & Ploberger testing procedures are given in Tables 11 through 14; see appendix B. The main conclusion is that the A&P statistics did not detect any parameter heterogeneity in the four macroeconomic series; the only hint that there might be something worth investigating further was given for the EU Total Turnover Index series. This result is not very surprising because the A&P statistics are designed to capture a particular form of non-stationarity, structural breaks.

6 Conclusion

A new misspecification testing procedure for assessing the presence of non-stationarity in the primary moments of a stochastic process is proposed. Motivated by the fact that model parameters are always functions of the underlying primary parameters, the proposed test is based on rolling (overlapping) window estimates of the means and the variances of the series involved. To be able to enhance the systematic information contained in each window of observations we use the maximum entropy bootstrap as an appropriate form of resampling in this context. The rationale for our perspective is provided by the fact that one needs to establish the presence of non-stationarity in the primary moments, before proceeding to establish its form; the latter cannot be established using a misspecification test. The effectiveness of the proposed procedure is assessed using Monte Carlo simulations and empirical examples of actual time series data. The Monte Carlo simulations and the empirical examples indicate that the proposed testing procedure has the capacity to detect non-stationarity even for small samples, and is able to distinguish between mean or/and variance non-stationarity. Although the proposed testing procedure is based on general forms of non-stationarity, it is shown to have good power against abrupt changes in the underlying moments. This testing procedure can be used in conjunction with traditional tests to explore a broader variety of possible departures from the t-homogeneity assumption.

References

[1] Andreou, E. and A. Spanos (2003), “Statistical adequacy and the testing of trend versus difference stationarity”, *Econometric Reviews*, 22, 217-252.

[2] Andrews, W.K., (1993), “Tests for Parameter Instability and Structural Change With Unknown Change Point”, *Econometrica*, 61, 821-856. (Corrigendum, 71, 395-397)

[3] Andrews, W.K., Ploberger W., (1994), “Optimal Tests when a Nuisance Parameter is Present only Under the Alternative”, *Econometrica*, 62, 1383-1414.

[4] Andrews, W.K., Lee, I., Ploberger, W., (1996), “Optimal change point tests for normal linear regression”, *Journal of Econometrics*, 70, 9-38.

[5] Bai, J., Perron P., (1998), “Estimating and Testing Linear Models with Multiple Structural Changes”, *Econometrica*, 66, 47-78.

[6] Bai, J., Perron P., (2003), “Computation and Analysis of Multiple Structural Change Models”, *Journal of Applied Econometrics*, 18, 1-22.

[7] Banerjee, A., Lumsdaine, R.L., Stock, J.H., (1992), “Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence”, *Journal of Business and Economic Statistics*, 10, 271-287.

[8] Berkowitz, J., Killian, L., (2000), “Recent developments in bootstrapping time series”, *Econometric Reviews*, 19, 1-48.

[9] Box, G. E. P. and G. M. Jenkins (1970), *Time series analysis: forecasting and control*, (revised edition 1976) Holden-Day, San Francisco.

[10] Brown, R., L., Durbin J., Evans, M., (1975), "Techniques for testing the Constancy of Regression Relationships over Time", *Journal of the Royal Statistical Society*, B 37, 149-192.

[11] Chow, G.C., (1960), "Tests of Equality Between Sets of Coefficients In Two Linear Regressions", *Econometrica*, 28, 591-605.

[12] Dickey, D.A. and W.A. Fuller, (1979) "Distributions of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427-31.

[13] Engle, R. F. and C. W. J. Granger (1987), "Cointegration and error-correction representation: estimation and testing," *Econometrica*, 55, 251-76.

[14] Granger, C. W. J. and P. Newbold (1974), "Spurious regressions in econometrics," *Journal of Econometrics*, 2, 111-20.

[15] Hansen, B.E., (1992), "Testing for Parameter Instability in Linear Models", *Journal of Policy Modeling*, 14, 517-533.

[16] Hansen, B.E., (1997) "Approximate asymptotic p-values for structural change tests.", *Journal of Business and Economic Statistics*, (1997), 15, 60-67.

[17] Hansen, B.E., (2000), "Testing for structural change in conditional models" *Journal of Econometrics*, 97, 93-115.

[18] Hansen, B.E., (2001), "The New Econometrics of Structural Change: Dating Breaks in U.S. labor Productivity", *Journal of Economic Perspectives*, 15, 117-128.

[19] Johansen, S. (1991), "Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-81.

[20] Lorentz, G. G. (1986), *Bernstein Polynomials*, Chelsea Publishers, London.

[21] Koutris, A., (2005), "Testing for Structural Change: Evaluation of the Current Methodologies, a Misspecification Testing Perspective and Applications". Ph.D. Dissertation, Virginia Polytechnic Institute and State University.

[22] Maasoumi, E. (2001), "On the relevance of first-order asymptotic theory to economics," *Journal of Econometrics*, 100, 83–86.

[23] Nelson, C.R. and C.I. Plosser (1982), "Trends and random walks in macroeconomic time series: some evidence and implications," *Journal of Monetary Economics*, 10, 139-62.

[24] Nyblom, J., (1989), "Testing for the Constancy of Parameters Over Time", *Journal of the American Statistical Association*, 84, 223-230.

[25] Perron, P. (1989), "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361-1401.

[26] Perron P., (2006), "Dealing with Structural Breaks", pp. 278-352 in Mills, T.C. and K. Patterson, *New Palgrave Handbook of Econometrics*, vol. 1, MacMillan, London..

[27] Phillips, P. C. B. (1986), "Understanding spurious regression in econometrics," *Journal of Econometrics*, 33, 311-40.

[28] Phillips, P. C. B. (1987), "Time series regressions with a unit root," *Econometrica*, 55**,** 227-301.

[29] Ploberger, W., Krämer, W., Kontrus, K.,(1989), "A new test for structural stability in the linear regression model", *Journal of Econometrics*, 40, 307-318.

[30] Quandt, R.E., (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", *Journal of the American Statistical Association*, 55, 324-330.

[31] Shewhart, W. A., (1939), "Statistical Method from the Viewpoint of Quality Control", Dover, NY.

[32] Sowel, F., (1996), "Optimal Tests for Parameter Instability in the Generalized Methods of Moments Framework", *Econometrica*, 64, 1085-1107.

[33] Spanos, A., (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press.

[34] Spanos, A., (1990), "Unit Roots and their Dependence on the Implicit Conditioning Information Set," Advances in Econometrics, 8, 271-292.

[35] Spanos, A., (1999), *Probability Theory and Statistical Inference*, Cambridge University Press.

[36] Spanos, A., (2000), "Revisiting Data Mining: 'Hunting' with or without a License," *Journal of Economic Methodology*, 7, 231-264.

[37] Spanos, A., (2001), "Time series and dynamic models," in: Baltagi, B., (Ed.), *A Companion to Theoretical Econometrics*, ch. 28. Blackwell Publishers, Oxford, pp. 585–609..

[38] Spanos, A., Kourtellos, A., (2002), "Model Validation and Resampling", Working Paper.

[39] Vinod, H.D., (2004), "Ranking Mutual Funds Using Unconventional Utility Theory and Stochastic Dominance," *Journal of Empirical Finance,* 11, 353-377.

7 Appendix A - Monte Carlo simulations

Table 2(b): Empirical Size for Andrews & Ploberger statistics for sample size n =100, based on R=10,000 replications

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	8.38	4.36	3.54	3.18
ExpF	2.61	3.88	2.80	2.51
AveF	2.01	2.11	2.29	3.31
Type I error for $\alpha = 1\%$				

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	8.38	25.170	24.800	23.540
ExpF	2.61	25.110	21.770	21.800
AveF	2.01	19.940	19.290	22.180
Type I error for $\alpha = 10\%$ *				

It reports the percentage of false rejections in R=10,000 experimental trials.

Ideally this should be equal to a%

Table 5: Empirical Size for the resampling testing procedure Type I error based on 10,000 replications							
Sample	Window	H_0: μ constant			H_0: σ^2 constant		
Size	Length	1%	5%	10%	1%	5%	10%
$n = 50$	4	1.21	5.42	9.75	1.01	4.75	8.54
	5	3.51	9.65	14.62	1.82	8.21	12.57
$n = 60$	4	0.86	3.27	6.43	1.32	4.55	7.79
	5	1.65	5.82	10.52	2.12	6.13	10.16
	6	3.07	8.74	14.44	3.11	5.09	13.56
$n = 70$	4	0.28	2.87	5.18	1.15	4.25	7.18
	5	0.91	4.68	9.05	2.10	4.81	9.77
	6	1.75	6.97	11.76	3.13	7.71	11.92
$n = 80$	5	0.65	3.78	7.11	1.83	4.45	7.82
	6	1.39	5.49	9.66	1.71	5.09	9.02
	7	3.35	7.04	11.69	2.87	7.32	13.51
$n = 90$	6	1.21	3.90	6.71	1.10	4.62	8.11
	7	1.42	5.61	9.68	1.58	5.99	10.46
	8	2.21	7.69	11.98	2.05	7.78	12.39
$n = 100$	7	1.39	4.78	8.51	1.71	5.10	8.84
	8	1.98	6.57	11.06	2.26	6.37	11.02
	9	3.10	7.72	12.07	2.59	8.42	13.21
$n = 110$	8	1.81	4.89	8.98	1.21	4.69	8.76
	9	2.23	5.69	10.81	1.91	5.38	11.65
	10	2.89	7.75	13.11	2.75	8.75	14.28
$n = 120$	9	1.35	3.75	7.78	2.18	5.89	8.97
	10	2.12	5.31	9.87	2.51	6.05	10.91
	11	2.63	6.98	12.31	3.81	9.71	14.02
It reports the percentage of false rejections in R=10,000 experimental trials. Ideally this should be equal to a%							

8 Appendix B - A&P tests, empirical results

Table 11: Exchange Rate Japan/US

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	6.4256	0.417	0.338	0.572
ExpF	1.1671	0.465	0.451	0.591
AveF	1.9141	0.417	0.403	0.425

Table 12: European Total Turnover Index

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	9.7149	0.277	0.285	0.086
ExpF	2.4727	0.269	0.335	0.101
AveF	2.8780	0.438	0.489	0.185

Table 13: US Industrial Production

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	2.0209	0.999	0.993	0.977
ExpF	0.45317	0.926	0.964	0.908
AveF	0.79433	0.907	0.961	0.884

Table 14: US Investment

Test Statistic		Andrews p-value	bootstrap p-value	Hetero-Corrected p-value
SupF	3.3021	0.992	0.980	0.995
ExpF	0.27832	1.000	1.000	1.000
AveF	0.39708	1.000	1.000	1.000