Persistence, Bias, Prediction and Averaging Estimators

Caroline JARDET [(1)] Alain MONFORT [(2)]

Banque de France CNAM, CREST and Banque de France

Fulvio PEGORARO[(3)]

Banque de France and CREST

May, 2009

Abstract

Persistence, Bias, Prediction and Averaging Estimators

Econometric modeling of persistent variables is not trivial given that, if the true dynamics is stationary but close to non-stationarity, non-stationarity tests may fail to reject it and, therefore, lead to serious flaws in the behavior of the model, in particular when long-run predictions are considered. This is a huge problem in many applications where the standard statistical tools do not provide a clear vision of the number of unit roots and of the number and the nature of cointegration relationships.

The objective of the paper is to see how to treat the stationarity vs. non-stationarity trade-off and the finite sample bias problem in order to optimize the prediction performances of the models. We will study and compare the in-sample and out-of-sample performances obtained from four "biased-corrected" estimators ("Kendall", Indirect Inference, Bootstrap and "Median-unbiased" estimators) with those coming from the averaging estimators *à la* B. Hansen (2009).

We will focus, first, on the simple AR(1) model, because we believe that the important problems show up still in this simple setting which can be studied accurately, even if we will propose solutions which are easily extended to more general models. Then, we will also consider the case of a bivariate "near-cointegrated" model.

Keywords : persistence, unit root, cointegration, near-cointegration, bias correction, averaging estimator, Kendall's bias approximation, Indirect Inference estimator, Bootstrap estimator, "Median-unbiased" estimator.

JEL classification : C52, C53.

[1]Banque de France, Financial Economics Research Service [DGEI-DEMFI-RECFIN; E-mail: Caroline.JARDET@banque-france.fr].

[2]CNAM and CREST, Laboratoire de Finance-Assurance, Banque de France, Financial Economics Research Service [E-mail: monfort@ensae.fr].

[3]Banque de France, Financial Economics Research Service [DGEI-DEMFI-RECFIN; E-mail: Fulvio.PEGORARO@banque-france.fr] and CREST, Laboratoire de Finance-Assurance [E-mail: pegoraro@ensae.fr]

1 Introduction

It is well known that many macroeconomic variables are persistent, in the sense that their dynamics imply high serial correlations. Examples of such variables are interest rates, exchange rates, inflation rates or price-dividend ratios. The econometric modeling of persistent variables is not trivial given that, if the true dynamics is stationary but close to non-stationarity, non-stationarity tests may fail to reject it and, therefore, lead to serious flaws in the behavior of the model, in particular when long-run predictions are considered. For instance, classical unit root tests like the Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) tests have size distortion and low power against several (persistent, for instance) alternatives when conventional sample sizes are considered [see, for instance, De Jong, Nankervis, Savin and Whiteman (1992a, 1992b), and Schwert (1989)]. Elliott, Rothenberg and Stock (1996) and Ng and Perron (2001) have proposed new unit root tests in order to improve the size and power of the classical ones. Nevertheless, even these more efficient tests tend to accept the presence of a unit root in highly persistent stationary time series [see, for instance, Jardet, Monfort and Pegoraro (2009)].

This is a huge problem in many applications where the standard statistical tools do not provide a clear vision of the number of unit roots and of the number and the nature of cointegration relationships. Let us, for instance, consider the simple univariate AR(1) model $y_t = \mu(1-\rho) + \rho y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim II(0, \sigma^2)$. If the true value of ρ is 0.95, the best prediction of y_{t+h} at time t is, for $h = 20$ (a five-year horizon with quarterly data), $\mu + 0.95^{20}(y_t - \mu)$ i.e. $\mu + 0.36(y_t - \mu)$, whereas, if $\rho = 1$ is accepted, the prediction is y_t and could be very misleading. The same "discontinuity" problem clearly applies also to the multivariate setting as indicated by Cochrane and Piazzesi (2008) and Jardet, Monfort and Pegoraro (2009).

This delicate situation may become even more complicated for two reasons. First, in many recent modeling strategies the persistent dynamics is captured through latent variables for which statistical tests are obviously not available given the lack of direct observations. Second, if the data generating process is stationary but close to non-stationarity, the finite sample bias of asymptotically efficient unconstrained estimators may be very large. For instance, there is an important literature on the bias of the parameters of an AR(p) model when a root is close to one [see, among others, Shaman and Stine (1988)].

In this paper we are mainly interested in the prediction problem. More precisely, the objective is to see how to treat the stationarity vs. non-stationarity trade-off and the finite sample bias problem in order to optimize the prediction performances of the models. We will focus, first, on the simple AR(1) model, because we believe that the important problems show up still in this simple setting which can be studied accurately, even if we will propose solutions which are easily extended to more general models. Then, we will also consider the case of a bivariate "near-cointegrated" model.

Finite sample distributions of the OLS estimator of ρ in the AR(1) model have been studied in a number of papers, mainly in the case $\mu = 0$ [see e.g. Evans and Savins (1981)]. In particular, the bias of order $\frac{1}{T}$, T being the number of observations, are derived in the AR(1) model [Kendall (1954), Marriott and Pope (1954)] and in AR(p) models [Shaman and Stine (1988)]. Here we will adopt a simulation analysis in order to get accurate descriptions of the finite sample distribution of the OLS estimator of ρ and we shall see that the bias of order $\frac{1}{T}$ is a bad approximation of the exact bias when ρ is close to 1, which is precisely the case we are interested in. We will also consider the median and the median bias, and we will concentrate on two sample sizes : T=160 and T=40 which are typical sizes when dealing, respectively, with quarterly and annual data. In these two cases, we will provide very good fitting of the mean bias and the median bias, based on quadratic splines.

Since it is natural to see whether bias corrected estimators are useful in terms of prediction, we will first define precisely four such estimators : the "Kendall" estimator (based on the Kendall's bias approximation), the Indirect Inference estimator [see Gourieroux and Monfort (1996) chapter 4, and Gourieroux, Renault and Touzi (2000)], the Bootstrap Estimator [see Hall (1997)] and the "Median-unbiased" estimator [see Andrews (1993)].

We will then compare these bias-corrected estimators to estimators taking into account both an unconstrained model and a unit root model. The pretest estimator is such an estimator but since its prediction performances are not clear from a practical point of view [see the conflicting conclusions in Stock (1996), Diebold and Kilian (2000) and B. Hansen (2007))], we will investigate the properties of the averaging estimators *à la* B. Hansen (2009) [see also B. Hansen (2007, 2008)].

The paper is organized as follows. In Section 2 we study the finite sample properties of the OLS estimator in an AR(1) model. In particular, we propose an accurate approximation of the mean bias and the median bias based on quadratic splines. In Section 3 we define the four "bias corrected" estimators: the Indirect Inference estimator, the Bootstrap estimator, the "Kendall" estimator and the "Median-unbiased" estimator. Section 4 introduces the class of averaging estimators and compare its prediction performances with those of the "bias corrected" estimators previously defined. Section 5 extends the study to a simple near-cointegrated bivariate model and Section 6 concludes.

2 Finite sample and asymptotic properties of the OLS estimator in an AR(1) model

Let us consider the model :

$$y_t = \mu(1-\rho) + \rho y_{t-1} + \varepsilon_t\,,\ t \in \{1, \ldots, T\}\,, \tag{1}$$

where the $\varepsilon_t' s$ are independently, identically distributed with $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = \sigma^2$. The initial value y_0 is fixed or random. The OLS estimator $\hat{\rho}_T$ of ρ is:

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_t(y_{t-1} - \bar{y})}{\sum_{t=1}^{T} (y_{t-1} - \bar{y})^2} \tag{2}$$

with $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}$.

It is well known that if $-1 < \rho < 1$, the asymptotic distribution of $\sqrt{T}(\hat{\rho}_T - \rho)$ is $N(0, 1-\rho^2)$, whereas if $\rho = 1$ the asymptotic distribution of $T(\hat{\rho}_T - 1)$ is non standard and function of a Brownian process [see Fuller (1976), Dickey and Fuller (1979, 1981)].

Moreover if ρ is smaller than 1 but close to 1, which the case of interest in this paper, the finite sample distribution of $\hat{\rho}_T$ remains far from a normal distribution even for large T. In order to tackle this problem many researchers have proposed a "local to unit root approach", that is a framework in which the true value of ρ depends on T and converges to 1 when $T \to \infty$. The results obtained heavily depend on several aspects of the retained setup : i) whether $\mu = 0$ or $\mu \neq 0$; ii) whether there is an intercept or not in the regression used to estimate ρ, i.e. whether $\hat{\rho}_T$ is taken equal to (2) or to $\sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2$, when $\mu = 0$; iii) the assumptions on the initial condition y_{0T}; iv) the rate of convergence of the true value ρ_T towards 1.

For instance, if we assume that $\mu = 0$ and $\hat{\rho}_T = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2$, the asymptotic behavior of $\hat{\rho}_T$ when T goes to infinity may still be quite different depending on the assumptions on y_{0T} and ρ_T. If y_{0T} is drawn in a fixed distribution and $\rho_T = 1 + \frac{c}{T}$ (with $c < 0$), the rate of convergence of $(\hat{\rho}_T - \rho_T)$ is $\frac{1}{T}$ and the asymptotic distribution of $T(\hat{\rho}_T - \rho_T)$ is a function of an Ornstein-Uhlenbeck process depending on c [see Phillips (1987)]. If we still assume that $\rho_T = 1 + \frac{c}{T}$ but y_{0T} is drawn in its unconditional distribution defined by $y_{0T} = \sum_{j=0}^{\infty} \rho_T^j \varepsilon_{T-j}$ [and therefore $V(y_{0T}) = O(T)$], an independent normal variable must be introduced in the asymptotic distribution [see Elliott (1999), Elliott and Stock (2001), Muller and Elliott (2003)]. If $T(1 - \rho_T) \to \infty$, that is ρ_T is "not too close to 1", for instance if $\rho_T = 1 + \frac{c}{T^\alpha}$ with $0 < \alpha < 1, c < 0$ and if $V(u_0^2) = o(T^{1/2})$ then $2[1 - \rho_T]^{-1/2}\ T^{1/2}(\hat{\rho}_T - \rho_T)$ converges in distribution to $N(0, 1)$ [see Giraitis and Phillips (2006), Phillips and Magdalinos (2007)], so we are back in a Gaussian asymptotic behavior. On the contrary if $T(1-\rho_T) \to 0$, that is if ρ_T is "very close to 1", and if y_{0T} is drawn in its unconditional distribution then $[2(1 - \rho_T)]^{-1/2} T^{1/2}(\hat{\rho}_T - \rho_T)$ converges in distribution to a Cauchy distribution [see Andrews and Guggenberger (2007)]. So, we see that the "near unit root" asymptotic results are very sensitive to initial conditions and to convergence rates, which have no concrete meanings for a practitioner. Even if we admit the usual assumption $\rho_T = 1 + \frac{c}{T}$, which provides non trivial asymptotic power for unit root tests, the asymptotic distribution still depends on the initial conditions and on the unknown value c.

Given this unclear practical message of the near unit root literature, we adopt a pragmatic solution based on simulation studies. More precisely, we assume that the ε_t's are independently distributed as $N(0, \sigma^2)$ and that $y_0 = \mu$. In this case the finite sample distribution of $\hat{\rho}_T$, given by (2), depends only on ρ (and not on μ and σ^2) since, from (1) we can equivalently write:

$$
\begin{aligned}
y_t &= \mu + \sigma z_t\,, \\
z_t &= \rho z_{t-1} + \eta_t\,,\ \eta_t \sim IIN(0,1)\,, \\
z_0 &= 0\,, \\
\text{and} \quad \hat{\rho}_T &= \frac{\sum_{t=1}^{T} z_t (z_{t-1} - \bar{z})}{\sum_{t=1}^{T} (z_{t-1} - \bar{z})^2}\,.
\end{aligned}
$$

Let us first consider the distribution of $\hat{\rho}_T$, for $\rho \in \{0.91, 0.95, 0.99\}$, with sample sizes $T = 160$

(see Figure 1) and $T = 40$ (see Figure 2). We can see the well known increasing left asymmetry of the distributions as far as ρ increases towards 0.99 [see, for instance, Evans and Savin (1981)]. Clearly, the distributions are more concentrated for $T = 160$ than for $T = 40$, but the non-normality does not seem to reduce when passing from $T = 40$ to $T = 160$.

	ρ	0.91	0.95	0.99
$T = 160$	*skew*	-0.89	-1.07	-1.35
	kurt	4.23	4.78	5.90
$T = 40$	*skew*	-1.05	-1.12	-1.20
	kurt	4.55	4.85	5.30

Table 1: Skewness (*skew*) and kurtosis (*kurt*) of the distribution of $\hat{\rho}_T$, computed from 5×10^5 simulations.

Table 1 gives the skewness and the kurtosis of the $\hat{\rho}_T$ distributions, computed from 5×10^5 simulations, and we can see that the negative skewness and the kurtosis increase with ρ, for given T. For $\rho = 0.91$ skewness ans kurtosis decrease with the sample size, for $\rho = 0.95$ they still slightly decrease but, for $\rho = 0.99$, they increase with T, stressing the specific behavior of the distributions around the values of ρ we are mainly interested in.

Let us now focus on the finite sample bias of $\hat{\rho}_T$. It is known, since Kendall (1954), that the bias of order $\dfrac{1}{T}$ is $-\dfrac{1+3\rho}{T}$, implying a downward bias of around -0.025 for $T = 160$ and 0.1 for $T = 40$, when ρ is close to 1. These biases are very large even for $T = 160$, since, for predicting purpose the behavior of ρ^h and $(\rho - 0.025)^h$, for ρ close to 1, are very different for large h. Kendall's formula is however an approximation and it is worth considering a more accurate estimation of the bias function :

$$b_T(\rho) = E(\hat{\rho}_T) - \rho \,,$$

based on 5×10^4 simulations.

This function is shown in Figure 3, for $T = 160$, and in Figure 4 for $T = 40$. It is seen from these figures that, for values of ρ close to 1, the bias is even much worse than the Kendall's approximation. For $\rho = 0.99$ the bias is approximately -0.034 instead of -0.025 when $T = 160$, and -0.13 instead of -0.1 when $T = 40$. Let us consider the case $T = 160$ and let us assume that the true value of ρ is 0.99. If we approximate the bias by -0.025, the expectation of $\hat{\rho}_T$ is evaluated as 0.975 whereas the true expectation is 0.956. The true mean reversion percentage of the prediction at horizon $h = 20$

(five years) is $100(1-0.99^{20}) \approx 18\%$, whereas this percentage evaluated at the approximated and the exact expectation of $\hat{\rho}_T$ are respectively $100(1-0.975^{20}) \approx 39\%$ and $100(1-0.956^{20}) \approx 59\%$. So, the consequences of the bias problem may be very severe in terms of predictions. We see that these bias functions are nonlinear in ρ and, in order to easily work with them, we approximate them by quadratic spline functions with a knot at $\rho = 0.9$. We obtain excellent fit in the domain of interest $\rho \in [0.4, 1]$ [see the supports of the distributions in Figure 1 and 2], as shown in Figures 3 and 4, using the following functions for the expectations $e_T(\rho) = E(\hat{\rho}_T)$:

$$e_{160}(\rho) = -0.010 + 0.996\rho - 0.013\rho^2 - 0.636(\rho - 0.9)^2 \mathbb{1}_{(\rho>0.9)}\,, \text{ for } T = 160\,, \tag{3}$$

$$e_{40}(\rho) = -0.057 + 1.039\rho - 0.113\rho^2 - 0.152(\rho - 0.9)^2 \mathbb{1}_{(\rho>0.9)}\,, \text{ for } T = 40\,. \tag{4}$$

It is also interesting to consider the median and the median bias of $\hat{\rho}_T$. Again, these biases are computed by simulation and given in Figure 5. As expected, since the distributions are negatively skewed, the median biases are smaller (in absolute value) than the mean biases. We also fitted a quadratic spline functions with one knot at 0.9 on these biases and they are obtained from the following fits for the median function $m_T(\rho)$:

$$m_{160}(\rho) = -0.0289 + 0.995\rho - 0.008\rho^2 - 0.533(r - 0.9)^2 \mathbb{1}_{(r>0.9)}\,, \text{ for } T = 160 \tag{5}$$

$$m_{40}(\rho) = -0.054 + 1.075\rho - 0.125\rho^2 - 0.238(r - 0.9)^2 \mathbb{1}_{(r>0.9)}\,, \text{ for } T = 40. \tag{6}$$

3 Bias-corrected estimators

The previous section showed that very large biases may appear both in the mean and in the median of $\hat{\rho}_T$. Since these biases may have an important impact at the prediction stage, it is important to carefully define various bias-corrected estimators.

3.1 The Indirect Inference estimator

Let us first consider the indirect inference estimator. Indirect inference is primarily designed to provide consistent estimators in models where the likelihood function is untractable and the method is based on an auxiliary tractable model. Nevertheless, indirect inference is also useful to remove the finite sample bias in models which are easily estimable. The idea is to propose the estimator :

$$\hat{\rho}_T^I = e_T^{-1}(\hat{\rho}_T)\,, \tag{7}$$

where $e_T(\rho)$ is the expectation function $e_T(\rho) = E_\rho(\hat{\rho}_T)$. Since $e_T(\hat{\rho}_T^I) = \hat{\rho}_T$, we obviously have :

$$E_\rho[e_T(\hat{\rho}_T^I)] = e_T(\rho)$$

and, therefore, $\hat{\rho}_T^I$ is e_T unbiased and, if e_T was linear, $\hat{\rho}_T^I$ would be exactly unbiased. In the general case, it can be shown that $\hat{\rho}_T^I$ is unbiased at order $\frac{1}{T}$ [see Gourieroux and Monfort (1996), Gourieroux, Renault and Touzi (2000)]. Moreover, this method has been very successful for removing bias both in time series context [Phillips and Yu (2005)] and panel context [Gourieroux, Phillips and Yu (2007)]. In the computation of e_T^{-1} we will use the quadratic spline approximations (3) and (4) which are easily inverted.

3.2 The Bootstrap and "Kendall" estimators

The bootstrap bias-corrected estimator is based on the "russian doll" principle [see Hall (1997) chap. 1]. Since the true bias $b_T(\rho) = e_T(\rho) - \rho$ is obviously unknown, the idea is to replace the unknown model by the estimated model, i.e. ρ by $\hat{\rho}_T$, and to replace $\hat{\rho}_T$ by the OLS estimator based on T pseudo observations drawn in the estimated model. In other words, $b_T(\rho)$ is replaced by $b_T(\hat{\rho}_T) = e_T(\hat{\rho}_T) - \hat{\rho}_T$ and, therefore, the bootstrap biais-corrected estimator is :

$$\begin{aligned} \hat{\rho}_T^B &= \hat{\rho}_T - [e_T(\hat{\rho}_T) - \hat{\rho}_T] \\ &= 2\hat{\rho}_T - e_T(\hat{\rho}_T)\,. \end{aligned} \tag{8}$$

Note that, contrary to the indirect inference estimator, $\hat{\rho}_T^B$ is not necessarily exactly unbiased if $e_T(\rho)$ is linear. Indeed, if $e_T(\rho) = \gamma_1 + \gamma_2\rho$ (say) we have :

$$\begin{aligned} E(\hat{\rho}_T^B) &= 2(\gamma_1 + \gamma_2\rho) - \gamma_1 - \gamma_2(\gamma_1 + \gamma_2\rho) \\ &= \gamma_1(1-\gamma_2) + \gamma_2(2-\gamma_2)\rho\,, \end{aligned}$$

which is equal to ρ only if $\gamma_2 = 1$. However, it can be shown [see Gourieroux, Renault and Touzi (2000)] that $\hat{\rho}_T^B$ is also unbiased at order $\frac{1}{T}$.

The bootstrap principle can also be applied to the Kendall approximation of the bias $-\frac{1+3\rho}{T}$ and we get the "Kendall" estimator:

$$\begin{aligned} \hat{\rho}_T^K &= \hat{\rho}_T - \left(-\frac{1+3\hat{\rho}_T}{T}\right) \\ &= \left(1 + \frac{3}{T}\right)\hat{\rho}_T + \frac{1}{T}\,. \end{aligned} \tag{9}$$

3.3 The "Median-unbiased" estimator

Finally, we can also apply the principle of the indirect inference method to the median function instead of the mean function. That is, we can define the estimator :

$$\hat{\rho}_T^M = m_T^{-1}(\hat{\rho}_T)\,, \tag{10}$$

where $m_T(\rho)$ is the median function. This estimator, proposed by Andrews (1993), is exactly median unbiased if m_T is increasing. Indeed, since m_T^{-1} is increasing, the median of $m_T^{-1}(\hat{\rho}_T)$ is equal to $m_T^{-1}[m_T(\rho)] = \rho$.

In practice, the various estimators will be bounded at 1. Note that the bounded median-unbiased is still median-unbiased since the median of $\hat{\rho}_T^M$ is ρ, smaller than 1, and an upper truncation above the median does not change it.

In Figures 6 and 7 we show, for $T = 160$ and $T = 40$ respectively, the functions $\hat{\rho}_T^I(\rho)$, $\hat{\rho}_T^B(\rho)$, $\hat{\rho}_T^K(\rho)$ and $\hat{\rho}_T^M(\rho)$ providing the corrections of the OLS estimators, in the range $\rho \in [0.4, 1]$. We also consider a zoom of these functions, in the range $\rho \in [0.8, 1]$, in Figures 8 and 9. We see that the indirect inference provides the more important correction, the lowest one being the mean unbiased estimator. The bootstrap correction is similar to the indirect inference except for high values of ρ, and the Kendall's correction is between the mean correction and the median correction (except for $T = 40$ and high values of ρ). We also see that the truncation at one occurs much more frequently for $T = 40$ than for $T = 160$, since the bias correction is larger.

4 Prediction performances of the bias-corrected estimators and of the averaging estimator

Since the main objective of this paper is to find improvements in prediction of persistent time series, we are going to investigate the forecast performance of the estimators introduced in the previous section. Moreover, we will include in the comparison the averaging estimator proposed by B. Hansen (2009). Indeed, using "local to unit root asymptotics" techniques, B. Hansen (2009) showed that this kind of estimators have nice properties in term of short-term predictions since the best weighting of the unconstrained and unit-root constrained estimator is strictly between 0 and 1. However, this study only considered the case of one-step ahead forecast associated to a scalar Gaussian AR(p) model. Moreover, the optimal weighting depends on the rate of convergence to non-stationarity which is difficult to evaluate in practice. In this paper, using simulation techniques, we will evaluate this averaging estimator for a general h-step ahead forecast both for scalar and bivariate (see Section 5) autoregressive models. In our setup the class of averaging estimators is :

$$\hat{\rho}_T^A(\lambda) = (1-\lambda) + \lambda \hat{\rho}_T, \;\; 0 \leq \lambda \leq 1 \,. \tag{11}$$

The methodology is as follows. For a given value of ρ and of T (160 or 40), we draw $S = 10^5$ simulated paths in the AR(1) model of length 180 if $T = 160$ and 45 if $T = 40$ (μ and σ are taken equal to 1) in order to keep 20 and 5 observations, respectively, for the out-of-sample forecast exercise. For each simulated path we compute the OLS estimator (with an intercept) $\hat{\rho}_T$, the estimators $\hat{\rho}_T^I, \hat{\rho}_T^B, \hat{\rho}_T^M$ and the class of averaging estimators $\{\hat{\rho}_T^A(\lambda), \lambda \in [0,1]\}$, from the simulations $y_1^s, \ldots, y_T^s$. We then use each estimator $\hat{\rho}_T^i$ [with $i = I, B, M, A$ denoting, respectively, the Indirect Inference, Bootstrap, "Median-unbiased" and Averaging estimator] to predict $y_{T+1}, \ldots, y_{T+h}, \ldots, y_{T+H}$. We get $\hat{y}_{Ti}^s(h) = \hat{\mu}_T + (\hat{\rho}_T^i)^h (y_{T+h}^s - \hat{\mu}_T)$ (with $H = 20$ if $T = 160$, and $H = 5$ if $T = 40$) and we compute the root mean square forecast error, for each estimator $\hat{\rho}_T^i$ and for $h = 1$ and $h = H$, by :

$$RMSFE_i(h) = \left\{ \frac{1}{S} \sum_{s=1}^{S} [\hat{y}_{Ti}^s(h) - y_{T+h}^s]^2 \right\}^{1/2} .$$

Finally, we calculate the ratio of this $RMSFE_i(h)$ to the root mean square forecast error obtained with the true value of ρ. The horizons $H = 5$, for $T = 40$, and $H = 20$, for $T = 160$, would correspond to a five-year horizon for annual and quarterly data respectively.

We also compute in-sample characteristics, in particular the bias and the in-sample root mean squared error $RMSE_i$ associated to the estimation of ρ (and μ) and for each estimation method $i = I, B, M, A$. Note that, for sake of clarity of the figures, we did not considered the Kendall estimator.

Let us first consider the case $T = 160$ and $\rho = 0.99$ (Figures 10 to 13). Among bias-corrected estimators, the best correction is obtained by $\hat{\rho}_T^I$, since the -0.034 bias of the OLS is reduced to -0.008. With regard to the $\hat{\rho}_T^A(\lambda)$ class, the bias is obviously the oblique line between 0.01 and -0.034. The RMSE (Figure 11) is minimal for the $\hat{\rho}_T^A(\lambda)$ estimator corresponding to $\lambda \simeq 0.15$ and the optimal RMSE is more than five time smaller than that associated to the other estimators. The OLS is particularly bad with a RMSE nine times larger than the optimal one. As far as the RMFSE ratios are concerned (Figures 12 and 13) they are optimal for the $\hat{\rho}_T^A(\lambda)$ estimator corresponding approximately to $\lambda = 0.25$ for $h = 1$ and $h = 20$. Moreover, for $h = 20$, the percentage of increase of the RMSFE (compared to the unfeasible one corresponding to the true value of ρ) is about four times smaller for the optimal $\hat{\rho}_T^A(0.25)$ one compared to those obtained from $\hat{\rho}_T^I, \hat{\rho}_T^\beta$ and $\hat{\rho}_T^M$ and six times smaller than the OLS one.

For $T = 40$ and $\rho = 0.99$ (Figures 14 to 17), the results are similar except that the optimal weight is approximately 0.05 for the in-sample RMSE and 0.1 for the out-of-sample RMSFE. In other words, when $T = 40$ and $\rho = 0.99$, the OLS is so bad that its weighted is smaller than in the case $T = 160$. In any case, the optimal $\hat{\rho}_T^A(\lambda)$ is by far the best estimator both in-sample and out-of-sample.

For $T = 160$ and $\rho = 0.95$ (Figures 18 to 21), again, among bias-corrected estimators, the best correction is obtained for $\hat{\rho}_T^I$ (Figure 18). All bias-corrected estimators dominate in-sample (Figure 19) the OLS ($\lambda = 1$) and the constrained estimator ($\lambda = 0$), but the best in-sample estimator is $\hat{\rho}^A(\lambda)$ with $\lambda \simeq 0.55$. The best predictions are also obtained for $\hat{\rho}^A(\lambda)$, with $\lambda \simeq 0.55$ when $h = 1$ (Figure 20), and with $\lambda \simeq 0.8$ when $h = 20$ (Figure 21). The OLS ($\lambda = 1$) is dominated by the bias-corrected estimators for short-term prediction ($h = 1$), while the converse it true for the long-term forecast ($h = 20$). In any case, the averaging estimator $\hat{\rho}_T^A(\lambda)$ remains the best solution, with a weight between 0.55 and 0.8, giving, as expected, much more importance to the OLS than when $\rho = 0.99$.

The last case is $T = 40$ and $\rho = 0.95$ (Figures 22 to 25). The bias-corrected estimators dominate in-sample the OLS (Figure 23) but not the constrained estimator. The averaging estimator is optimal in-sample (with $\lambda \simeq 0.2$) and out-of-sample (with $\lambda \simeq 0.25$ for $h = 1$, $\lambda \simeq 0.30$ for $h = 20$). The OLS is dominated by all the bias-corrected estimators in-sample and out-of-sample. The optimal weight in $\hat{\rho}_T^A(\lambda)$ is larger, compared to the case $\rho = 0.99$, but remains much below 0.5.

An obvious global conclusion is that, in all the situations, the averaging estimator $\hat{\rho}_T^A(\lambda)$ is by far the best. Moreover, for a given T and ρ, the optimal λ does not change too much when considering the in-sample behavior or the short-term and the long-term prediction behavior.

These results suggest that, in practice, we could adopt a pragmatic averaging estimator strategy, when facing the choice among two kinds of models, one estimated without constraints the other one with unit root or cointegration constraints. If we denote by $\underline{y_T} = (y_1, \dots, y_T)$ the observations and by $g(\underline{y_t})$, $t \in \{1, \dots, T\}$ a variable of interest that we want to predict accurately at horizon h, the strategy we suggest is as follows :

- define a sequence of increasing windows $\{1, \ldots, t\}$, with $t \in \{t_0, \ldots, T-h\}$;
- for each t compute the unconstrained estimator $\hat{\theta}_t^{(u)}$ and the constrained estimator $\hat{\theta}_t^{(c)}$ of the parameter θ;
- for each t compute the class of averaging estimators $\hat{\theta}_t(\lambda) = (1-\lambda)\hat{\theta}_t^{(c)} + \lambda\hat{\theta}_t^{(u)}$, the corresponding predictions $\hat{g}_{t,h}(\lambda)$ of $g(\underline{y_{t+h}})$ and the prediction error $[g(\underline{y_{t+h}}) - \hat{g}_{t,h}(\lambda)]$;
- compute $Q_T(\lambda, h) = \sum_{t=t_0}^{T-h}[g(\underline{y_{t+h}}) - \hat{g}_{t,h}(\lambda)]^2$;
- calculate $\lambda^*(h) = argmin_{\lambda\in[0,1]}Q_T(\lambda, h)$;
- compute $\hat{\theta}_T(\lambda^*(h))$.

5 A example of bivariate near-cointegrated model

In order to give an example of the flexibility of our averaging estimator strategy, let us consider three bivariate data generating processes :

$$\begin{aligned} y_{1t} &= (1-\rho) + \rho y_{1,t-1} + \varepsilon_{1t} \quad (12) \\ y_{2t} &= 2y_{1t} + \varepsilon_{2t} \quad (13) \end{aligned}$$

with $\rho \in \{0.97, 0.98, 0.99\}$, where ε_{1t} and ε_{2t} are independent standard Gaussian white noises.

The bivariate process $y_t = (y_{1t}, y_{2t})'$ is "nearly cointegrated" since in the limit case $\rho = 1$, both processes y_{1t} and y_{2t} are $I(1)$ and $(y_{1t} - 2y_{2t})$ is stationary.

The unconstrained model is the VAR(1) defined by :

$$y_t = \nu + Ay_{t-1} + \eta_t \, , \quad (14)$$

and the unconstrained estimators of ν and A, denoted by $\hat{\nu}_T^{(u)}$ and $\hat{A}_T^{(u)}$ are just the OLS estimators. The constrained model is the error correction model imposing one cointegration relationship, namely :

$$\Delta y_t = \mu + \alpha(y_{1,t-1} - \beta y_{2,t-1}) + \xi_t \, ,$$

with $\mu = (\mu_1, \mu_2)', \alpha = (\alpha_1, \alpha_2)'$ where the estimator $\hat{\beta}_T$ of β is obtained by regressing y_{1t} on y_{2t}, and the estimators $(\hat{\mu}_{1T}, \hat{\alpha}_{1T})$ [resp. $(\hat{\mu}_{2T}, \hat{\alpha}_{2T})$] of (μ_1, α_1) [resp. (μ_2, α_2)] are obtained by regressing Δy_{1t} [resp. Δy_{2t}] on $(1, y_{1,t-1} - \hat{\beta}_T y_{2,t-1})$.

So the constrained estimators of ν and A are :

$$\hat{\nu}_T^{(c)} = \hat{\mu}_T \text{ and } \hat{A}_T^{(c)} = I + \hat{\alpha}_T(1, -\hat{\beta}_T)\,,$$

and the class of averaging estimators is :

$$\begin{aligned}\hat{\nu}_T(\lambda) &= (1-\lambda)\hat{\nu}_T^{(c)} + \lambda\hat{\nu}_T^{(u)}\,,\\ \hat{A}_T(\lambda) &= (1-\lambda)\hat{A}_T^{(c)} + \lambda A_T^{(u)},\ \ 0 \leq \lambda \leq 1.\end{aligned}$$

The class of predictions of y_{T+h} at T using the averaging estimators is :

$$\hat{y}_{T,h}(\lambda) = [I - \hat{A}_T(\lambda)]^{-1}[I - \hat{A}_T^h(\lambda)]\hat{\nu}_T(\lambda) + \hat{A}_T^h(\lambda)y_T\,. \tag{15}$$

Using the strategy described in the previous section we can compute the RMFSE ratios for the predictions of $y_{1,T+h}$ and $y_{2,T+h}$, i.e. the ratios of the RMSFE to the RMSFE obtained with the true values of the parameters.

Again the number of simulations is 10^5. In Figures 26 and 27 we consider the RMSFE ratios for $y_{1,T+h}$ and $y_{2,T+h}$, respectively, in the case $T = 160$ and $h = 1$, with $\rho \in \{0.97, 0.98, 0.99\}$. For $y_{1,T+h}$ the optimal values of λ are approximately 0.55 for $\rho = 0.97$, 0.45 for $\rho = 0.98$, and 0.3 for $\rho = 0.99$, whereas for $y_{2,T+h}$ these values are, respectively, 0.7, 0.65 and 0.55. So, for given values of ρ, the optimal values of λ are not exactly the same for $y_{1,T+h}$ and $y_{2,T+h}$ and therefore, there is some trade-off between these variables.

Figures 28 and 29 consider the same situations but for long term predictions ($h = 20$). Here, the optimal value of λ is approximately the same for $y_{1,T+h}$ and $y_{2,T+h}$, namely 0.8 for $\rho = 0.97, 0.65$ for $\rho = 0.98$ and 0.5 for $\rho = 0.99$. Thus, we see that for a given practical situation where T and ρ are fixed, the optimal value of λ must be fine tuned according to the particular forecasting horizon we want to put forward. It is clear that, we could also consider a criterion $Q_T(\lambda)$ (say), based on averaging $Q_T(\lambda, h)$ over h, selecting a weighting parameter λ^* independent from the forecasting horizon.

Figures 30 to 33 are the equivalent to Figures 26 to 29 but for the case $T = 40$. For $h = 1$, the optimal value of λ is approximately 0.3 for $y_{1,T+h}$ and any ρ, whereas it is approximately 0.5 for $y_{2,T+h}$ and any ρ. For $h = 5$, the optimal value of λ for $y_{1,T+h}$ it decreases from 0.4 to 0.3 when

ρ moves from 0.97 to 0.99 and, for $y_{2,T+h}$ it decreases from 0.45 to 0.40. So, the trade-off is much less important than for the case $T = 160$.

6 Concluding remarks

We have compared the prediction performances of various estimators when high persistence characterizes the time series of interest. The bias-corrected estimators like the Indirect Inference estimator, the Bootstrap estimator, the "Kendall" estimator and the "Median-unbiased" estimator, generally improve the performances of the OLS estimator, but all these estimators are dominated by the optimal averaging estimator. B. Hansen (2009)'s work gives a theoretical explanation of this phenomenon at least for short-term (one-step ahead) prediction and for a simple univariate AR(1) model. It would be interesting to investigate further the theoretical properties of the averaging estimator in a more general framework. However, in a multivariate framework, there is obviously a very large number of paths from an unrestricted model to a model restricted by some unit roots or cointegration constraints and, even for a particular path, the results are likely to depend on the converge rate. So, in practice, a pragmatic approach based on simulation results like the one proposed in this paper, seems, for the moment, an encouraging alternative which can be used in an univariate and multivariate setting and for a general autoregressive order. Moreover, this kind of approach, based on constrained modelling of the conditional expectation, could a priori be extended to the case where constraints also operate on the conditional variance-covariance matrix. In other words, the strategy described in this paper could also be applied to volatility persistence. Another extension could consists in making endogenous the choice between the unconstrained and the constrained estimator, using an hidden Markov chain (possibly non-homogeneous).

R E F E R E N C E S

Andrews, D.W.K. (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models", Econometrica, 61, 1, 139-165.

Andrews, D.W.K., Guggenberger, P. (2007): "Asymptotics for stationary very nearly unit root processes", Cowles Foundation D.P. n. 167.

Cochrane, J., and Piazzesi, M., (2008): "Decomposing the Yield Curve", Graduate School of Business, University of Chicago, Working Paper.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., (1992a): "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors", Journal of Econometrics, 53, 323-343.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., (1992b): "Integration Versus Trend Stationary in Time Series", Econometrica, 60(2), 423-433.

Dickey, D.A., Fuller, W.A. (1979): "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American Statistical Association, 74, 427-431.

Dickey, D.A., Fuller, W.A. (1981): "Likelihood ratio statistics for autoregressive time series with unit root", Econometrica, 49, 1057-1072.

Diebold, F.X., Lopez, J.A. (2000): "Unit-root tests are useful for selecting forecasting models", Journal of Business and Economic Statistics, 18, 265-273.

Elliott, G. (1999): "Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution", International Economic Review, 40, 767-783.

Elliott, G., Rothenberg, T.J., Stock, J.H., (1996): "Efficient Tests for an Autoregressive Unit Root", Econometrica, 64(4), 813-836.

Elliott, G., Stock, J.H. (2001): "Confidence intervals for autoregressive coefficients near one", Journal of Econometrics, 103, 155-181.

Elliott, G., and Muller, U. (2003): "Tests for unit roots and the initial conditions", Econometrica, 71, 1269-1286.

Evans, G.B.A., Savin, E. (1981): "Testing for unit roots I", Econometrica, 49, 3, 753-779.

Fuller, W.A. (1976): "Introduction to statistical time series", Wiley.

Giraitis, L., Phillips, P.C.B. (2006): "Uniform limit theory for stationary autoregressions", Journal of Time Series Analysis, 27, 51-60.

Gourieroux, C., Monfort, A. (1996): "Simulation based econometric methods", Oxford University Press.

Gourieroux, C., Renault, E., Touzi, N. (2000): "Calibration by simulation for small sample's bias correction", in Simulation-Based Inference in Econometrics, Mariano, Schuermann, Weeks eds, Cambridge University Press.

Gourieroux, C., Phillips, P.C.A., Yu, J. (2007): "Indirect Inference of Dynamic Panel Models", Journal of Econometrics, forthcoming.

Hall, P. (1997): "The Bootstrap and Edgeworth expansion", Springer Verlag.

Hansen, B., (2007): "Least Square Model Averaging", Econometrica, 75(4), 1175-1189.

Hansen, B., (2008): "Least Square Forecast Averaging", Journal of Econometrics, 146, 342-350.

Hansen, B., (2009): "Averaging Estimators for Autoregressions with a Near Unit Root", Journal of Econometrics, forthcoming.

Jardet C., Monfort A., Pegoraro, F. (2009): "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth", working paper Banque de France.

Kendall, M.G. (1954): "A note on bias in the estimation of autocorrelation", Biometrika, 41, 403-404.

Magdalinos, T., Phillips, P.C.B. (2007): "Limit theory for moderate deviations from a unit root", Journal of Econometrics, 136, 115-130.

Marriott, F.H.C., Pope, J.A. (1954): "Bias in the estimation of autocorrelations", Biometrika, 41, 390-402.

Ng, S., Perron, P., (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", Econometrica, 69(6), 1519-1554.

Phillips, P.C.B. (1987): "Towards a unified asymptotic theory for autoregression", Biometrika, 74, 535-547.

Phillips, P.C.B., Yu, J. (2009): "Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance", Handbook of Financial Time Series, 497-530.

Shaman, P., Stine, P. (1988): “The bias of autoregressive coefficient estimators”, Journal of the American Statistical Association, 83, 403, 842-848.

Schwert, G.W., (1989): ”Test for unit roots: A Monte Carlo investigation”, Journal of Business and Economic Statistics, 7, 147-158.

Stock, J.H. (1996): “VAR, error correction and pretest forecast at long horizons”, Oxford Bulletin of Economics and Statistics, 58, 685-701.

FIGURE 1: PDF of the OLS Estimators of Rho , T=160
Rho : .91(short dashes), .95(dashes), .99(solid line)

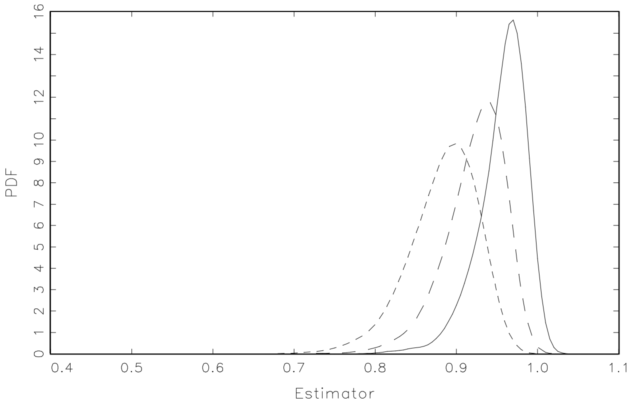

FIGURE 2: PDF of the OLS Estimators of Rho , T=40
Rho : .91(short dashes), .95(dashes), .99(solid line)

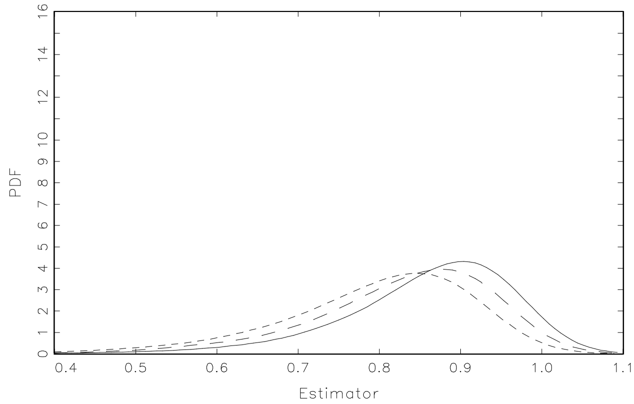

FIGURE 3: Bias of the OLS Estimator ,T=160
Exact(Solid Line) ,Spline Approximation(Dashed),
Kendall Approximation(Short Dashes)

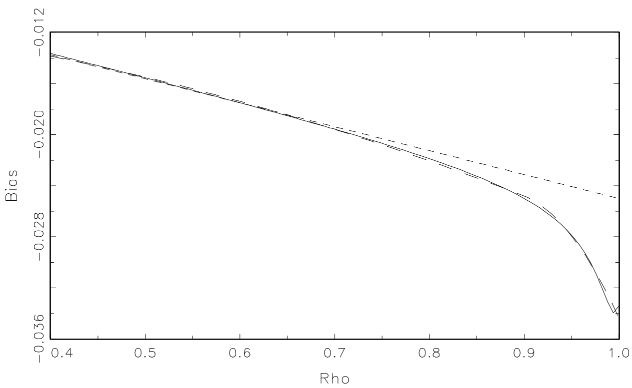

FIGURE 4: Bias of the OLS Estimator ,T=40
Exact(Solid Line) ,Spline Approximation(Dashed),
Kendall Approximation(Short Dashes)

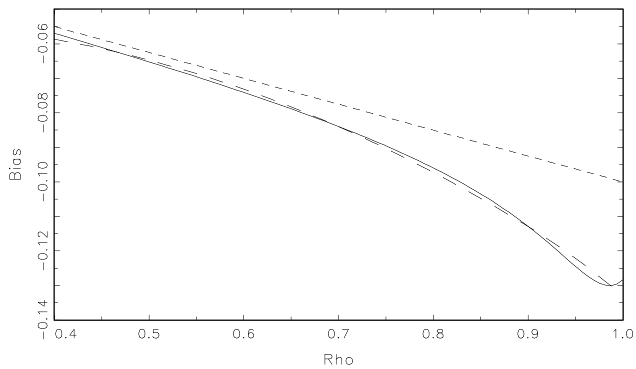

FIGURE 5 : Median Bias (solid lines),
Mean Bias(dashes) , Kendall's Mean bias(short dashes)
T=160(highest curves), T=40(lowest curves)

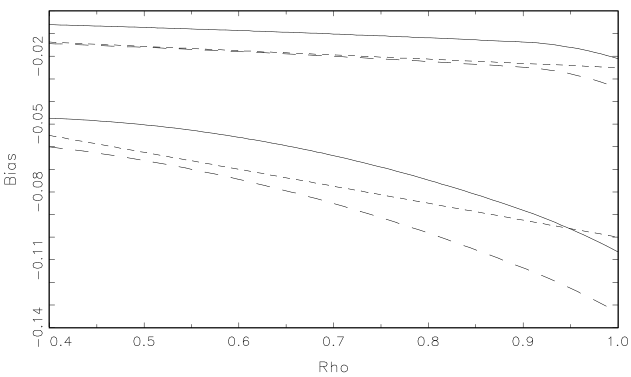

FIGURE 6 : Bias Corrected Estimators,T=160
Indirect Inference(Upper Solid Line),Bootstrap(Dashed),
Median(Short Dashes),Kendall(Dots and Dashes),OLS(Lower Solid Line)

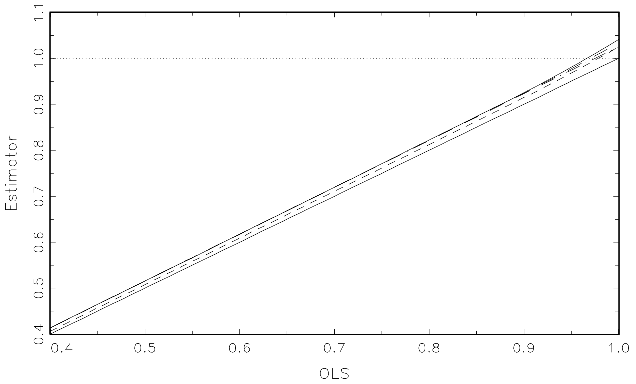

FIGURE 7 : Bias Corrected Estimators,T=40
Indirect Inference(Upper Solid Line),Bootstrap(Dashed),
Median(Short Dashes),Kendall(Dots and Dashes),OLS(Lower Solid Line)

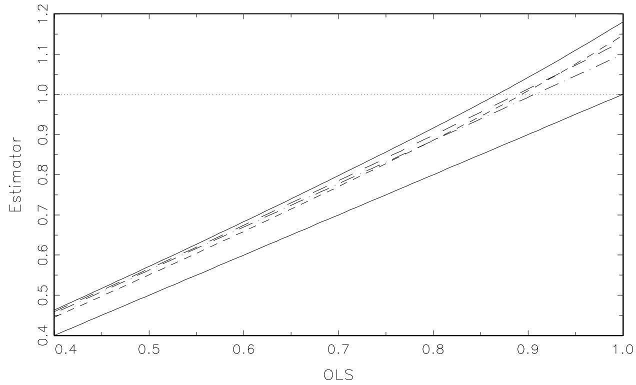

FIGURE 8 : Zoom, Bias Corrected Estimators ,T=160
Indirect Inference(Upper Solid Line),Bootstrap(Dashed),
Median(Short Dashes),Kendall(Dots and Dashes),OLS(Lower Solid Line)

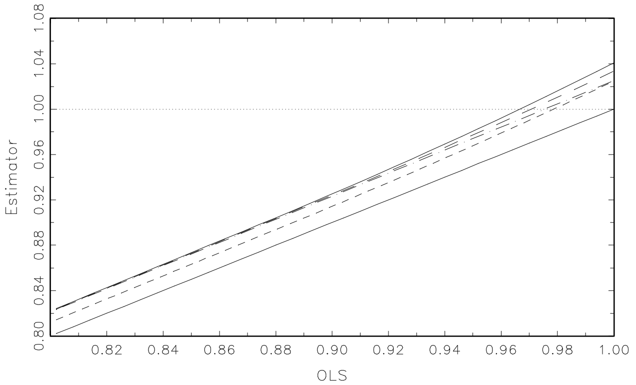

FIGURE 9 : Zoom, Bias Corrected Estimators ,T=40
Indirect Inference(Upper Solid Line),Bootstrap(Dashed),
Median(Short Dashes),Kendall(Dots and Dashes),OLS(Lower Solid Line)

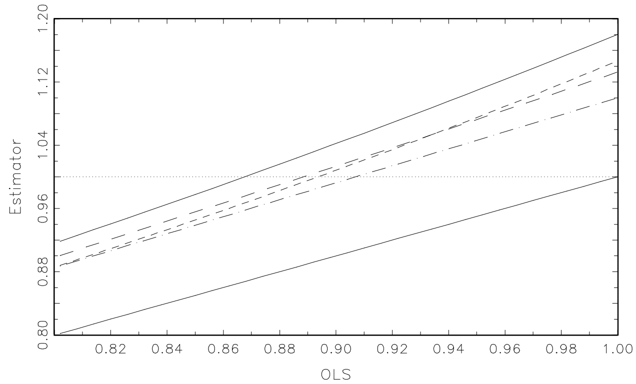

FIGURE 10 :Bias,T=160 ,Rho=.99
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

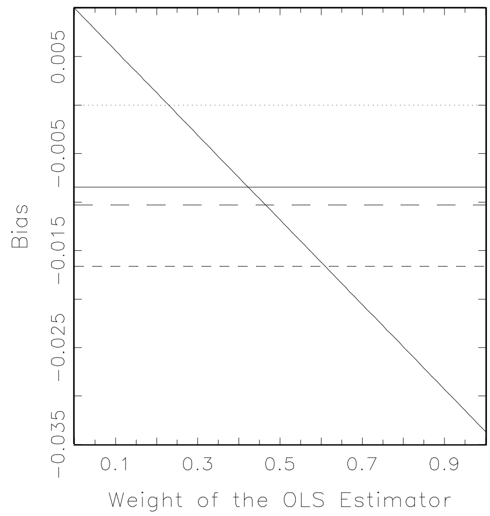

FIGURE 11 :RMSE,T=160 ,Rho=.99
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

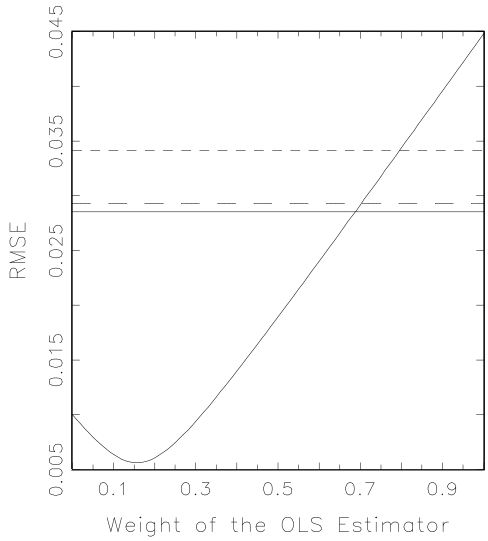

FIGURE 12:RMSFE Ratio,T=160,Rho=.99,h=1
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

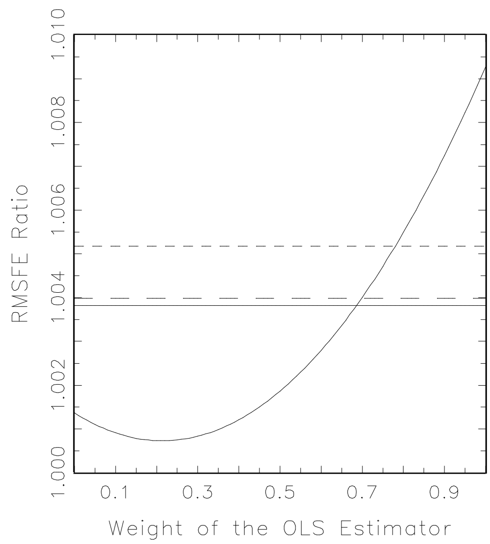

FIGURE 13:RMSFE Ratio,T=160,Rho=.99,h=20
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

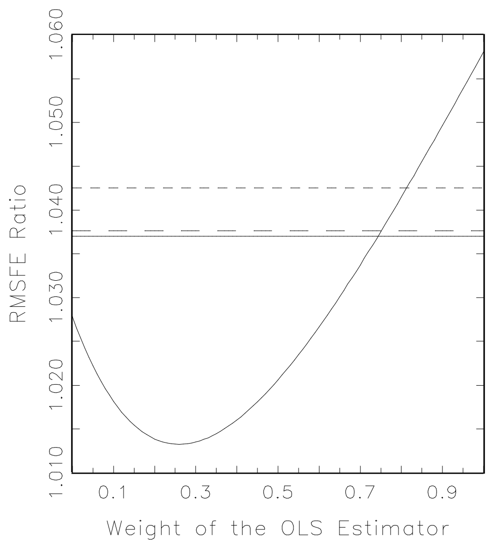

FIGURE 14 :Bias,T=40 ,Rho=.99
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

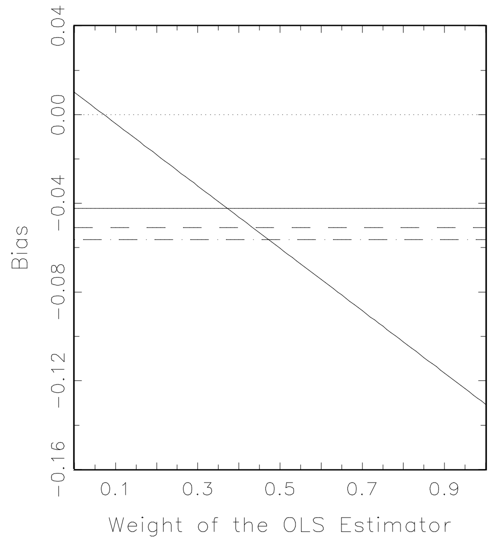

FIGURE 15 :RMSE,T=40 ,Rho=.99
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

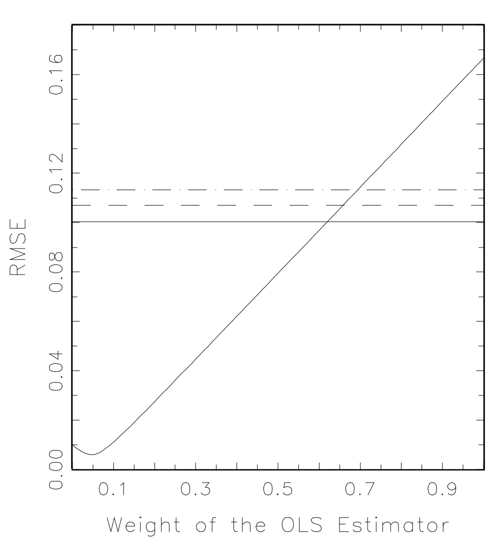

FIGURE 16:RMSFE Ratio,T=40,Rho=.99,h=1
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

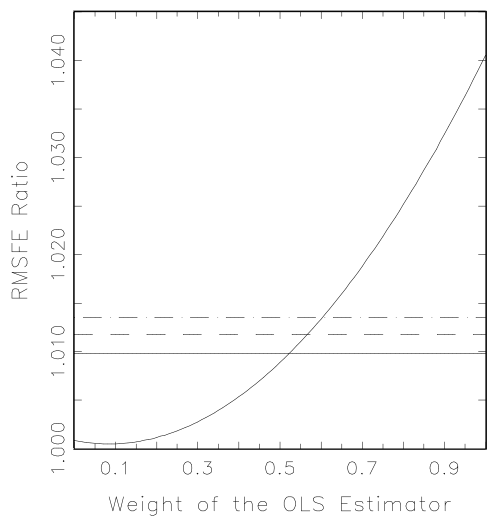

FIGURE 17:RMSFE Ratio,T=40,Rho=.99,h=5
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

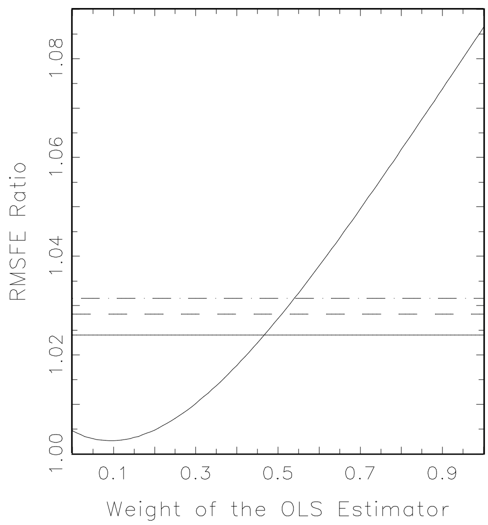

FIGURE 18 :Bias,T=160 ,Rho=.95
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

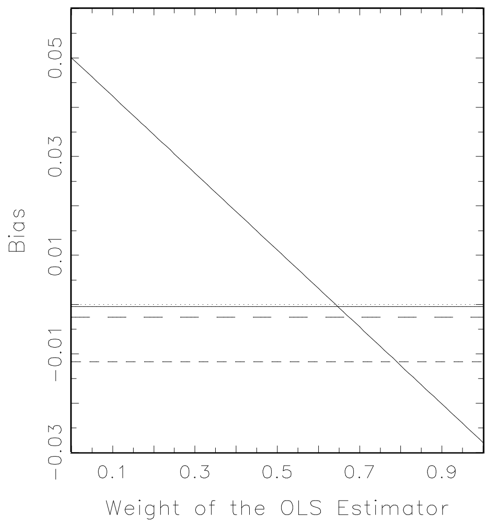

FIGURE 19 :RMSE,T=160 ,Rho=.95
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

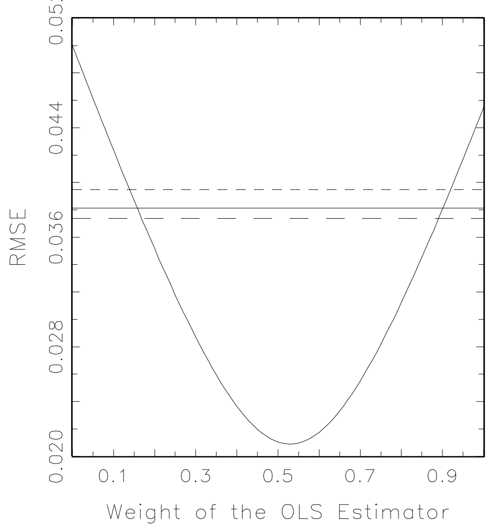

FIGURE 20:RMSFE Ratio,T=160,Rho=.95,h=1
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

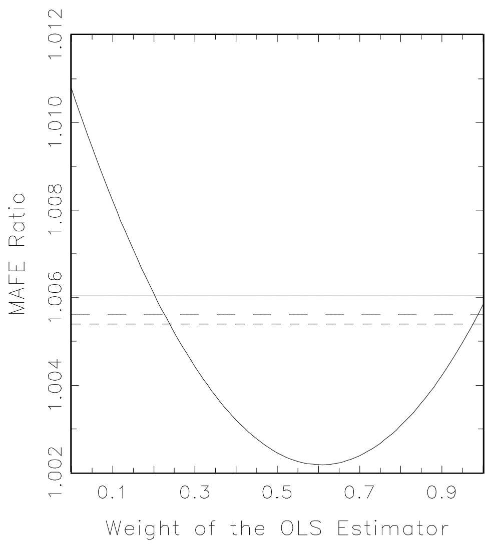

FIGURE 21:RMSFE Ratio,T=160,Rho=.95,h=20
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

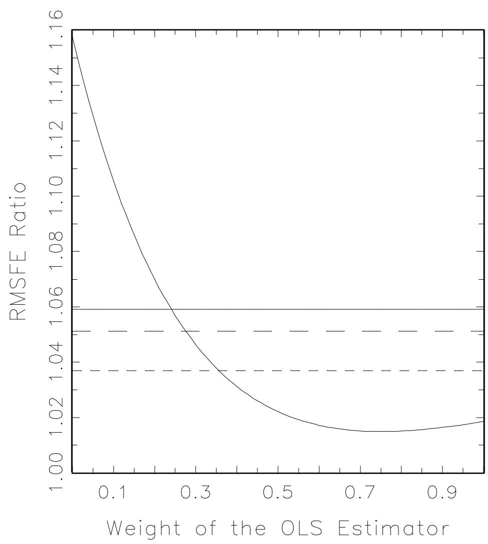

FIGURE 22 :Bias,T=40 ,Rho=.95
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

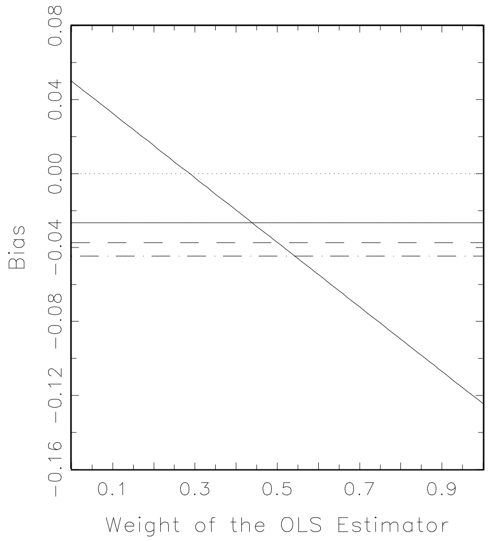

FIGURE 23 :RMSE,T=40 ,Rho=.95
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

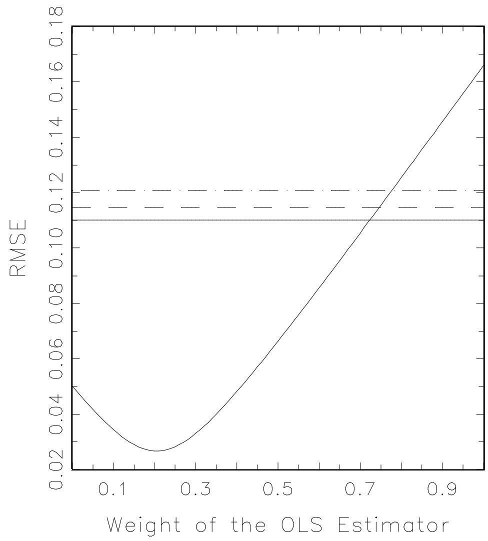

FIGURE 24:RMSFE Ratio,T=40,Rho=.95,h=1
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

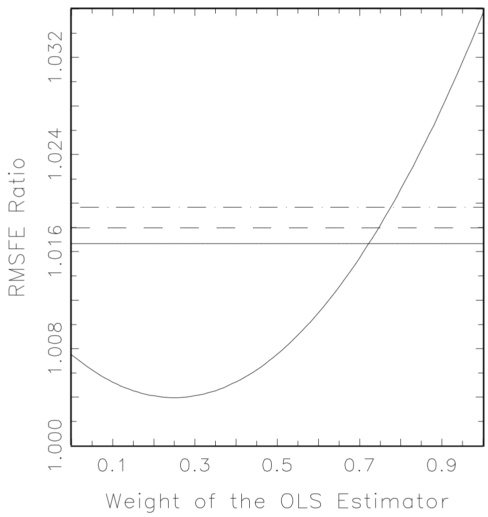

FIGURE 25:RMSFE Ratio,T=40,Rho=.95,h=5
RhoA(Solid Curve),RhoI(Solid Line)
RhoB(Dashes),RhoM(Short Dashes)

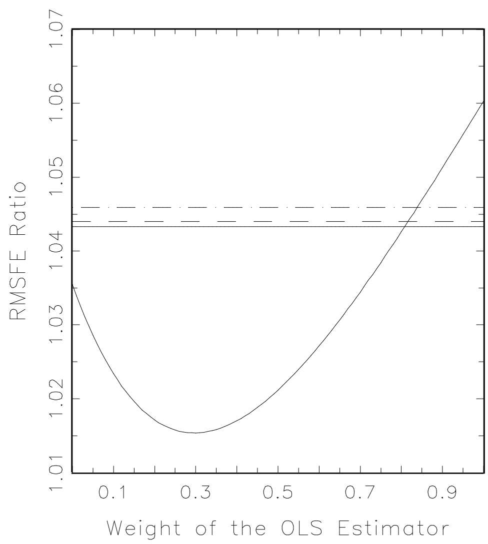

FIGURE 26 : RMSFE Ratio for Y1
T=160, h=1, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

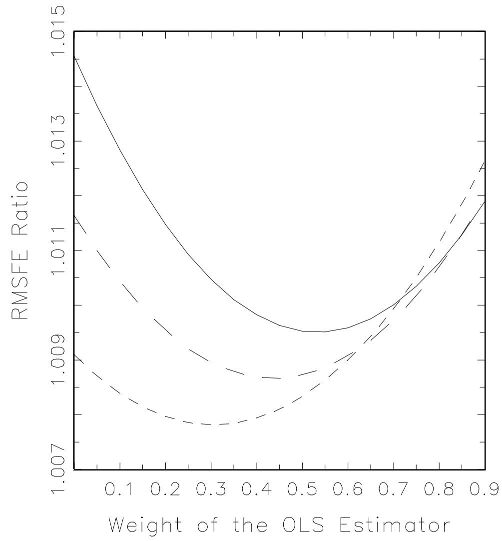

FIGURE 27 : RMSFE Ratio for Y2
T=160, h=1, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

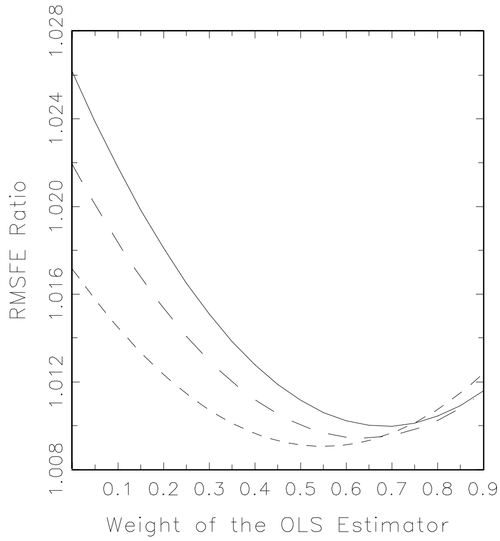

FIGURE 28 : RMSFE Ratio for Y1
T=160, h=20, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

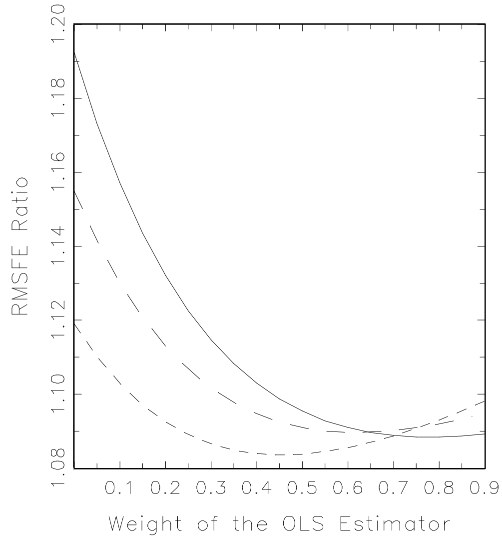

FIGURE 29 : RMSFE Ratio for Y2
T=160, h=20, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

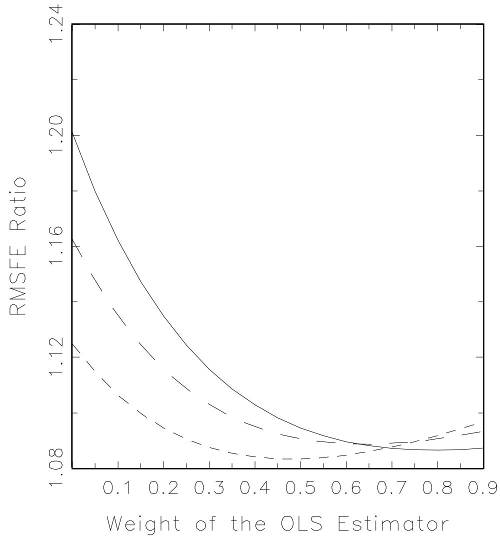

FIGURE 30 : RMSFE Ratio for Y1
T=40, h=1, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

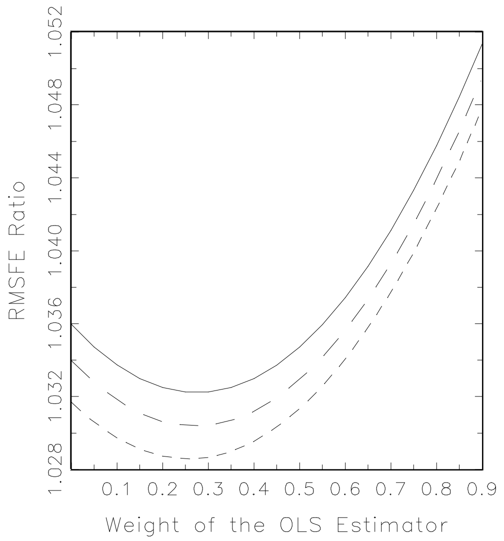

FIGURE 31 : RMSFE Ratio for Y2
T=40, h=1, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

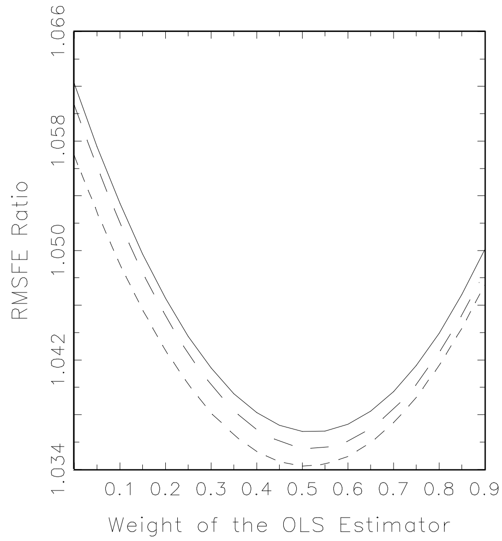

FIGURE 32 : RMSFE Ratio for Y1
T=40, h=5, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)

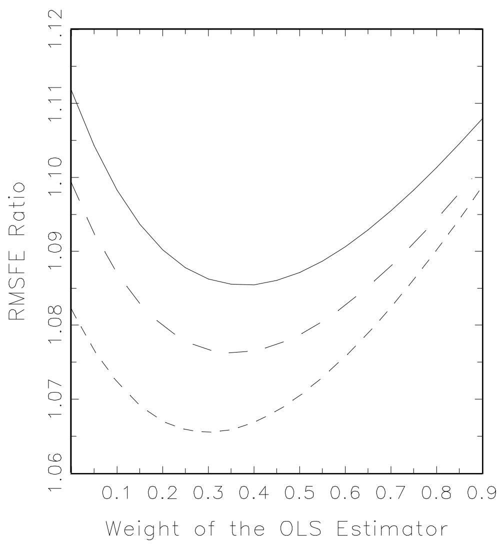

FIGURE 33 : RMSFE Ratio for Y2
T=40, h=5, Rho=.97(solid)
Rho=.98(dashes), .99(short dashes)