School of Economics
UNSW, Sydney 2052
Australia

http://www.economics.unsw.edu.au

Finite Sample Properties of the Dependent Bootstrap for Conditional Moment Models

Rachida Ouysse

School of Economics Discussion Paper: 2007/02

ISSN 1323-8949

ISBN 978 0 7334 2428 1

Finite Sample Properties of the Dependent Bootstrap for Conditional Moment Models

Rachida Ouysse*

PRELIMINARY AND INCOMPLETE

September 6, 2006

Abstract

This paper assesses the finite sample refinements of the block bootstrap and the Non-Parametric Bootstrap for conditional moment models. The study recononsiders inference in the generalized method of moments estimation of the consumption asset pricing model of Singleton (1986). These dependent bootstrap resampling schemes are proposed as an alternative to the asymptotic approximation in small samples and as an improvement upon the conventional bootstrap for time series data. This paper is a comparative simulation study of these resampling methods in terms of the differences between the nominal and true rejection probabilities of the test statistics and the nominal and true coverage probabilities of symmetrical confidence intervals.

JEL Classification: C1, C22, C52

Keywords: GMM estimation, Block Bootstrap, Markov Bootstrap, Power and size of a test.

*School of Economics, The University Of New South Wales, Sydney 2052 Australia. Email: rouysse@unsw.edu.au

1 Introduction

This paper addresses the finite sample statistical inference in the context of generalized method of moments estimation (GMM) for the nonlinear rational expectations model of Hansen and Singleton (1982) (HS henceforth) using dependent bootstrap for time series data. The paper will focus on two bootstrapping schemes: the block bootstrap as in Hall and Horowitz (1996) (HH) and the nonparametric bootstrap using Markovian local resampling. The emphasis lies on assessing the ability of these dependent bootstrap methods to improve upon the asymptotic approximation and the conventional bootstrap method. This is a comparative study of the performance of these resampling schemes in the context of a consumption asset pricing model with GMM estimation. The purpose is to compare the distortions of the sampling distribution of symmetrical (t-statistic) and non-symmetrical tests (test for overidentifying restrictions) from their asymptotic approximation, as well as the differences between their empirical and nominal rejection probabilities. The study also addresses the differences between the nominal and empirical coverage probabilities of symmetrical confidence interval.

Since Efron [17] first introduced the bootstrap in 1979, many applied statisticians and econometricians embraced this resampling procedure for estimating the distributions of estimators and test statistics. The bootstrap provides an alternative to tedious mathematical statistics when evaluating the asymptotic distribution is difficult. It resamples from one's data by treating the observed sample as if it was the population. The appeal of the bootstrap stems from its ability to provide higher-order accuracy and improvements over first order asymptotic theory[1] by reducing the bias and the mean square error especially in finite samples.

The relevance and validity of the bootstrap has been the center of focus and concern in both applied and theoretical literature. The work of Bickel and Freedman (1981) which established the asymptotic consistency, was followed by a stream of research on its higher order accuracy properties [see, e.g., Efron (1987), Singh (1981) or Beran (1988)]. The literature is far more rich when it comes to elucidating the comparative asymptotic properties of refined bootstrap methods. The asymptotic refinements of the bootstrap are achieved in situations where the statistic of interest is pivotal, meaning that its asymptotic distribution does not depend on any unknown parameters of the data generating process.

The use of bootstrap in hypothesis testing was highly advocated by several authors [see, e.g. Horowitz (1997, 2003a), Davidson and Mackinnon (1999)]. First order asymptotic theory often gives poor approximations to the distributions of test statistics in limited size samples. As a result, the discrepancy between the nominal probability that a test rejects a true null hypothesis using the asymptotic critical value and the true rejection probability can be very significant [See Davidson and MacKinnon (1999, 2006),Efron and LePage (1992)]. The bootstrap can also be used in confidence interval to reduce errors in coverage probabilities[2].

The focus of the literature in the recent decades has shifted away from the polishing

[1] Efron and Lepage (1992) summarize some of the results concerning higher order accuracy of the bootstrap.

[2] The difference between the true and nominal coverage probabilities is often lower when the bootstrap is used than when first order approximations are used.

of the bootstrap in the independent data setting towards extending the procedure to dependent observations. The well known accuracy of the Efron bootstrap to approximate the distributions of test statistics in the case of independent data is shown to fail in the case of dependent data in a number of studies [see, e.g., Singh (1992)]. Leger, Politis and Romano (1992) and Lepage and Billard (1992) show that the distribution of a statistic will generally heavily depend on the joint distribution of the observations. The difficulty to use the asymptotic distribution in this case opens the doors for the bootstrap to play a more fundamental enabling role to bypass problems associated to the use of asymptotics. However, it is clear that the resampling scheme appropriate to independent data will fail to provide consistent approximations even with weakly dependent data. The validity of the bootstrap in the dependent data case is a more delicate matter. Developments to dependent data models have considered both parametric and nonparametric alternatives. Parametrically, the bootstrap can be applied to structured dependent data models by assuming or estimating the structure of the dependence in the sample. Papers by Athreya and Fuh (1992), Datta and McCormick (1992) consider type of parametric methods in the case of Markov chain models.

Most of the interest lies however in nonparametric resampling schemes which divide the data into blocks. Künsch (1989) [41] and Liu and Singh (1992) propose a "moving blocks" resampling procedure for stationary time series data. The basic idea is to break the data into overlapping blocks of observations and the bootstrap data are obtained by independent sampling with replacement from among these blocks. Consistency of this bootstrap scheme under $m - dependence$ is achieved if the number of blocks increases with the sample size. Lahiri (1992) demonstrates that the rate of approximation of Künsch's method may be worse than the normal approximation and that second order correct approximation is obtained only through suitable modification in the bootstrap statistic. Hall and Horowitz (1996) [24] extends the non-overlapping blocking scheme first introduced by Carlstein (1986) [6] to the case of generalized method of moments estimator (GMM) and proposes some corrections to the formulae used for the $t - statistic$ and the test for overidentifying restrictions. The authors show that the new formulae will ensure the asymptotic refinements of the block bootstrap. All the resampling procedures based on blocking schemes depend on a "tuning parameter", which is the number of blocks and the length of each block. Hall, Horowitz and Jing (1995) [23] proposes a rule which depend on whether the distribution of the test statistic is symmetrical or not. Hansen (1999) suggests that the bootstrap distribution based on blocks may be an inefficient estimate of the sampling distribution. The author proposes a non-parametric data-dependent bootstrap which incorporates the moment restrictions by using a weighted empirical-likelihood of the conditional distribution. Applied to an $AR(1)$ model, Hansen's procedure performed better than HH block bootstrap in terms of test size in the monte Carlo study. This paper reexamines the results of Hansen (1999) and compares the two bootstrap methods in the context of GMM estimation of a nonlinear rational expectation model. The present study will evaluate the role of the error term structure in the sampling behavior of the sampling distribution of the statistics of interest.

The finding of the simulation study can be summarized in the following. First, the asymptotic approximation results in severe size distortion. The test statistic based on the asymptotic critical values reject more often than the nominal level. Second, both the block bootstrap and the Markov bootstrap outperformed the asymptotic approximation.

The second section introduces *HS* rational expectation model and sets the *GMM* estimation frame work along with the asymptotic results. Section 3 lays out the details of the block bootstrap procedure and summarizes the finding of Hall and Horowitz (1996). Section 4, discusses the Markov bootstrap and introduces Hansen's procedure. The results of the monte Carlo study are presented in section 4. Section 5 concludes with some remarks and future extensions.

2 The rational expectation Model

2.1 Hansen and Singleton model

Hansen and Singleton (1982) propose a new method which permit direct estimation and inference of nonlinear rational expectation model. The authors use the population orthogonality conditions[3] implied by the equilibrium stochastic Euler equations to construct a generalized method of moments estimator for a multi-period asset pricing model. *HS* estimation method circumvents the need of a complete, explicit characterization of the environment and do not require strong assumptions on forcing variables of the equilibrium path.

The model to be estimated is a consumption *CAPM*. As in *HS*, the representative consumer chooses stochastic consumption and investment so as maximizes the present value of his lifetime utility.

$$\max_{C_t} \; E_0[\sum_{t=0}^{\infty} \beta^t U(C_t)] \tag{1}$$

where β is a discount factor in the interval [0,1] and $U(.)$ is a strictly concave function. Suppose that the consumer has the choice to invest in N assets with maturity M_j, $j = 1, .., N$. The feasible set of consumption and investment plans must satisfy the sequence of budget constraints:

$$C_t + \sum_{t=1}^{N} P_{jt} Q_{jt} \leqslant R_{jt} Q_{jt-M_j} + W_t \tag{2}$$

where Q_{jt} is the quantity of asset j held at the end of period t. P_{jt} is the price of asset j at period t. R_{jt} is the payoff from holding asset j[4].

The Euler equation for this maximization problem is:

$$E_t \left[\beta^{n_j} \frac{U'(C_{t+n_j}, \gamma)}{U'(C_t, \gamma)} x_{jt+n_j} - 1 \right] = 0, \tag{3}$$

[3]These orthogonality conditions will depend in a nonlinear way on the parameters of the models and on the variables in the information set.

[4]In the case of stocks, the one period return from holding one unit of stock j is defined as:

$$x_{jt+1} = \frac{P_{jt+1} + D_{jt+1}}{P_{jt}}$$

where $x_{jt+n_j} = \frac{R_{n_j t+n_j}}{P_{n_j t}}$, $j = 1,..,m$ $(m \leqslant N)$.

Assume that he preferences are of the constant relative risk aversion type $U(C_t) = \frac{C_t^\gamma}{\gamma}$, and that the assets are held for one period $n_j = 1$, then the marginal rate of substitution $\frac{U'(C_{t+1},\gamma)}{U'(C_t,\gamma)} = \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha}$. Let x_{kt+1} be the ratio of the ratio of consumption at time $t+1$ to the consumption at time t, equation [3] becomes:

$$E_t\left[\beta(x_{kt+1})^{-\alpha}x_{jt+1} - 1\right] = 0 \tag{4}$$

where $\alpha = 1-\gamma$ is the risk aversion parameter (>0) and β is an impatience parameter. The stochastic Euler equation (4) implies a family of population orthogonality conditions. The *GMM* estimation method uses these moment conditions to construct an objective function whose optimizer is the estimate of the asset pricing model's parameters, α and β. Euler equation in (4) can be written as an error term. Let u_{t+1} be an m dimensional error term which measure the deviation from the equilibrium condition. This error term is a function of the model parameters and the observed variables. For the simple of two assets, case of $m = 2$:

$$u_{t+1} = \begin{bmatrix} \beta(x_{kt+1})^{\alpha}x_{1t+1} - 1 \\ \beta(x_{kt+1})^{\alpha}x_{2t+1} - 1 \end{bmatrix} = h(x_{t+1}, b)$$

where $b = (\alpha, \beta)$. The first order condition is then: $E_t[h(x_{t+1}, b)] = 0$, where, $x_{t+1} = (x_{kt+1}, x_{jt+1})'$.Given a set z_t of q instruments, available at time t, a family of population orthogonality conditions can be constructed based on the following moments functions[5]:

$$\begin{aligned} E[f(x_{t+1}, z_t, b_0)] &= 0 \\ f(x_{t+1}, z_t, b_0) &= h(x_{t+1}, b_0) \otimes z_t \end{aligned}$$

The *GMM* estimation method sets the sample versions of these moments conditions close to zero by minimizing the criterion function $J(b)$, at the point estimate b of the model parameters:

$$\min_b \; J(b) = \underset{1\times mq}{g(X,b)'} \; \underset{mq\times mq}{\Omega_n^{-1}} \; \underset{mq\times 1}{g(X,b)}$$

where $g(X,b) = \frac{1}{n}\sum_{t=1}^n f(x_{t+1}, z_t, b)$, and Ω_n is a weighting matrix.

The first order condition to the minimization problem:

$$\left[\frac{\partial g(X,b)'}{\partial b}\Omega_n^{-1}\right] g(X,b) = 0 \tag{5}$$

The *GMM* point estimates have the important feature of being consistent and have a limiting normal distribution under fairly weak conditions[6]. The asymptotic covariance of the estimators will depend on the choice of the weighting matrix. The most efficient choice is to set

$$\Omega_n = \frac{1}{n}\sum_{t=1}^n\sum_{s=1}^n E[g_t(X,b_0)g_s(X,b_0)']$$

[5] $f(x_{t+1}, z_t, b_0) = h(x_{t+1}, b_0) \otimes z_t$ stands for the Kronecker product of the $m\times 1$ vector $h(x_{t+1}, b_0)$ and the $qx1$ vector of instruments z_t. The product is an $mq \times 1$ vector.

[6] Sufficient conditions for strong consistency and asymptotic normality are provided in Hansen (1982).

Efficient estimation can be achieved by using the heteroskedasticity and autocorrelation consistent (HAC) covariance estimators. This paper uses Andrews and Monahan (1992) *(AM)* prewhitened HAC estimators which estimates the spectral density at frequency zero by using a prewhitening procedure to transform the data into an uncorrelated sequence before constructing the kernel HAC estimator as in Newey and West (1987). The estimators considered are prewhitened kernel estimators with vector autoregressions employed in the prewhitening stage. AM propose the use a vector autoregression VAR model to filter some of the temporal dependence in the series of sample moments $g_t(X,b)$. The standard HAC estimator Ω_n^{HAC} is then computed using the VAR residuals $\widetilde{g}_t(b)$, from the regression model $g_t(X,b) = \sum_{s=1}^{\kappa} \widehat{A}_s g_{t-s}(X,b) + \widetilde{g}_t(b)$:

$$\Omega_n^{HAC} = \frac{n}{n-k} \sum_{j=-n+1}^{n-1} \varphi\left(\frac{j}{D_n}\right) \left[\frac{1}{n-k} \sum_{t=\iota}^{n} \widetilde{g}_t(b) \widetilde{g}_{t+q}(b)'\right] \tag{6}$$

$$\left\{ \begin{array}{c} \iota = j+1,\ q = -j \text{ if } j \geqslant 0 \\ \iota = -j+1,\ q = j \text{ if } j < 0 \end{array} \right\}$$

The kernel used in this paper is the QS kernel defined as $\varphi(x) = \frac{25}{12\pi^2 x^2} \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)$. The QS kernel yields a positive semi-definite HAC covariance matrix estimator and Andrews (1991) argues the QS kernel possesses some large sample optimality properties. The data-dependent bandwidth D_n is defined in Andrews (1991) as $D_n = 1.3221(\widehat{a}(2)n)^{1/5}$, where

$$\widehat{a}(2) = \frac{\sum_{s=1}^{p} w_s \frac{4\widehat{\varrho}_s^2 \widehat{\sigma}_s^2}{(1-\widehat{\varrho}_s)^8}}{\sum_{s=1}^{p} w_s \frac{\widehat{\sigma}_s^4}{(1-\widehat{\varrho}_s)^4}}; \text{where } p = mq \tag{7}$$

The parameters $(\widehat{\varrho}_c, \widehat{\sigma}_c^2)$ are the autoregressive and innovation variance estimates from the first order autoregressive approximation model for the series $g_{s,t}(X,b)$, with $s = 1,...,p$. The weights w_s represent the relative importance of each of the disturbance vectors and are set to be equal to the inverse of the standard deviation of the corresponding residuals.

The prewhitened HAC estimator is computed by recoloring the usual HAC estimator Ω_n^{HAC} to recover the correlation properties of the data. Consequently, the AM "optimal" weighting matrix being considered is the following:

$$\begin{aligned} \Omega_n &= \widehat{\Lambda} \Omega_n^{HAC} \widehat{\Lambda}' \\ \widehat{\Lambda} &= \left(I - \sum_{s=1}^{\kappa} \widehat{A}_s\right)^{-1}; \qquad W_n = \Omega_n^{-1} \end{aligned} \tag{8}$$

The covariance matrix of the parameters estimates is given by:

$$V(\widehat{b}) = (M_n' W_n M_n)^{-1} M_n' W_n \Omega_0 W_n M_n (M_n' W_n M_n)^{-1}$$

with

$$\begin{aligned} M_n &= \frac{\partial g(X,b)'}{\partial b} \\ \Omega_0 &= E\left[\Omega_n(b_0)\right] \end{aligned} \tag{9}$$

Notice that, if the objective weight matrix is set as a function of the parameters in the objective function, e.g., $J(b) = g(X,b)'\Omega_n^{-1}(b)g(X,b)$, then the covariance matrix of the parameters is equal to $V(\widehat{b}) = (M_n'\Omega_0^{-1}M_n)^{-1}$.

Asymptotic theory suggests that the *GMM* estimator $\widehat{b} = (\widehat{\alpha}, \widehat{\beta})$ is consistent and normally distributed with mean $b_0 = (\alpha, \beta)$ and covariance matrix $V(\widehat{b})$. The results hold under conditional heteroskedasticity and serial correlation of the error terms. The instruments need not to be "econometrically" exogenous, they are only required to be predetermined at time t.

For the model to be identified and the estimation feasible, the number of orthogonality conditions $r = m \times q$, should be at least equal to the number k of unknown parameters in the model, in this case it is required that $r \geq 2$. The first order condition [5] sets the k linear combination of the r moment conditions equal to zero[7]. Consequently, there are $r - k$ remaining linearly independent moment conditions that are not set to zero in estimation, therefore should be close to zero if the model restrictions are true. Hansen (1982) uses a test statistic J_n, called $J - test$, to perform a test for these overidentifying restrictions. Under the null hypothesis of correct specification, the test statistic, defined as $J_n = nJ(\widehat{b})$, is shown by Hansen (1982) to be asymptotically distributed as a $\chi^2(r-k)$.

2.2 Estimation results

This section presents the empirical results for the stock prices model using the generalized method of moments to estimate the parameters of preferences, α and β in [4]. The real per capita consumption series is obtained by dividing each observation of the monthly and seasonally adjusted aggregate real consumption [nondurable plus services, obtained from *DRI* data base] by the corresponding observation on population [also from *DRI* data][8]. The consumption series is then paired with two measures of stock returns: the Equally-weighted and the Value-weighted returns obtained from *CRSP* data set. The estimation period considered in this exercise is $1960:01 - 1990 - 12$.

The vector of instruments is formed using lagged values of consumption and stock returns paired with a constant term. The number of lagged values, lag , is chosen to be 1, 2 and 4. The number of instruments[9] q, increases the number of orthogonality conditions and thus increases the number of overidentifying restrictions being tested. *HS* argues that the asymptotic covariance matrix of the estimates becomes smaller as the number of orthogonality conditions and might affect the sampling properties of the point estimates. *Table 1*1 represents the estimation results for the model parameters as well as the test for model fit. The point estimate for the discounted factor β, is stable and does not vary significantly with the number of lags considered in the instruments. However the corresponding $p - value$ changes dramatically from 71% for $lag = 1$ to 1.9% for $lag = 4$. The null hypothesis that $\beta = 1$ is almost never rejected for

[7]Through the columns of the $k \times r$ matrix $\frac{\partial g(X,b)'}{\partial b}W$.

[8]The *DRI* and *CRSP* data were provided by wrds.wharton.upenn.edu.

[9]In this application, since the instruments used are lagged values of consumption and a constant term, the number of instruments is equal to: $q = (lag \times 3) + 1$. The number of degrees of freedom is therfore: $df = m \times q - k$, where m is the number of returns, here $m = 2$ and k is the number of parameters, here $k = 2$.

Table 1: GMM Estimation for the CAPM with Value- and Equally-weighted Returns

lag	$\widehat{\alpha}$	$p_{\widehat{\alpha}}$	$\widehat{\beta}$	$\widehat{J}$	df	$p_{\widehat{J}}$
1	0.197 ($\widehat{se}_\alpha$=0.1028) t_α=1.9208 **CI(90%)=[−0.027;0.367]** ME=0.394	0.054	0.9999 ($\widehat{se}_\beta$=0.0004) (t_β=−0.3676) $p_{\widehat{\beta}} = 0.713$ **CI(90%)=[0.9993;1.000]** ME=0.0007	17.35	6	0.0084
2	0.0610 ($\widehat{se}_\alpha$=0.0942) t_α=0.647 **CI(90%)=[−0.0953;0.2173]** ME=0.311	0.0809	0.9994 ($\widehat{se}_\beta$=0.0003) t_β=−1.745 $p_{\widehat{\beta}} = 0.5174$ **CI(90%)=[0.9989;1.000]** ME=0.0011	25.12	12	0.0342
4	0.0815 ($\widehat{se}_\alpha$=0.0542) t_α=1.504 **CI(90%)=[−0.0083;0.1713]** ME=0.18	0.1336	0.9994 ($\widehat{se}_\beta$=0.0003) t_β=−2.345 $p_{\widehat{\beta}} = 0.0190$ **CI(90%)=[0.9989;0.9998]** ME=0.0009	29.19	24	0.2211

$lag = 1$ and 2. When the number of orthogonality conditions increases to $lag = 4$, the null hypothesis will be rejected for significance level larger than 1.9%. The confidence interval for the estimated discounted factor is very narrow and precise with a maximum width of 0.0011.

The results are somewhat different for the coefficient of relative risk aversion (RRA) α. The estimated values are significantly affected by the number of lags in the estimation. The $p - value$ for the null hypothesis $\alpha = 0$ ranges from 5.4% to 13.3% and thus the null is more likely to be rejected in the case with fewer lags and more likely to be true for the multiple lags case. The precision of the 90% confidence intervals for the coefficient of RRA significantly increases [53% reduction in the width when lag goes from 1 to 4] with the number of lags in the instruments. The increase in the precision is due to the decrease in the standard errors of the estimates. This confirms the effect of the number of orthogonality conditions on the covariance matrix as argued by Hansen and Singleton (1982).

Table-1 displays the estimated values $\widehat{J}$ for the test statistic for overidentifying restrictions. The number of overidentifying restrictions is given by the number of degrees of freedom df $(= r - k)$. The $p - value$ for the $J - statistic$ is based on the asymptotic theory that under the null hypothesis $\widehat{J} \sim \chi^2(df)$. Therefore, the *p-value* is defined as $p_{\widehat{J}} = Prob[\chi^2(df) \geqslant \widehat{J}_{obs}]$, where $\widehat{J}_{obs}$ is the observed value of the $J - statistic$ from the sample data. The $J - test$ based on the asymptotic theory provides greater evidence against the model when fewer lags are included the instruments vector. For example, with $lag = 1$, there is little evidence to support the null hypothesis since the $p_{\widehat{J}} = 0.85\%$. However, when 4 lags are included in the instruments, the $p - value$ increased to 22%.

3 Asymptotic refinements of Efron's Bootstrap

The asymptotic distribution theory is an approximation to the distribution of an estimator or a test statistic obtained from first-order asymptotic expansions [Edgeworth expansion]. Most of the test statistic are asymptotically pivotal, which means that the asymptotic approximation does not depend on unknown population parameters. However this approximation can be very poor in finite samples[10]. In the context of generalized-method-of-moments, monte Carlo experiments have shown that tests have the true levels that differ greatly from their nominal levels when asymptotic critical values are used. Moreover, not all statistics are pivotal and some distributions are difficult or impossible to estimate analytically.

The bootstrap is an alternative to first order expansion to approximating the distribution of a function $g_n(X, F_0)$ of the observations and the underlying distribution F by a bootstrap distribution $g_n(X, F_n)$. This resampling method consists of replacing the unknown distribution by the empirical distribution of the data in the statistical function, and then resampling the data to obtain a Monte Carlo distribution for the resulting random variable.

In the case of *GMM* estimation, the bootstrap method operates as follows. It is required to estimate the distribution of the test of overidentifying restriction, $J = J(X_1, .., X_t, .., X_n, F_0)$ and the $t-statistic$, $t = t(X_1, .., X_t, .., X_n, F_0)$ for the parameters estimates. These random variables depend on the data $X_1, .., X_t, .., X_n$, which are considered as a random sample drawn from an unknown probability distribution whose cumulative distribution is F. This probability distribution must satisfy the null hypothesis for the tests to have any power. The bootstrap method approximates the sampling distribution of $J(X_1, .., X_t, .., X_n, F_0)$ (respectively $t(.)$) under F_0 by that of $J(Y_1, .., Y_t, .., Y_n, F_T)$ under the empirical distribution F_n, where $\{Y_1, .., Y_t, .., Y_n\}$ denotes a random sample from F_n.

The bootstrap method is shown to improve upon first asymptotic approximations under mild regularity conditions[11]. The bootstrap ability to provide asymptotic refinements is very important for bias reduction and hypothesis testing. It is not unusual that an asymptotically unbiased estimator to have a large finite-sample bias. There is a well established body of the literature on the asymptotic refinements of the bootstrap over the first order approximation. In many cases, the bootstrap can be used to reduce both the finite sample bias and the mean-square error.

Theorem 3.1 (Horowitz 2001, [30]) *Suppose the data are a random sample,*$\{X_i, i = 1, ..., n\}$*. Assume that the parameter of interest* θ *can be expressed as a smooth function of a vector of population moments. Suppose that the true value of* θ *is* $\theta_0 = g(\mu)$*, where* $\mu = E(X)$*.*

- θ *can be consistently estimated by* $\theta_n = g(\overline{X})$ *. The bias of* θ_n*,*$B_n = E[\theta_n - \theta_0]$ *is of order* $O\left(n^{-1}\right)$*.*

[10] The information matrix test of White (1982) [55] is an example of a test in which large finite-sample distortions of level can occur when asymptotic critical values are used. See Taylor (1987), Orme (1990)

[11] Assumption of $g_n(X, F_n)$ a smooth function, continuously differentiable with respect to any mixture of components of F_n in the neighborhood of F_0.

- *Let* $X_i^*, i = 1, ..., n$ *be the bootstrap sample. The bootstrap estimate for* θ *is* $\theta_n^* = \overline{X}^*$, *where* $\overline{X}^* = \frac{1}{n}\sum_{j=1}^{n} X_j^*$.
- *The bootstrap bias is given by:* $B_n^* = E[\theta_n^* - \theta_n]$
- *Horowitz (2001) shows that:* $E(B_n^*) = B_n + O(n^{-2})$
- *The use of the bias corrected estimator,* $\theta_n - B_n^*$ *will result in a bias reduction of order* $O(n^{-1})$.

In small samples, the nominal probability that a test based on asymptotic critical value rejects a true null hypothesis can be very different from the true rejection probability. The use of first-order approximation may also result in a significant gap between the nominal and the true coverage probabilities when estimating confidence intervals. The first-order asymptotic approximation to the exact finite sample distribution of a symmetrical statistic (e.g., the $t-statistic$) makes an error of size $O(n^{-\frac{1}{2}})$, respectively $O(n^{-1})$ for a one sided pivotal distribution. In contrast, the error made by the bootstrap approximation is $O(n^{-\frac{3}{2}})$, respectively $O(n^{-1})$ for the one sided function (See, e.g., Horowitz (2001, 2003a)).

The bootstrap also provides an improved approximation to the finite sample distribution of an asymptotically pivotal statistic. Under the assumptions of smooth function model, the following theorem explains why the bootstrap provides higher order accuracy.

Theorem 3.2 (Horowitz 2001, [30]) *Suppose the data are a random sample,*$\{X_i, i = 1, ..., n\}$ *from a probability distribution whose CDF is* F_0. *Let* $T_n = T_n(X_1, ..., X_n)$ *be a statistic. Let* $G_n(\tau, F_0) = P(T_n \leqslant \tau)$, *be the exact , finite sample CDF of* T_n.

- *The Asymptotic CDF* $G_\infty(\tau, F_0)$ *of* T_n *satisfies* $sup_\tau |G_n(\tau, F_0) - G_\infty(\tau, F_0)| = O\left(n^{-1/2}\right)$.
- *The bootstrap CDF* $G_n(\tau, F_n)$ *provides a good approximation for the asymptotic CDF. Indeed,* $sup_\tau |G_n(\tau, F_n) - G_\infty(\tau, F_0)| \to 0$
- *The error of the bootstrap approximation to a one sided distribution function is* $G_n(\tau, F_n) - G_n(\tau, F_0) = O\left(n^{-1}\right)$ *almost surely over* τ.
- *The error of the bootstrap approximation to a symmetrical distribution function is* $[G_n(\tau, F_n) - G_n(-\tau, F_n)] - [G_n(\tau, F_0) - G_n(-\tau, F_0)] = O\left(n^{-3/2}\right)$ *almost surely over* τ.

However, in the case of GMM estimation, this is not necessarily true for the tests of overidentifying restrictions. In an overidentified model, the population moment condition $E(h(X, \theta_0))$ does not hold in the bootstrap sample, $E(h(X^*, \theta_n)) \neq 0$. The bootstrap estimator of the distribution of the $J-test$ is inconsistent, [5]. One way to solve this problem is to use of the recentered bootstrap moment condition $E(h^*(X_t, \theta)) = 0$, where $h^*(X_t, \theta) = h(X_t^*, \widehat{\theta}^*) - h(X, \widehat{\theta})$. Hall and Horowitz (1996) shows that recentering provides asymptotic refinements in terms the rejection probabilities for hypothesis testing and the distribution of the $J-test$. Brown and Newey (1995) propose an alternative approach to recentering. The authors replace the empirical distribution of X with an empirical likelihood estimator which satisfies the moment condition $E(h^*(X_t, \theta)) = 0$. This likelihood estimator assigns probability mass π_t to each observation X_t such that $\sum_{t=1}^n \pi_t = 1$ and $\sum_{t=1}^n \pi_t h(X_t, \widehat{\theta}) = 0$. Horowitz (2001) argues that with either method, the differences between the nominal and true rejection probabilities and between the nominal and true coverage probabilities of symmetrical confidence intervals is of order $O(n^{-2})$.

4 Bootstrap with dependent data

The independent bootstrap resampling method raises serious concerns in the case of dependent data. Resampling with replacement individual observations will destroy any dependence properties that might exists in the original data sample. To capture the dependence in he data, a number of alternative modifications to the bootstrap are proposed. If the dependence structure is known, a parametric model can be used to estimate this structure and reduce the data generation process to a transformation of independent random variables. This equivalent to first filtering the data before resampling and then recoloring the bootstrap[sample. If there is no parametric model to filter data dependence, an alternative bootstrap can be implemented by dividing the sample data into blocks, either non-overlapping [See, Carlestein (1986)] or overlapping [See, Hall (1985), Paparodits and Politis(2001,2002), and randomly resample the blocks with replacement.

In this section, the block bootstrap method will be implemented following Hall and Horowitz (1996). This section will describe the blocking rule and lay out the corrections proposed by the authors to achieve the asymptotic refinements in the context of generalized method of moment estimator. A simulation experiment is carried out to compare the performance of the asymptotic approximation, the recentered bootstrap and the corrected block bootstrap.

The relatively poor performance of block bootstrap in terms of the convergence rates which are only slightly faster than the first order approximation. The Markov bootstrap offers an alternative to the block bootstrap for processes which can be approximated by a Markov process. Horowitz (2003) derives the rates of convergence of the Markov bootstrap and shows that the errors made by Markov bootstrap converge to zero faster than the block bootstrap errors. However, this result is true under stronger conditions than the conditions required for the block bootstrap, mainly the Markov structure of the data. If the distribution of the data is not sufficiently smooth, the Markov bootstrap performs poorly and its errors converge slower than the block bootstrap.

4.1 Block Bootstrap

4.1.1 Non-overlapping Block resampling

Let $\{X_t, t = 1, .., n\}$ denote the observations from a stationary time series. The non-overlapping blocking method divides the sample into blocks of length l such that, the first block consists of observations $\{X_t, , t = 1, .., l\}$, block s consists of observations

$\{X_t, , t = l(s-1)+1, .., ls\}$ and so forth. The blocks (there are $\phi = n/l$) are then sampled randomly with replacement and laid end to end to obtain the bootstrap sample. This blocking methods does not replicate the exact the dependence structure in the data. As a result, the asymptotic refinements through $O(n^{-1})$ cannot be achieved by using the usual formulae for the bootstrap statistics. *HH* develops new versions of the later and proposes new formulae which correct for the differences between the asymptotic covariances of the original sample and bootstrap versions of the test statistics without distorting the higher order expansions that produce refinements.

The revised formulae of the bootstrap statistics depend on the blocking rule (number and length of blocks). Consider the symmetrical test statistic for the parameters hypothesis testing. Let H_n be the exact *CDF* of $t-statistic$: $H_T(z, F) = P(t \leqslant z)$. The rejection rule for the double sided hypothesis $H_0 : \theta = \theta_0$ at the level p is as follows: reject H_0 if $|t| > z_p$, where z_p satisfies $H_n(z_p, F) - H_n(-z_p, F) = 1 - p$. HH shows that the corrected formula for the bootstrap $t-statistic$ is: $\widetilde{t}^*(X^*, F_n) = \frac{S_n}{S_\phi} t(X^*, F_n)$, $t(X^*, F_n)$ is the usual $t - statistic$ applied to the bootstrap sample, S_n is the usual *GMM* standard errors of the parameter being tested from the original sample, the second term in the correction factor, S_ϕ, represents the standard deviation of the bootstrap estimate of the parameter of interest.

Since the estimation procedure in this paper uses a continuously updated *GMM*, in the sense that the weighting matrix is not fixed but is replaced with an asymptotically optimal estimate, then the covariance matrix of the parameters estimates is given by:

$$V(\widehat{b}) = (M_n' W_n M_n)^{-1} M_n' W_n \Omega_0 W_n M_n (M_n' W_n M_n)^{-1} = \begin{bmatrix} \widehat{\sigma}_{11} & \widehat{\sigma}_{12} \\ \widehat{\sigma}_{21} & \widehat{\sigma}_{22} \end{bmatrix}$$

where W_n and M_n are given by [8] and [9] respectively. Let ϕ be the number of blocks and l be the corresponding length, the bootstrap estimate for the covariance matrix of the parameters is shown by *HH* to be equal to:

$$\widetilde{V(b^*)} = (M_n' W_n M_n)^{-1} M_n' W_n \widetilde{W}_n W_n M_n (M_n' W_n M_n)^{-1} = \begin{bmatrix} \widetilde{\sigma}_{11} & \widetilde{\sigma}_{12} \\ \widetilde{\sigma}_{21} & \widetilde{\sigma}_{22} \end{bmatrix}$$

where

$$\widetilde{W}_n = \frac{1}{n} \sum_{i=1}^{\phi-1} \sum_{j=1}^{l} \sum_{k=1}^{l} g^*(X_{il+j}, \widehat{b}) g^*(X_{il+k}, \widehat{b})'$$

The *GMM* covariance matrix applied to the bootstrap sample is given by: $V(b^*)^* = (M_n^{*\prime} W_n^* M_n^*)^{-1} = \begin{bmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{22}^* \end{bmatrix}$, where $M_n^* = \frac{\partial g^*(X^*,b)'}{\partial b}$ and W_n^* is the prewhitened HAC estimator for the bootstrap covariance matrix.

$$\begin{aligned}
\widetilde{\Omega}_n^* &= \frac{n}{n-k} \sum_{j=-n+1}^{n-1} \varphi\left(\frac{j}{B_n}\right) \left[\frac{1}{n-k} \sum_{t=\iota}^{n} \widetilde{g}_t^*(b^*) \widetilde{g}_{t+q}^*(b^*)'\right] \\
g_t^*(X, b) &= \sum_{s=1}^{\kappa} \widehat{A}_s g_{t-s}^*(X, b) + \widetilde{g}_t^*(b) \\
\Omega_n^* &= \widehat{\Lambda} \widetilde{\Omega}_n^* \widehat{\Lambda}' \text{ where } \widehat{\Lambda} = \left(I - \sum_{s=1}^{\kappa} \widehat{A}_s\right)^{-1} \\
W_n^* &= (\Omega_n^*)^{-1}
\end{aligned}$$

The QS kernel and the bandwidth are computed by applying the formulae in [2.1] and [7] to the bootstrap data. Following this notation, the correction factor for the first parameter will be $\tau_1 = \frac{\widehat{\sigma}_{11}}{\widetilde{\sigma}_{11}}$ and for the second $\tau_2 = \frac{\widehat{\sigma}_{22}}{\widetilde{\sigma}_{22}}$. The corrected bootstrap $t - statistic$ for the parameter θ_i :

$$\widetilde{t}^* = \tau_i \frac{\sqrt{n}(\theta_i^* - \widehat{\theta}_i)}{\sqrt{\sigma_{ii}^*}} = \frac{\widehat{\sigma}_{ii}}{\widetilde{\sigma}_{ii}} \frac{\sqrt{n}(\theta_i^* - \widehat{\theta}_i)}{\sqrt{\sigma_{ii}^*}} \tag{10}$$

Hall and Horowitz (1996) demonstrate that this corrected version of the bootstrap test statistic for the parameters estimate provides asymptotic refinements to the asymptotic approximation. The levels for the test based on the corrected bootstrap critical values, $\widetilde{z}_t^*$, for $\widetilde{t}^*(X^*, F_n)$ are correct through $O(n^{-1})$, e.g., $P\left(\left|\widetilde{t}^*\right| > \widetilde{z}_t^*(p)\right) = p + O(n^{-1})$.

The usual formula for the bootstrap test for overidentifying restriction is obtained by applying the test formula to the bootstrap sample. In the case of dependent data, HH proposes the use of a new version to the re-sampled blocks of data. The corrected formula is given by:

$$\widetilde{J}^*(X^*, F_n) = ng^*(X^*, b)' W^{*\frac{1}{2}} V_N W^{*\frac{1}{2}} g^*(X^*, b)$$

where the correction matrix V_N is equal to:

$$V_N = M_N W_n^{\frac{1}{2}} \widetilde{W}_n W_n^{\frac{1}{2}} M_N \tag{11}$$

$$M_N = I - \Omega_n^{-\frac{1}{2}} M_n [M_n' \Omega_n^{-1} M_n]^{-1} M_n' \Omega_n^{-\frac{1}{2}} \tag{12}$$

The level of the test for overidentifying restrictions based on the corrected bootstrap critical value $\widetilde{z}_J^*(p)$ is also shown in HH to converge to the nominal level at a rate $O(T^{-1})$, thus

$$P\left(\widetilde{J}^*(X, F_n) > \widetilde{z}_J^*(p)\right) = p + O(n^{-1}).$$

4.1.2 Moving Block resampling

let $\{X_t, t = 1, .., n\}$ denote the observations from a stationary time series. Let $Q_i = \{X_i, ..., X_{i+l-1}\}$,with $1 \leqslant i \leqslant N$ be the collection of all overlapping blocks of length l, where $N = n - l + 1$. From these moving N blocks Q_i's, a sample of $\phi = \frac{n}{l}$ blocks is drawn with replacement. These blocks are then paste together to form a bootstrap sample.

The overlapping block resampling method was first introduced by Künsch (1989). [41] gave conditions under which the overlapping-blocks bootstrap consistently estimates the CDF in the case of a sample average. The asymptotic refinements to the estimation of distribution function were later investigated by Lahiri ([34], [35]).

Theorem 4.1 (Lahiri's Results [28], [34] and [35]) *In the case of estimating the CDF of a normalized or a Studentized function of a sample mean for an M-dependent DGP, the error in the bootstrap estimator of the one-sided distribution function is* $o\left(n^{-\frac{1}{2}}\right)$ *almost surely. This is an improvement over the asymptotic approximation which makes errors of order* $O\left(n^{-\frac{1}{2}}\right)$.

The relative accuracy of the Bootstrap with overlapping and non-overlapping blocks has been investigated by, Hall, *et al.* (1995) [23] and Lahiri (1999) [37]. Using the asymptotically optimal block length, their results show that the rates of convergence of the two blocking methods are the same. The bias of the bootstrap estimator is the same for non-overlapping and overlapping blocks. However, the asymptotic mean squared error, AMSE, is approximately 31% larger with non-overlapping blocks. In terms of AMSE in estimating a one-sided distribution function of a normalized sample average, [23] find that the AMSE is 22% larger with non-overlapping blocks.

4.1.3 Blocking rule

One of the issues which arises when using the block based resampling method to preserve the asymptotic refinements of the bootstrap is the choice of the blocking rule. The sampling properties of the bootstrap statistic and their performances as approximations to the true distributions depends on the number of blocks and their lengths.

The block length must increase with the sample size to enable the block bootstrap to produce consistent estimators of moments and and distribution functions and to achieve asymptotically correct rejection probabilities and coverage probabilities for confidence intervals.

The asymptotically optimal block length will depend on the function to be estimated. If the obejective is to estimate a moment condition or distribution function, the asymptotically optimal block length is defined as the one that minimizes the AMSE of the block bootstrap estimator. If however, the aim is to form confidence intervals (resp. test a hypothesis), an optimal block is one which minimizes the error in coverage probabilities (resp. the error in the rejection probability). Hall textitet al. (1995) show that the asymptotically optimal blocking rule is defined as: $l \sim n^{\omega}$ where ω is set to minimize the mean square error of the block bootstrap estimator. The authors show that setting $\omega = \frac{1}{5}$ is the optimal blocking rule for estimating a double sided distribution, e.g., the $t - statistic$ for parameters estimates, and $\omega = \frac{1}{4}$ for estimating a one-sided distribution such as the test of over-identifying restrictions.

4.2 The Markov Bootstrap

If the parametric structure of the dependence in the data is not available, then the block bootstrap works reasonably well under very week conditions on the dependency structure and no specific assumptions are made on the structure of the data generating process. However blocking will distort the dependence structure, and the resulting serial dependence patterns in the bootstrap samples can be very different from the original data. This will increase the error made by the bootstrap and will result in inefficient estimates of the sampling distribution. This poor performance of the block bootstrap methods has led to a search for alternative nonparametric resampling methods for time series data. If the data can be approximated by high order Markov process, then the Markov bootstrap is an attractive alternative with faster convergence rates than the block bootstrap.

The Markovian local resampling scheme introduced by Paparodits and Politis (1997) generates bootstrap data by reshuffling the original data according to a particular probability mechanism. This method applies to Markov processes or processes which can be approximated by one with sufficient accuracy. The resampling is based on nonparametric estimation of the Markov transition density, and sampling the process implied by this estimated density.

In the case of a conditional moment restrictions model like the CAPM example of this paper, the Markovian structure is derived from the conditional moments implied by the model. Hansen (1999) describes this resampling method based on the localized empirical likelihood and the moment restrictions of the model.

The resampling method is based on a nonparametric estimate of the transition density $f(X_{t+1}|X_t)$. Using a kernel density estimate for the density of X_t and the joint density of (X_t, X_{t+1}), the transition density can be estimated by:

$$w_s = P(X_{t+1} = X_s | X_t = x) = \frac{K_h(X_{s-1} - x)}{\sum_{r=1}^{n-1} K_h(X_r - x)};\ 2 \leq s \leq n$$

where $K_h(u)$ is the kernel function with bandwidth h. Following Hansen (1999), we choose a multivariate kernel $K_h(u) = r^{-1} \exp\left(-\left(\frac{u'\Sigma^{-1}u}{2r}\right)\right)$ with $\Sigma = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})(x_t - \overline{x})'$. The localized empirical likelihood must satisfy the conditional moment conditions implied by the theoretical model. Therefore the likelihood estimator assigns probability mass π_t to each observation X_t such that $\sum_{t=1}^{n} \pi_t = 1$ and $\sum_{t=1}^{n} \pi_t h(X_t, \widehat{\theta}) = 0$. The weights are found by solving the maximization problem given by:

$$\begin{aligned} \widehat{\pi} &= \arg\max_{\pi} \sum_{t=1}^{n} w_t \log(\pi_t) \\ s.t. \quad \pi_t &> 0;\ t = 1, .., n \\ \sum_{t=1}^{n} \pi_t &= 1 \text{ and } \sum_{t=1}^{n} \pi_t h(X_t, \widehat{\theta}) = 0 \end{aligned}$$

The localized resampling scheme can be summarized in the following steps.

1. Select a bandwidth h for the kernel estimator $K_h(u)$. In this paper we follow Hansen's suggestion to use the plug-in-rule of Silverman (1986) and set $h = \left(\frac{4}{1+2m}\right)^{\frac{1}{4+m}} T^{-\frac{1}{4+m}}$ for a Markov process of order m. Select a set of starting values $Y^*_{m,m} = \{X^*_j, j = 1, .., m\}$. A simple choice for the starting sequence is $Y_{m,m} = \{X_j, j = 1, .., m\}$, or as in Hansen (1999), select randomly a block of size m from the original sample.

2. To draw a bootstrap sample of size n, for any time point $t+1 \in \{m+1, ..., n+m\}$, the kernel weights for the transition from the actual state $Y^*_{t,m} = \{X^*_j, j = t, t-1, .., t-p\}$ to the state $Y^*_{s,m} = \{X^*_j, j = s, st-1, .., s-p\}$ for $s \in \{m, ..., n+m-1\}$ are given by
$$w_s = \frac{K_h(Y^*_{s,m} - Y^*_{t,m})}{\sum_{r=m}^{n+m-1} K_h(Y^*_{r,m} - Y^*_{t,m})}$$

3. The transition probabilities π_s are then estimated using the localized empirical likelihood as described above.

4. Draw a random number U from the uniform $(0,1)$. The bootstrap replicate is drawn from this discrete transition probability by selecting $X^*_{t+1} = X_{J+1}$, where $\sum_{s=1}^{J-1} \pi_s < U < \sum_{s=1}^{J} \pi_s$.

5 Simulation Study

5.1 Monte Carlo Experiment

This paper investigates the small sample properties of the *GMM* based Wald test ($J - test$) and the coverage probabilities of the parameters confidence intervals and compares the performances of the asymptotic theory, recentered bootstrap and the block bootstrap.

The data generating process considered in the experiment is based on the *CAPM* model. Given the series of consumption, the series of asset returns are generated under the null hypothesis that Euler equation [4] holds for given values of alpha and beta. The simulation experiment uses the observed series of the monthly consumption growth $\varkappa_t = \frac{C_{t+1}}{C_t}$ and the returns R are randomly generated as follows:

$$\begin{aligned}
\beta R_{i,t+1}\varkappa_{t+1}^{-\alpha} &= 1 + u_{i,t+1} \text{ where } E_t\left(u_{i,t+1}\right) = 0 \\
R_{i,t+1} &= \frac{\varkappa_{t+1}^{\alpha}}{\beta}(1 + u_{i,t+1}) \qquad (13)\\
\log(x_t) &= c + \rho \log(x_{t-1}) + \epsilon_t \qquad (14)\\
where, \epsilon_t &\sim IID(0, \sigma^2) \qquad (15)
\end{aligned}$$

Euler equation implies $E_t\left(\zeta_{i,t+1}\right) = 1$, which is equivalent to setting $E_t\left(u_{i,t+1}\right) = 0$. For the null model, we will consider the simple case of $u_{i,t} \sim iid\ N(0,1)$.

In the alternative model we will consider two potential situations where the equilibrium condition represented by the Euler equation does not hold.

Case 1 of H_a The alternative hypothesis corresponds to a situation where there is measurement errors in the consumption series. Suppose that the consumption series is observed subject to a multiplicative error: $C_t = C_t^*\eta_t$. The measurement error is stationary and independent of all information in the model. In this case $\varkappa_t = \frac{C_{t+1}}{C_t} = \frac{C^*_{t+1}}{C^*_t}\frac{\eta_{t+1}}{\eta_t}$.

$$\beta R_{i,t+1}\varkappa_{t+1}^{-\alpha} = \zeta_{i,t+1}\left(\frac{\eta_{t+1}}{\eta_t}\right)^{-\alpha}$$

$$\begin{aligned}
\zeta_{i,t+1} &= 1 + u_{i,t+1} \\
E_t\left(\zeta_{i,t+1}\frac{\eta_{t+1}}{\eta_t}\right) &= E_t\left(\zeta_{i,t+1}\right) \\
E_t\left(\frac{\eta_{t+1}}{\eta_t}\right)^{-\alpha} &= E_t\left(\frac{\eta_{t+1}}{\eta_t}\right)^{-\alpha}
\end{aligned}$$

If we assume that the measurement error is $iid\ N(a,\delta)$, then the conditional mean of the ratio of the measurement errors, $E_t\left(\frac{\eta_{t+1}}{\eta_t}\right)^{-\alpha} = \exp(\alpha^2\delta) = \kappa$, $a = 0$, $\delta = 2$.

Case 2 of H_a , The alternative model represents a situation where the expectation errors are serially correlated:

$$\begin{aligned} u_{i,t} &= w_i + \delta_i u_{i,t-1} + \nu_{i.t} \text{ where } \nu_{i.t} = e_{i,t}\sqrt{w + \gamma e_{i,t-1}^2} \\ e_{i,t} &\sim iid\ N(0,1) \text{ and let } w = 1,\ \gamma = 0 \\ w_i &= \{0, 2\};\ \ \delta_i = 0.95 \end{aligned}$$

The comparison of interest in this study is that between the performance of the three different methods of inference, asymptotic, block bootstrap and Markov bootstrap. The performance is measured through the distortion from the nominal level of the test statistic and the coverage probabilities of the confidence interval.

Let the model in [13] be the null hypothesis H_0, we wish to test. Let $\widehat{J}$ be the realization calculated from the data of the test statistic for overidentifying restrictions. The $p-value$ for inference based on $\widehat{J}$ is defined as: $p_{\widehat{J}} = Prob_{\theta_0}(J(X, F_0) \succeq \widehat{J})$. Since the true parameters of the population are unknown, the bootstrap $p - value,$ defined as $p^*_{\widehat{J}} = Prob_{\widehat{\theta}}(J^*(X, F_0) \succeq \widehat{J})$ for the recentered bootstrap and $\widetilde{p}^*_{\widehat{J}} = Prob_{\widehat{\theta}}(\widetilde{J}^*(X, F_0) \succeq \widehat{J})$ for the blocking method, is used to estimate $p_{\widehat{J}}$. The size of the test is measured by the rejection probabilities of the null hypothesis under the DGP in [13]. Let z^*_J and $\widetilde{z}^*_J$ be the $J - test$ critical values based on the recentered bootstrap and the block bootstrap respectively. The rejection probability for the recentered bootstrap, $RP^* = P(\widehat{J} > z^*_J)$ and $\widetilde{RP}^* = P(\widehat{J} > \widetilde{z}^*_J)$ for HH block bootstrap. Notice that the asymptotic rejection probabilities are given by the $\chi^2(df)$ critical values.

The asymptotic confidence interval for the parameter θ_i, at confidence level $1 - p$, is $\widehat{\theta}_i \pm z_{\frac{p}{2}} s_{\widehat{\theta}_i}$, where $z_{\frac{p}{2}}$ is the standard normal critical value, $Prob(N(0, 1) > z_{p/2}) = \frac{p}{2}$. With the bootstrap critical values, the confidence interval for the recentered bootstrap is $\widehat{\theta}_i \pm z^*_{\frac{p}{2}} \widehat{s}^*_{\theta_i}$, where $z^*_{\frac{p}{2}}$ is defined as $Prob(|t^*| > z^*_{\frac{p}{2}}) = \frac{p}{2}$ with $t^* = \frac{\sqrt{T}(\widehat{\theta}^*_i - \widehat{\theta}_i)}{\widehat{s}^*_{\theta_i}}$, and $\widehat{\theta}_i \pm \widetilde{z}^*_{\frac{p}{2}} \widehat{s}^*_{\theta_i}$ $Prob(|\widetilde{t}^*| > \widetilde{z}^*_{\frac{p}{2}}) = \frac{p}{2}$ with $\widetilde{t}^*(X^*, F_n) = \frac{\widehat{\sigma}_{ii}}{\widetilde{\sigma}_{ii}} \frac{\sqrt{n}(\widehat{\theta}^*_i - \widehat{\theta}_i)}{\widehat{s}^*_{\theta_i}}$ in equation [10]. The coverage probabilities of these confidence intervals are given by the likelihood that the estimated interval will cover the true population values.

The sample size considered in the experiment is $n = [60, 200, 330]$ and the number of assets $N = 2$. The number of lags in the instruments is set to $1, 2$ and 4. The number of bootstrap iterations is set to $B = 500$. Because the computations are time consuming, the number of monte Carlo replications is limited to $M = 5000$. The experiment is conducted as described in the following steps:

Step 1 Generate a random sample of size T based on the DGP in equation [13]. Use the data to compute the parameters estimates $\widehat{\theta}_i$, the standard deviations $s_{\widehat{\theta}_i}$ and the test statistic for overidentifying restrictions $\widehat{J}$.

Step 2 Generate a bootstrap sample by resampling with replacement the individual observations or the blocks from the sample data or using the Markov process. Compute the bootstrap estimates for the parameters, standard deviations, the $t - statistic$ and the $J - test$ from the bootstrap sample, e.g., $\widehat{\theta}^*_i$, t^* and $\widehat{J}^*$ for both the conventional bootstrap and the block bootstrap using the formulas in [10] and [11].

Step 3

- Repeat *step2* for $B = 300$ times and use the results to compute the empirical distribution of the bootstrap $t-statistic$, t^* and $\widetilde{t}^*$and the bootstrap $J-statistic$, J^* and $\widetilde{J}^*$.
- Set the critical values z^*_J, $\widetilde{z}^*_J$ equal to the $1-p$ percentile of the distribution of J^* and $\widetilde{J}^*$, for nominal level $p = 1\%$, 5% and 10%.
- Compute the critical values for the $t-statistic$ by setting $z^*_{\frac{p}{2}}$ (resp. $\widetilde{z}^*_{\frac{p}{2}}$) equal to the $1-p$ quantile of the empirical distribution of $|t^*|$ (resp. $|\widetilde{t}^*|$).
- Compute the bootstrap $p-value$, for both the conventional and the block bootstrap, by setting $p^*_{\widehat{J}}$ (resp. $\widetilde{p}^*_{\widehat{J}}$) equal to the percentage of J^* (resp. $\widetilde{J}^*$) which are greater than $\widehat{J}$.

Step 4

- Replicate *step1-3* for $M = 5000$ times. Use the results to compute the rejection probabilities by setting RP^* equal to the percentage of $\widehat{J}$ that are greater than z^*_J and $\widetilde{RP}^*$ to the fraction of $\widehat{J}$ larger than $\widetilde{z}^*_J$.
- The expected bootstrap *p-value* can be estimated by its sample average over the 5000 replications: $E(p^*_{\widehat{J}}) \simeq \frac{1}{500} \sum p^*_{\widehat{J}}$ and $E(\widetilde{p}^*_{\widehat{J}}) \simeq \frac{1}{5000} \sum \widetilde{p}^*_{\widehat{J}}$.
- Set the coverage probability equal to the fraction of intervals covering the true values of parameters used to generate the data.

5.2 Fast Methods to compute the RP and the Modified P values

The algorithm described above is computationally extremely costly especially with the Markovian bootstrap. For each of the M monte Carlo samples, we need to compute $B+1$ test statistic. Thus a total of $M(B+1)$ test statistics.

In this section we will describe the simulation techniques proposed by Davidson and Mackinnon([16], [13])to estimate the rejection probabilities and (modified) bootstrap P value at a fraction of the cost.

i. For each iteration m, $m = 1, ..., M$, generate a random sample of size n based on the DGP in equation [13]. Use the data to compute the parameters estimates $\widehat{\theta}^{(m)}_i$, the standard deviations $s_{\widehat{\theta}_i}$ and the test statistic for over-identifying restrictions $\widehat{J}_m$.

ii. Generate **one single** (instead of B) bootstrap sample for each resampling scheme (blocking or Markovian method). Compute the corresponding bootstrap test statistics t^*_m and $\widehat{J}^*_m$.

iii. Set the critical values $\widehat{Q}_J(p)$ and $\widehat{Q}^*_J(p)$ equal to the $1-p$ percentile of the sampling distribution of J and J^*.

Davidson and Mackinnon (2006) proposed the following two approximations for the rejection probabilities:

$$\widehat{RP}_1 \equiv \frac{1}{M}\sum_{m=1}^{M} I\left(J_m > \widehat{Q}_J^*(p)\right)$$
$$\widehat{RP}_2 \equiv 2p - \frac{1}{M}\sum_{m=1}^{M} I\left(J_m^* > \widehat{Q}_J(p)\right)$$

where $I(.)$ is an indicator function taking a value of 1 if its argument is true. These approximations can be used to compute the errors in rejection probabilities, which measures the difference between the actual and nominal level.

$$\widehat{ERP}_1 \equiv \widehat{RP}_1 - p$$
$$\widehat{ERP}_2 \equiv \widehat{RP}_2 - p$$

This method is quite inexpensive and requires computing only $2M$ test statistics. The accuracy of these fast approximations requires that the test statistics and the bootstrap DGP are independent or asymptotically independent, see [16]. $\widehat{RP}_1$ is shown to be more accurate than $\widehat{RP}_2$. However, the authors suggest that substantial differences between the two approximations may indicate that neither of them is accurate. Based on the same idea, [16] proposed modified bootstrap P value that are potentially more accurate than the conventional procedure described in *Step 3.*The procedure is described as follows:

1. For each of the B bootstrap replications, a first bootstrap sample is generated in similar way as in *Step-2* and is used to compute the bootstrap test statistic τ_j^* and a second-level bootstrap DGP. The latter is used to generate a new bootstrap sample and a second level bootstrap statistic τ_j^{**} is computed.

2. The conventional bootstrap *P value* is

$$\widehat{p}^* \equiv \frac{1}{B}\sum_{j=1}^{B} I\left(\tau_j^* > \widehat{\tau}\right)$$

3. Let $\widehat{Q}^*(\widehat{p}^*)$ be the $1-\widehat{p}^*$ percentile of τ_j^{**},

$$\widehat{p}^* \equiv \frac{1}{B}\sum_{j=1}^{B} I\left(\tau_j^{**} > \widehat{Q}^*(\widehat{p}^*)\right)$$

4. Davidson and Mackinnon (2000) proposed two versions of the fast double bootstrap P value:

$$\widehat{p}_1^{**} \equiv \frac{1}{B}\sum_{j=1}^{B} I\left(\tau_j^* > \widehat{Q}^*(\widehat{p}^*)\right)$$

$$\widehat{p}_2^{**} \equiv 2\widehat{p}^* - \frac{1}{B}\sum_{j=1}^{B} I\left(\tau_j^{**} > \widehat{\tau}\right)$$

In total, $2B+1$ test statistics are computed compared to $B+1$ for the conventional bootstrap *P value*. [16] found that this modified version has the potential of reducing quite substantially the errors in rejection probabilities which may occur for bootstrap tests. Once gain, comparing the improved *P value* to the conventional *P value* may shed light on their accuracy.

5.3 Results and discussion

5.3.1 Results for Non-overlapping Blocks using the full monte carlo

First, lets look at the sampling distribution of the estimated statistic for the $J-test$. *Panel 1* represents a graphical representation of the kernel density estimates of the probability density function for the random variable $\widehat{J}$. Different values of serial correlation in the errors are considered and the *GMM* estimation in done with $lag = 1$ and thus a plot of the theoretical $\chi^2(6)$ is also included. The series of consumption used in the simulation is the same observed data series used in the empirical estimation.

This graphical representation of $\widehat{J}$ and $\chi^2(6)$ confirms the finding in the literature that the asymptotic theory is a poor approximation in small samples. More precisely, the empirical distribution of $\widehat{J}$ is more skewed to the right and flatter than the asymptotic $\chi^2(6)$. As a result, the asymptotic theory will overestimate the rejection probability of the null hypothesis. For any value of $\widehat{J}$, the true empirical *p-value* is larger than the [estimated] asymptotic *p-value*. The sampling distribution becomes flatter with thicker tails as the degree of correlation in the errors increases. In the extreme case of $\rho_u = 0.9$, the distribution spans a larger range and is almost flat (the $p.d.f$ at the modal value is 16×10^{-6}).

Panel 4 represents the distribution of the bootstrap *J-statistic* along with the empirical distribution of the random variable $\widehat{J}$. The *GMM* estimation is performed with four lags in the instruments, the series of consumption is *iid* normal and the error terms have a moderate degree of correlation, $\rho_u = 0.25$. The graph suggests that the distribution of conventional bootstrap test ($\phi = T$) is very close to the true empirical distribution.

Secondly, the numerical results in tables below reinforce the graphical finding. The level of the asymptotic rejection probability *RP* is far larger than the nominal level. For example, in the case of $lag = 1$ and $\rho_u = 0$, *Table-2* shows that the true rejection probability for the asymptotic $\chi^2(6)$ is 79% which significantly larger than the 10% nominal level. The confidence level for interval estimation is equal to 81% for α and only 24% for β, which is very low compared to the 95% nominal confidence level. The same conclusion holds when the *GMM* estimation includes more lags, e.g., $lag = 2$, 4. The discrepancy between the true level nominal levels when using the asymptotic approximation becomes more severe in the case of serial correlation, e.g., $\rho_u = 0.70$.

5.3.2 Results under the null using ***Fast*** methods

In this section, we consider a DGP under the null model 13. We consider two processes for the consumption growth rate in 14, an $AR(1)$ with $\rho = 0.95$ and an $AR(4)$ with $rho = (0.95, 0.8, 0.01, 0.7)$ with normal innovations. In these experiments, we consider two sample sizes,$n = 100$ and $n = 200$. We use the *Fast* bootstrap method to esti-

Table 2: Block Bootstrap Confidence Intervals and Bootstrap *p-value*,$lag = 1, T = 330$

$CI(90\%)$	$\widehat{\alpha}$	$\widehat{\beta}$	$p_{\widehat{J}}$
Asy	$[-0.027; 0.368]$	$[0.999; 1.000]$	0.0084
$Conv-Boot$	$[-0.239; 0.525]$	$[0.998; 1.004]$	0.90
ϕ	**Hall & Horowitz**		
$n^{1-\rho_1} = 234$	$[-0.136; -0.031]$	$[0.9981; 0.9985]$	0.53
$n^{\frac{4}{5}} = 104$	$[-0.262; 0.1793]$	$[0.9984; 0.9987]$	0.99
$n^{\frac{3}{4}} = 78$	$[-0.170; -0.107]$	$[0.9988; 0.9990]$	1.00
$n^{1-\rho_2} = 79$	$[-0.202; -0.154]$	$[0.9985; 0.9988]$	1.00
$n^{1-\rho_3} = 27$	$[-0.280; -0.180]$	$[0.9985; 0.9988]$	1.00

Table 3: Block Bootstrap Confidence Intervals and Bootstrap *p-value*,$lag = 4, T = 330$

$CI(90\%)$	$\widehat{\alpha}$	$\widehat{\beta}$	$p_{\widehat{J}}$
Asy	$[-0.008; 0.171]$	$[0.9989; 0.9998]$	0.22
$Conv-Boot$	$[-0.540; 0.341]$	$[0.9978; 1.0000]$	0.88
ϕ	**Hall & Horowitz**		
$n^{1-\rho_1} = 232$	$[-0.695; -0.485]$	$[0.9962; 0.9970]$	0.54
$n^{\frac{4}{5}} = 104$	$[-0.435; -0.399]$	$[0.9977; 0.9979]$	0.22
$n^{\frac{3}{4}} = 78$	$[-0.567; -0.414]$	$[0.9975; 0.9979]$	0.23
$n^{1-\rho_2} = 76$	$[-0.424; -0.243]$	$[0.9974; 0.9984]$	0.24
$n^{1-\rho_3} = 25$	$[-0.510; -0.390]$	$[0.9972; 0.9979]$	0.49

mate the two versions of rejection probabilities, *RP1* and *RP2*. We also report the corresponding size discrepency, *ERP1* and *ERP2*.

5.4 Asset Pricing: Updated Inference

In light of the simulation results discussed in the previous section, it is apparent that using the asymptotic theory to conduct statistical inference will most likely lead to incorrect conclusions. The monte Carlo study clearly states that the asymptotic approximation will inflate the rejection probabilities and the resulting $p-value$ of the test statistics will be severely bias downwards.

In this section, an estimation of the bootstrap *p-value* and the bootstrap confidence intervals is conducted. The results are represented in *Table 2* and *Table 3* and are compared to the asymptotic inference of *Table-1*.

The bootstrap *p-value* shown in *Table 2* and *Table 3* suggest that the test statistic of overidentifying restrictions almost never rejects the null hypothesis of correct model specification. For the case of *GMM* estimation with $lag = 1$, the asymptotic *p-value* leads to a rejection for all nominal levels higher than 0.8%, however the *p-value* for the block bootstrap exceeds 53% for all values of the tuning parameter ϕ. The results are similar for $lag = 2$ and 4. The bootstrap interval estimates for the population parameters are more precise than the asymptotic intervals. Furthermore, the bootstrap

Table 4: Inference using Asymptotic Theory, Recentered and Block Bootstrap, $\rho_u = 0.70, \eta = 0, lag = 1, T = 330$

	RP			$E[p_{\hat{J}}]$	$PREJ(5\%)$	CI_α	CI_β
$Nominal - level$	**0.01**	**0.05**	**0.10**	**-**	**0.05**	**0.95**	**0.95**
$Asymptotic$	0.98	1.00	1.00	$-$		0.68	0.11
$Recentered - Boot$	0.00	0.00	0.02	0.67	0.00	0.73	0.98
$number - block$			**Hall**	**&**	**Horowitz**		
$T^{4/5} = 104$	0.03	0.06	0.11	0.44	0.10	0.99	0.98
$T^{3/4} = 78$	0.02	0.03	0.08	0.35	0.07	0.96	0.97
48	0.05	0.09	0.16	0.31	0.06	0.94	0.94
37	0.03	0.04	0.11	0.31	0.06	0.97	0.96
30	0.03	0.06	0.11	0.32	0.06	0.96	0.98
$T^{1-\rho_u} = 10$	0.02	0.04	0.08	0.31	0.06	0.99	0.97
8	0.02	0.13	0.27	0.19	0.06	0.97	0.99

confidence intervals achieve a higher level of coverage compare to the extremely low levels of coverage probability of the asymptotic interval estimates.

6 Concluding Remarks

This paper has provided an empirical investigation of the sampling properties of the test statistics and symmetrical confidence intervals in the case of nonlinear model using *GMM* estimation. It is well known that the first order asymptotic may be a poor approximation of the empirical distribution of the static of interest in small samples and that the bootstrap provides higher order refinements over the asymptotic test. In the case of temporal dependence, block bootstrap is used as an alternative resampling method to preserve the ability of the bootstrap to improve upon the asymptotic theory. However, little is known about the sampling properties of block bootstrap and its ability to improve upon the conventional (recentered) bootstrap method. This paper takes a close look at the properties of the three approximation methods and provides a comparative analysis of their performance in the context of the consumption asset pricing model.

The simulation analysis confirms the well known size distortion of the asymptotic approximation. The asymptotic test of overidentifying restrictions has a very small *p-value* leading to a high level of rejection probability. The differences in terms of coverage probabilities of the confidence intervals are significantly large when using the asymptotic theory.

The conventional bootstrap test generally tends to underestimate the rejection probability but the error is significantly low compared to the asymptotic approximation. The interval estimation of the population parameters using the usual bootstrap *percentile-t* method results in high levels of coverage probabilities which are close to the nominal confidence level.

The results concerning the block bootstrap are somewhat inconclusive. In general, the size distortion is lower than the conventional bootstrap, especially when only one or two lags are included as instruments in the estimation. The coverage probabilities are higher and although the discrepancy is low, they tend to overestimate the nominal levels.

References

[1] Andrews, D., W., K. and Monahan, C., J. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica, Vol. 60, No. 4, pp. 953-966.*

[2] Athreya, K. B. and Fuh, C., D. (1992), Bootstrapping Markov Chains, *In Exploring the Limits of Bootstrap (R. Lepage and L. Billard, eds.) pp. 49-64. Wiley, New York.*

[3] Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements, *Journal of the American Statistical Association, Vol.83, pp.687-697.*

[4] Bickel, P. J. and Freedman, D. A. (1981), Some Asymptotic Theory for the Bootstrap, *Annals of Statistics, 9, pp. 1196-1217.*

[5] Brown, B. and Newey, W. (1995), Bootstrapping for GMM, *Unpublished Manuscript, Deaprtment of Economics, Rice University.*

[6] Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series, *Annals of Statistics, Vol. 14, pp. 1171-1179.*

[7] Datta, S., and McCormick, W., P. (1992), Bootstrap for Finite State Markov Chain Based on i.i.d Resampling. *In Exploring the Limits of Bootstrap (R. Lepage and L. Billard, eds.) pp. 77-97. Wiley, New York.*

[8] Davison, A.C., Hinkley, D.V. and Young, G.V. (2003), *Recent Developments in Bootstrap Methodology. Statist. Science, Vol. 18, pp.141-157.*

[9] Davidson, R. and MacKinnon, J., G (1998), Graphical Methods for Investigating the Size and Power of Test Statistics, *The Manchester School, 66, pp. 1-26.*

[10] Davidson, R. and MacKinnon, J., G (1999), The Size Distortion of Bootstrap Tests, *Econometric Theory, 15, pp. 361-376.*

[11] Davidson, R. and MacKinnon, J., G. (1999), Bootstrap Testing in Nonlinear Models, *International Economic Review, 40, pp.487-508.*

[12] Davidson, R. and MacKinnon, J., G. (2000), Bootstrap Tests: How Many Bootstraps?, *Econometric Reviews, 19, pp. 55-68.*

[13] Davidson, R. and MacKinnon, J., G. (2002), Fast Double Bootstrap Tests of Nonnested Linear Regression Models, *Econometric Reviews, 21, pp. 417-427.*

[14] Davidson,R. and MacKinnon, J., G. (2002), Bootstrap J tests of nonnested linear regression models, *Journal of Econometrics, 109, pp. 167-193.*

[15] Davidson, R. and MacKinnon, J., G. (2006), The Power of Bootstrap and Asymptotic Tests, *Journal of Econometrics, forthcoming.*

[16] Davidson, R. and MacKinnon, J., G. (2006), Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap, *Computational Statistics and Data Analysis, forthcoming.*

[17] Efron, B. (1979), Bootstrap Methods, *Annals of Statistics, 7, pp. 1-26.*

[18] Efron, B. (1987), Better Bootstrap Confidence Intervals, *Journal od American Statistical Association, 82, pp. 171-200.*

[19] Efron, B. (2003). Second Thoughts on the Bootstrap, *Statistical Science, Vol. 18, pp. 135-140.*

[20] Efron, B. and R. LePage (1992), Introduction to Bootstrap, *In Exploring the Limits of Bootstrap, R. LePage and L. Billard (eds.), John Wiley and Sons.*

[21] Efron, B., and Tibshirani, R. J., (1993) An Introduction to the Bootstrap, *Chapman and Hall, New York.*

[22] Hall, P. (1985), Resampling a coverage process, *Stochastic Processes and their Applications, Vol.19, pp.259-269.*

[23] Hall, P., Horowitz, J., L. and Jing, B., Y. (1995), On Blocking Rules for the Bootstrap with Dependent Data, *Biometrika, 82, pp. 561-574.*

[24] Hall, P. and Horowitz, J., L. (1996), Bootstrap Critical Values for Tests Based on Generalized Method of- Moments Estimators, *Econometrica, 64, pp. 891-916.*

[25] Hansen, E., B. (1999), Non-Parmetric Dependent Data Bootstrap for Conditional Moment Models, *University of Wisconsin, www.ssc.wisc.edu/˜bhansen.*

[26] Hansen, L., P. (1982), Large Sample Properties of Generalized Method Of Moments Estimators, *Econometrica, Vol. 50, No. 4, pp. 1029-1054.*

[27] Hansen, L., P. and Singleton, K. (1982), Generalized Instrumental Variables Estimation Of Nonlinear Rational Expectations Models, *Econometrica, Vol. 50, No. 5, pp.1269-1286.*

[28] Hrdle, W., Horowitz, J.L. and Kreiss, J.-P. (2003), Bootstrap methods for time series, *International Statistical Review, Vol.71, pp.435-459.*

[29] Horowitz, J.L. (1997). Bootstrap Methods in Econometrics: Theory and Numerical Performance. In Kreps, D.M. and Wallis, K.F. (eds), *Advances in Economics and Econometrics: Theory and Applications, pp. 188-222, Cambridge University Press.*

[30] Horowitz, J.L. (2001), The Bootstrap, *In Heckman, J.J. and Leamer, E.E. (eds), Handbook of Econometrics, Vol. 5, Chap. 52, pp. 3159-3228, Elsevier Science.*

[31] Horowitz, J.L. (2003a). The bootstrap in econometrics. *Statistical Science, 18:211-218.*

[32] Horowitz, J. L.(2003b), Bootstrap Methods for Markov Processes, *Econometrica, Vol. 71, No. 4, pp. 1049-1082.*

[33] Horowitz, J., L. and Savin, N., E. (2000), Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach, *Journal of Econometrics 95, pp. 375-389.*

[34] Lahiri, S. N. (1991), Second order optimality of stationary bootstrap, *Statistics and Probability Letters, Vol. 11, pp. 335-341.*

[35] Lahiri, S. N. (1992), Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Nonstationary Data, *R. LePage and L. Billard (eds), Exploring the Limits of Bootstrap, Wiley, New York.*

[36] Lahiri, S., N. (1995), On The Asymptotic Behavior Of The Moving Block Bootstrap For Normalized Sums Of Heavy-Tail Random Variables, *The Annals od Statistics, Vol. 23, No. 4, pp. 1331-134.*

[37] Lahiri, S.N. (1999), Theoretical comparison of block bootstrap methods, *Annals of Statistics, Vol. 27 pp. 386-404.*

[38] Lahiri, S.N. (2003), Resampling Methods for Dependent data, *Springer, New York.*

[39] Leger, C., Politis, D., N., and Romano, J., P. (1992), Bootstrap Technology and Applications, *Technometrics, 34, pp.378-398.*

[40] Lepage, R., and Billard, L. (1992), Exploring the Limits of Bootstrap, *Wiley, New York.*

[41] Künsch, H., R. (1989), The Jacknife and the Bootstrap for General Stationary Observations, *Annals of Statistics, 17, pp. 1217-1241.*

[42] Liu, R. Y. and Singh, K. (1992), Moving Blocks Jacknife and Bootstrap Capture Weak Dependence. *In Exploring the Limits of Bootstrap (R. Lepage and L. Billard, eds.) pp. 225-248. Wiley, New York.*

[43] MacKinnon, J., G. (2002a), "Bootstrap J tests of nonnested linear regression models," *Journal of Econometrics, 109, pp. 167-193.*

[44] MacKinnon, J, G. (2002b), "Bootstrap inference in econometrics," *Canadian Journal of Economics, 35 pp. 615-645.*

[45] Newey, W., K. and Brown, B., W. (1998), Efficient Semiparametric Estimation of Expectations, *Econometrica, Vol. 66, No. 2, pp. 453-464.*

[46] Newey, W., K. and West, K., D. (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica, Vol. 55, No. 3, pp. 703-708.*

[47] Orme, C., (1990), The Small-Sample Performance of the Information Matrix Test, *Journal of Econometrics, Vol. 46, pp. 309-311.*

[48] Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis, Vol. 57, pp.277-296.*

[49] Paparodits, E. and Politis, N., D. (2001), A Markovian Local Resampling Scheme for Nonparametric Estimators in Time Series Analysis, *Econometric Theory, Vol. 17, pp. 540-566.*

[50] Paparodits, E. and Politis, N., D. (2002), The Local Bootstrap for Markov Processes, *Journal of Statistical Planning and Inference, Vol. 108, pp. 301-328.*

[51] Rajarshi, M.B. (1990). Bootstrap in Markov sequences based on estimates of transition density. Annals. Inst. Statist. Math., 42:253-268.

[52] Singh, K. (1981), On The Accuracy Of Efron's Bootstrap, *The Annals of Statistics, Vol. 9, N0. 6, pp. 1187-1195*

[53] Stomberg, C. and White, H.(2000), Bootstrapping the Information Matrix Test,*University of California San Diego Discussion Paper 2000-04.*

[54] Taylor, L., (1987), The Size Bias of White's Information Matrix Test, textitEconometrics Letters, Vol. 24, pp. 63-67.

[55] White, H. (1982), Maximum Likelihood Estimation of Misspecified Models,*Econometrica, Vol, 50, pp. 1-25.*

[56] Young, G., A. (1994), Bootstrap: More than a Stab in the Dark?, *Statistical Science, Vol. 9, No. 3, pp. 382-415.*

[57] Young, G., A. and Daniels, H., E. (1990), Bootstrap Bias *Biometrika 77, 1, pp. 179-85.*

Table 5: Inference using Asymptotic Theory, Recentered and Block Bootstrap, $\rho_u = 0.70, \eta = 0, lag = 2, T = 330$

	RP			$E[p_{\hat{J}}]$	$PREJ(5\%)$	CI_α	CI_β
$Nominal - level$	**0.01**	**0.05**	**0.10**	–	**0.05**	**0.95**	**0.95**
$Asymptotic$	0.86	0.91	0.92	–	–	0.68	0.11
$Recentered - Boot$	0.015	0.037	0.071	0.77	0.02	0.78	0.81
$Number - block$		**Hall**		**&**	**Horowitz**		
$T^{4/5} = 104$	0.09	0.10	0.13	0.44	0.11	0.97	0.83
$T^{3/4} = 78$	0.01	0.03	0.03	0.81	0.11	0.99	0.84
47	0.02	0.02	0.03	0.73	0.11	0.99	0.88
37	0.02	0.03	0.03	0.73	0.11	0.99	0.90
30	0.01	0.02	0.03	0.76	0.11	0.99	0.90
10	0.03	0.07	0.15	0.23	0.11	0.99	0.87
8	0.02	0.08	0.17	0.21	0.11	0.98	0.93

Table 6: Inference using Asymptotic Theory, Recentered and Block Bootstrap, $\rho_u = 0.70, \eta = 0, lag = 2, T = 330$

	RP			$E[p_{\hat{J}}]$	$PREJ(5\%)$	CI_α	CI_β
Nominal – level	**0.01**	**0.05**	**0.10**	–	**0.05**	**0.95**	**0.95**
$Asymptotic$	1.00	1.00	1.00				
$Recentered - Boot$	0.00	0.01	0.02	0.69	0.00	0.70	0.94
$Number - block$		**Hall**		**&**	**Horowitz**		
$T^{4/5} = 104$	0.16	0.23	0.25	0.37	0.22	0.99	0.98
$T^{3/4} = 78$	0.04	0.08	0.13	0.49	0.22	0.97	0.98
47	0.06	0.12	0.20	0.33	0.22	0.96	0.96
37	0.05	0.11	0.22	0.33	0.22	0.95	0.96
30	0.03	0.11	0.18	0.36	0.22	0.96	0.98
$T^{1-\rho_u} = 10$	0.09	0.37	0.77	0.08	0.22	0.97	0.96
8	0.12	0.37	0.65	0.10	0.22	0.98	0.99

Figure 1: Fast Bootstrap $RP1$ for J-test under the Null. The consumption growth is $AR(1)$ process with coefficients, $\rho = 0.95$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{1-\omega}, \omega = [0.4, 0.5, 0.6, 0.75, 0.9]$ and Markov lag $m = 2$

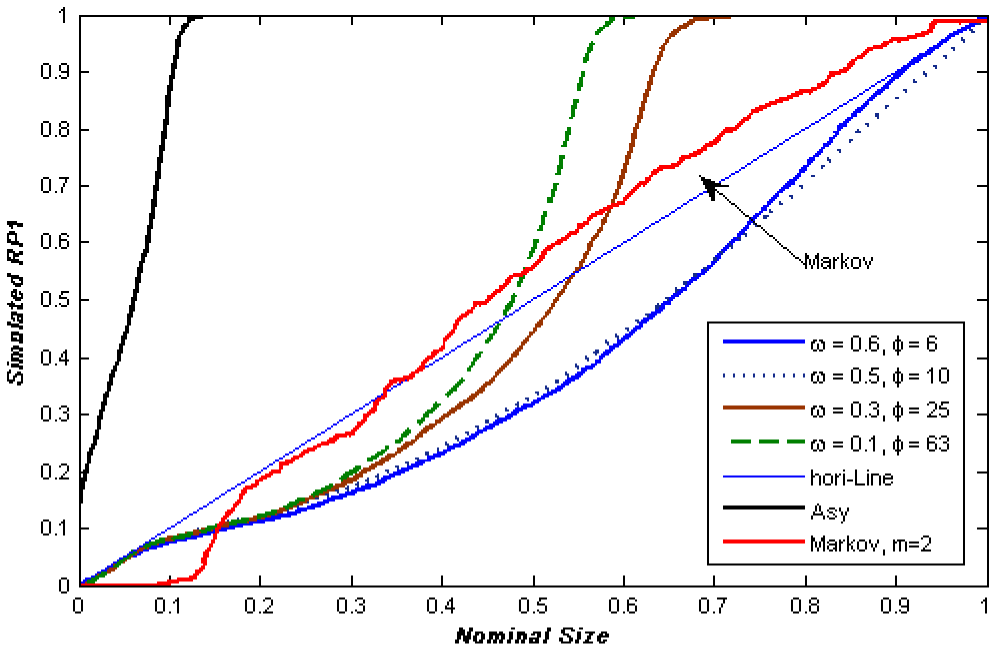

Figure 2: Fast Bootstrap $ERP1$ for J-test under the Null. The consumption growth is $AR(1)$ process with coefficients, $\rho = 0.95$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

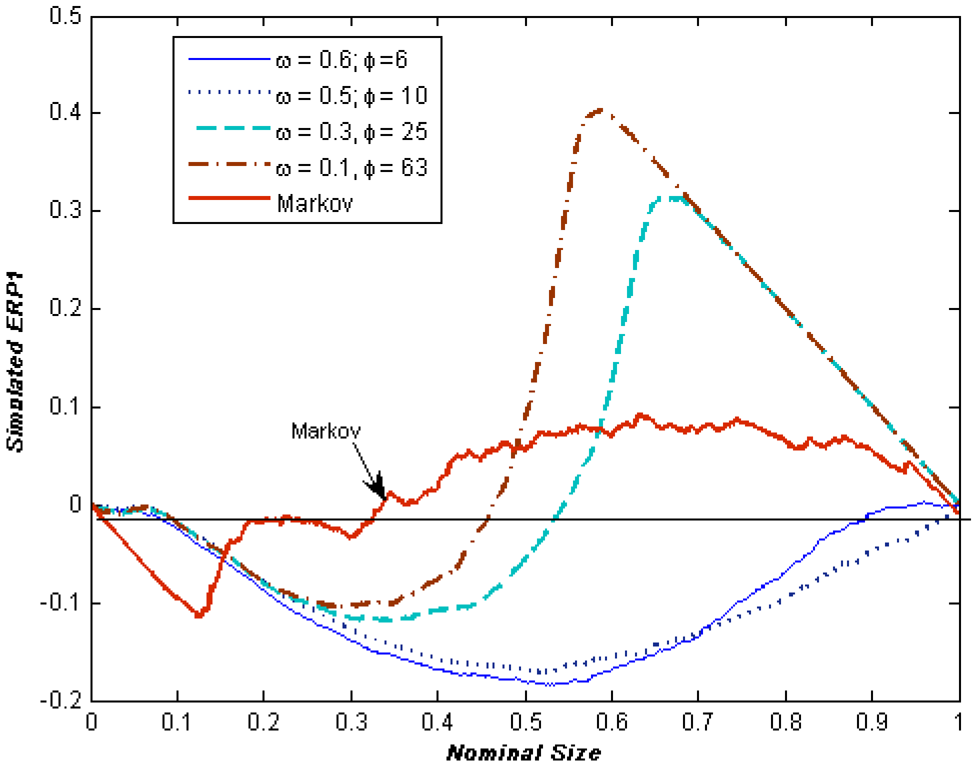

Figure 3: Fast Bootstrap $RP2$ and $RP1$ for J-test under the Null. The consumption growth is $AR(1)$ process with coefficients, $\rho = 0.95$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

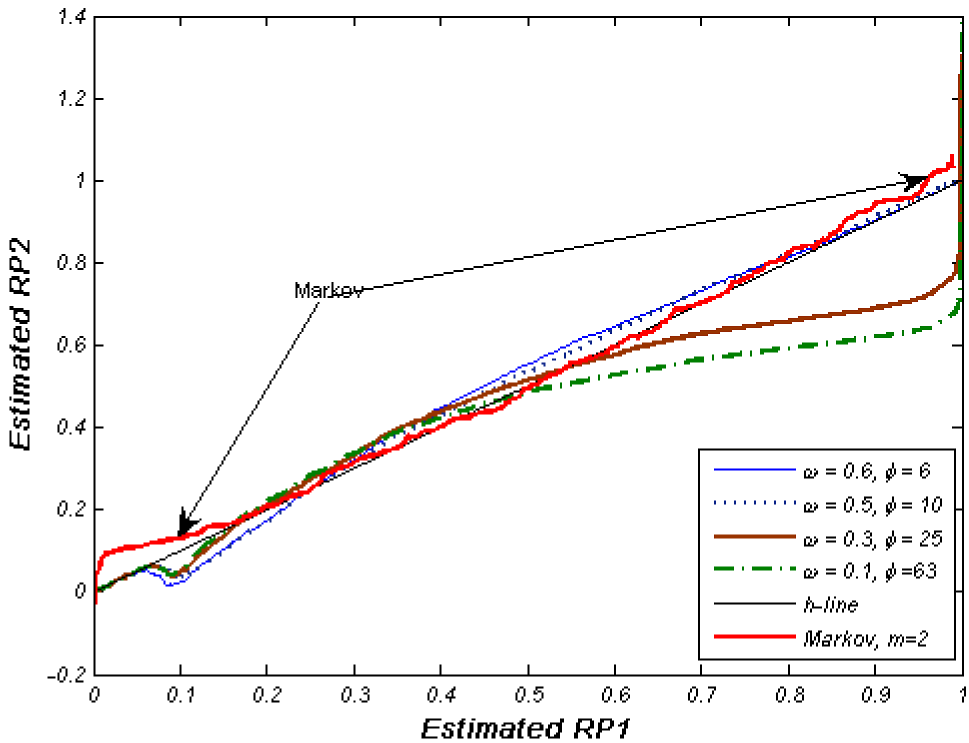

Figure 4: Fast Bootstrap $RP2$ for J-test under the Null. The consumption growth is $AR(1)$ process with coefficients, $\rho = 0.95$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

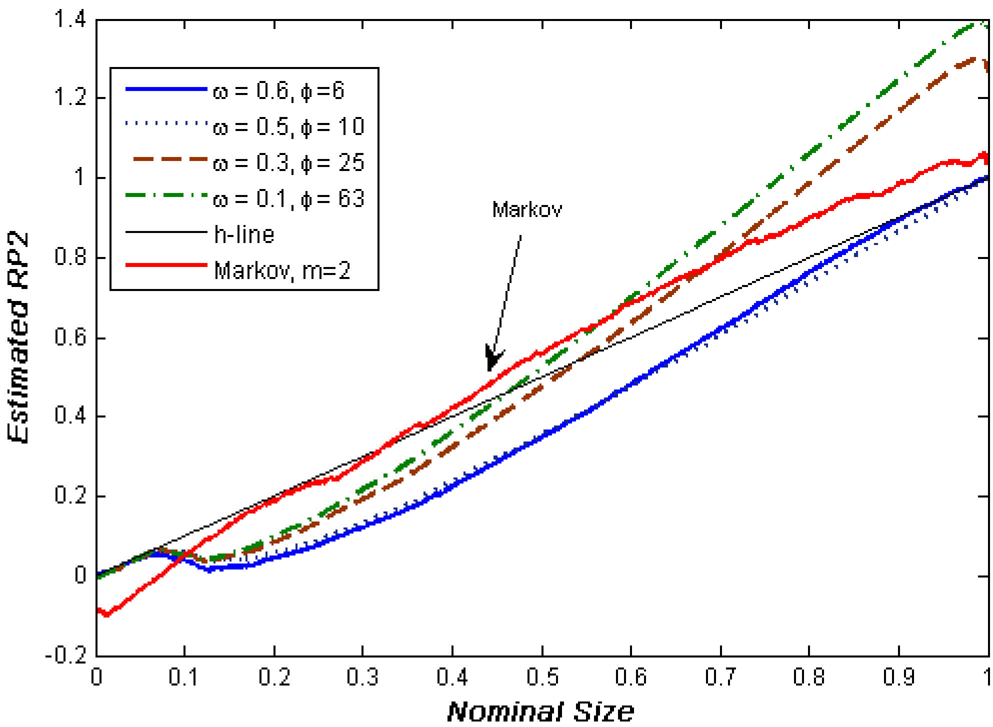

Figure 5: Fast Bootstrap $ERP2$ for J-test under the Null. The consumption growth is $AR(1)$ process with coefficients, $\rho = 0.95$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

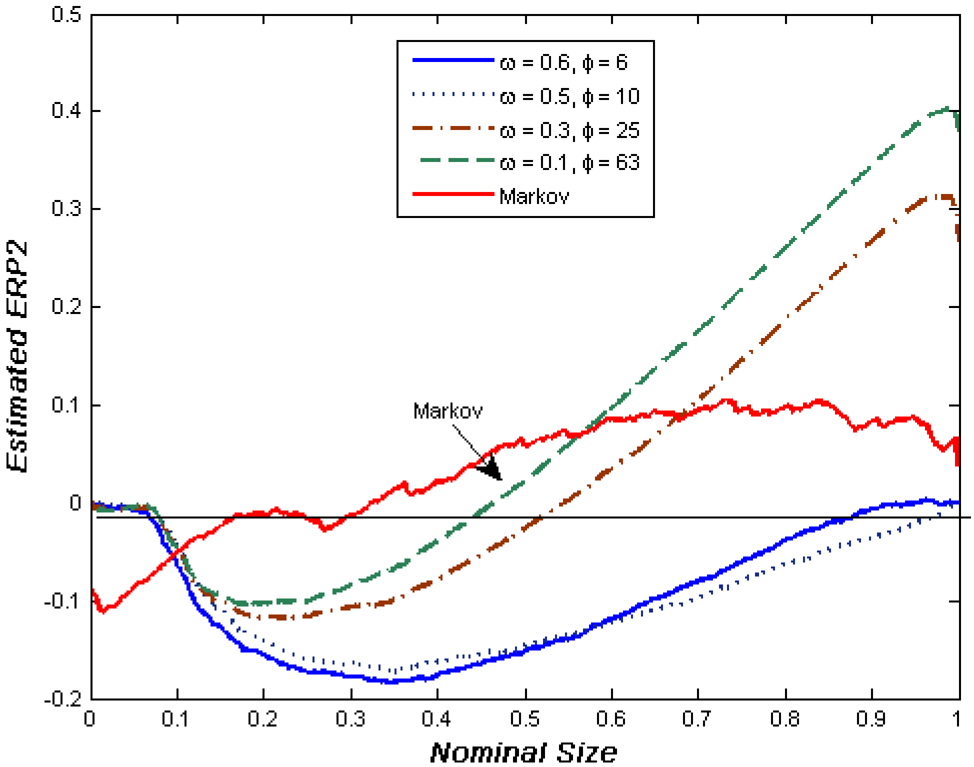

Figure 6: Fast Bootstrap $RP1$ for J-test under the Null. The consumption growth is $AR(4)$ process with coefficients, $\rho = [0.95, 0.8, 0.01, 0.7]$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^{1-\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

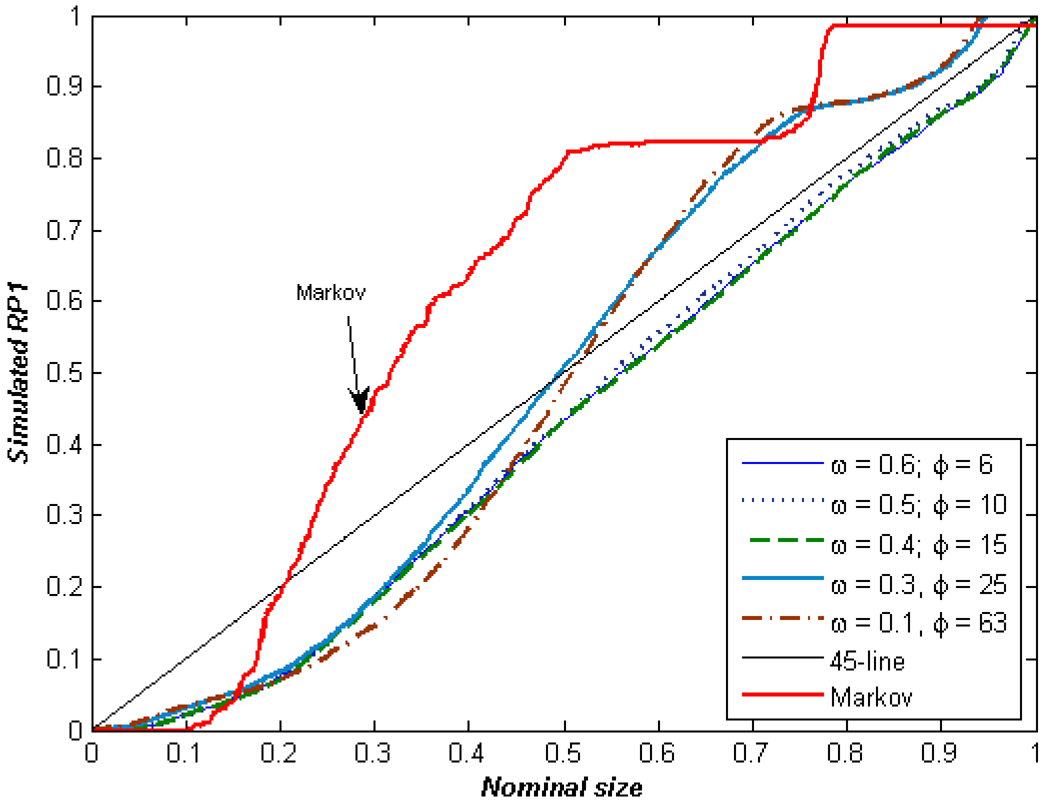

Figure 7: $ERP1$ for J-test under the Null: $AR(4)$, $\rho = [0.95, 0.8, 0.01, 0.7]$. The sample size is $n = 100$, $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

Figure C-2.

ERP1 for J-test. Null Model with $\phi = T^{1-\omega}$ blocks and Markov order m=2.
x_t is an AR(4) with $\rho = [0.95,0.8, 0.01, 0.70]$; T= 100

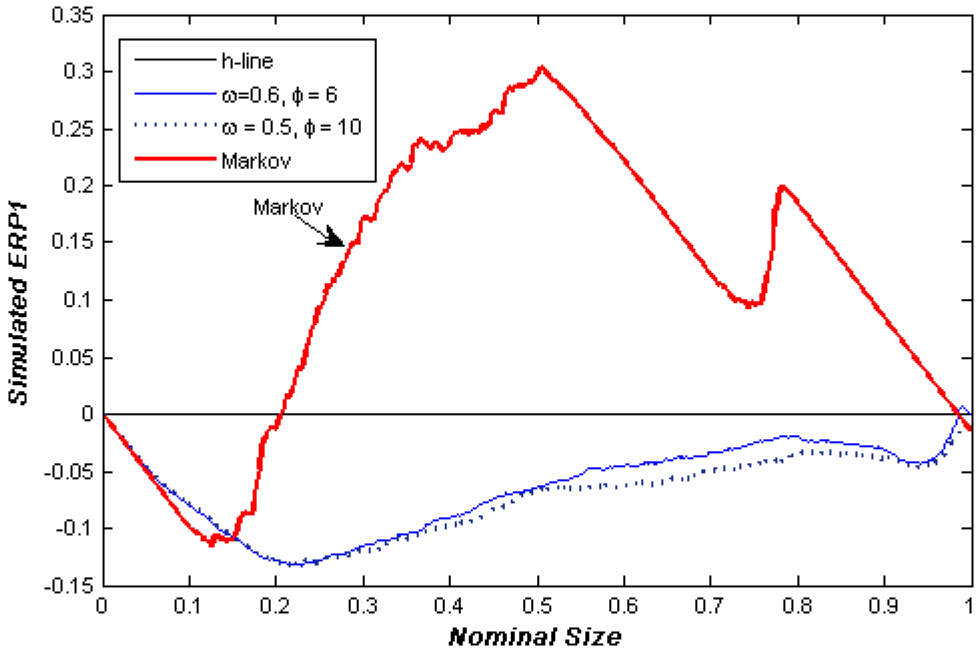

Figure 8: J-test under the Null: $AR(4)$, $\rho = [0.95, 0.8, 0.01, 0.7]$. The sample size is $n = 100$, $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

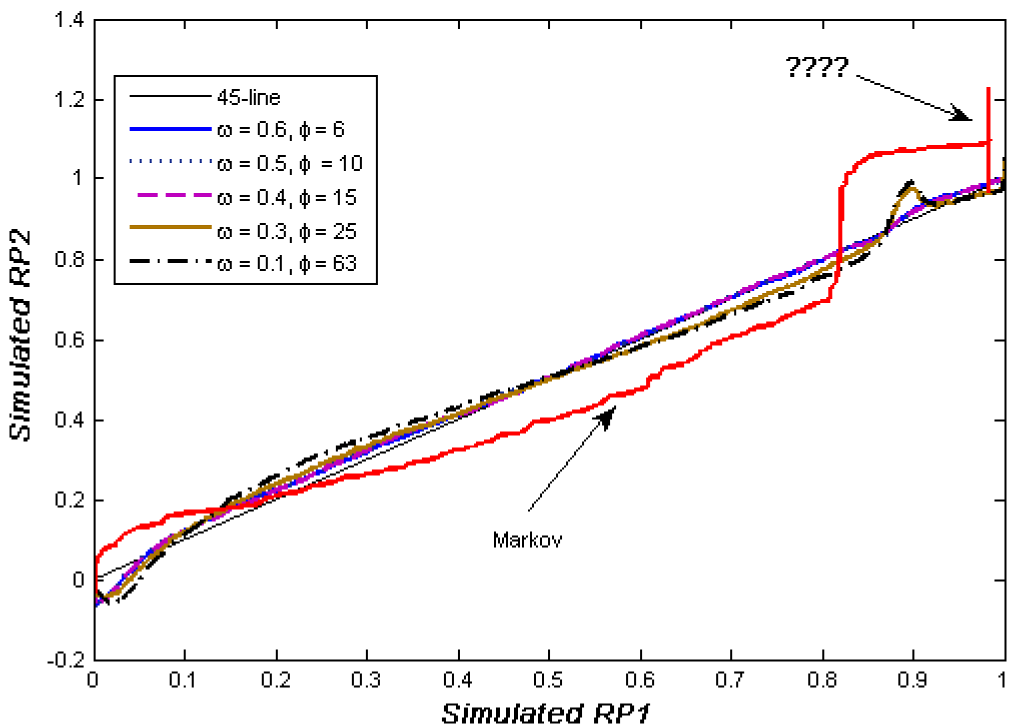

Figure 9: Fast Bootstrap $RP2$ for J-test under the Null. The consumption growth is $AR(4)$ process with coefficients, $\rho = [0.95, 0.8, 0.01, 0.7]$. The sample size is $T = 100$ and the number of simulation is $M = 10^3$. The block size is $l = T^\omega, \omega = [0.6, 0.5, 0.4, 0.3, 0.1]$ and Markov lag $m = 2$

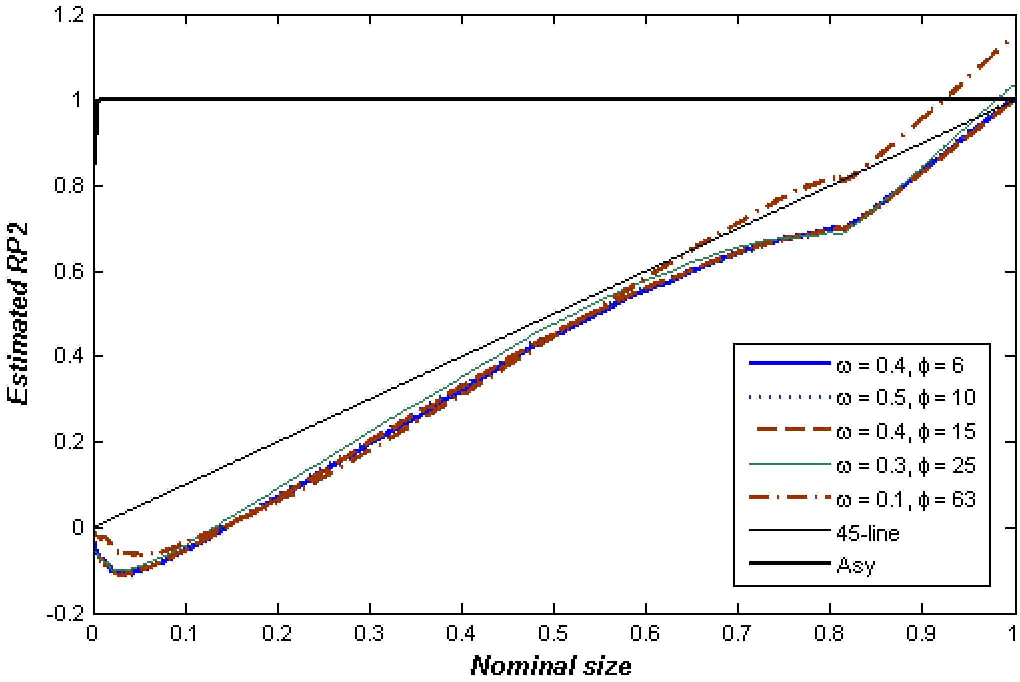

Figure 10: Effect of sample size on $RP1$ for J-test under the Null. The consumption growth is $AR(4)$ process with coefficients, $\rho = [0.95, 0.8, 0.01, 0.7]$. The sample size is $T = 100, 200$ and the number of simulation is $M = 10^3$. The block size is $l = T^{\omega}, \omega = [0.1, 0.3]$ and Markov lag $m = 2$