Persistence, Bias, Prediction and Averaging Estimators

Caroline JARDET [1]
Banque de France

Alain MONFORT [2]
CREST, Banque de France
and Maastricht University

Fulvio PEGORARO[3]
Banque de France, CREST and HEC Lausanne

First version : May, 2009. This version: July, 2011.

Abstract

Persistence, Bias, Prediction and Averaging Estimators

The objective of the paper is to propose a class of estimators dealing with the stationarity vs. non-stationarity trade-off and the finite sample bias problem when persistent variables are considered, in order to optimize the prediction performances of the models of interest, namely autoregressive models with or without conditional heteroscedasticity. We first study and compare, by means of Monte Carlo experiments, the in-sample fit and out-of-sample (one-step and multiple-step ahead) forecast performances obtained from several "biased-corrected" estimators with those coming from the class of estimators averaging unconstrained parameter estimates and parameter estimates with unit root or cointegration constraints [B. Hansen (2010)]. For both univariate and bivariate specifications, we find that in-sample and out-of-sample prediction performances are strongly in favor of this class of averaging estimators. The second part of the analysis studies a more general setting where the large level of persistence is generated by both the conditional mean and the time-varying conditional variance of the variable of interest. More precisely, we show how to generalize this averaging estimator approach to the problem of estimating long-term quantiles, or Value-at-Risk (*VaR*), in presence of persistence and conditional heteroscedasticity. We find that our proposed estimation strategy may dramatically reduce the estimation uncertainty (RMSEs) of the *VaR* level.

Keywords : persistence, conditional heteroscedasticity, bias correction, averaging estimator, estimation uncertainty, Value-at-Risk.

JEL classification : C52, C53.

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

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

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

1 Introduction

It is well known that many macroeconomic variables are persistent, in the sense that their dynamics feature high serial correlations. Examples of such variables are interest rates, inflation rates, exchange rates or price-dividend ratios [see, among the others, Kozicki and Tinsley (2001a, 2001b), Rossi (2005), Cochrane (2006), Jardet, Monfort and Pegoraro (2010a) and Joslin, Priebsch and Singleton (2010)]. Other examples are mortality rates, which are basic variables in the literature of longevity risk [see Gourieroux and Monfort (2008) and the references therein]. The econometric modeling of persistent variables is challenging because, if the true dynamics is stationary but close to non-stationarity, non-stationarity tests may fail to reject it, and this may imply 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 a stationary but persistent alternative, 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 properties 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 (2010a, 2010b) for an application to interest rates]. More generally, when persistent variables are present, standard statistical tools do not provide a clear vision of the number and the nature of cointegration relationships, and this may imply severe consequences for macro-finance applications, as indicated by Cochrane and Piazzesi (2008) and Jardet, Monfort and Pegoraro (2010a) [JMP (2010), hereafter].

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 estimators of the parameters of an AR(p) model when a root is close to one [see, among others, Shaman and Stine (1988)]. In particular, finite sample distributions of the OLS estimator of the autoregressive parameter ρ (say) in the AR(1) model have been studied in a number of papers, mainly in the case of a constant term $\mu = 0$ [see e.g. Evans and Savins (1981)]. In addition, the bias of order $\frac{1}{T}$, T being the number of observations, has been derived for AR(1) [Kendall (1954), Marriott and Pope (1954)] and AR(p) models [Shaman and Stine (1988)].

The objective of this paper is to propose a class of estimators dealing with the stationarity vs. non-stationarity trade-off and the finite sample bias problem (both induced by persistence) in order to optimize the multiple-step ahead prediction performances of models of interest, namely autoregressive processes with or without conditional heteroscedasticity.

We first work with homoscedastic AR(1) and VAR(1) models and our analysis is based on the following steps. We start with a simulation analysis in order to get accurate descriptions of the finite sample distribution of the OLS estimator of ρ and we detect 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 also consider the median bias, and we focus on two sample sizes : T=160 and T=40, which are typical sizes when dealing, respectively, with quarterly and annual macroeconomic data. In these two cases, we provide a very good fitting of the mean bias and the median bias, based on quadratic splines. Given the large and nonlinear downward bias affecting the OLS estimator, we study whether "bias-corrected" estimators are useful to forecast persistent variables. We consider the "Kendall" estimator (based on the Kendall's bias approximation), the Indirect Inference estimator [see Gourieroux, Monfort and Renault (1993), Gourieroux and Monfort (1996) chapter 4, and Gourieroux, Renault and Touzi (2000)], the Bootstrap Estimator [see Hall (1997)] and the "Median-unbiased" estimator [see Rudebusch (1992), Andrews (1993)].

Then, we compare the prediction performances of these bias-corrected estimators with those of estimators taking into account both the estimates of an unconstrained model and a unit root or a cointegrated model. More precisely, we investigate the usefulness of the averaging estimators *à la* B.

Hansen (2010) for multiple-step ahead forecasts [see also B. Hansen (2007, 2008)]. The basic idea of this class of estimators is to consider weighted averages of unconstrained parameter estimates (for instance, OLS-based estimates) and of parameter estimates taking into account constraints of non-stationarity (unit roots) or of cointegration. B. Hansen (2010) has shown, in a univariate setting and using a "local to unit root asymptotics" approach, that these estimators have attractive properties in terms of short-term (one-step ahead) prediction, in the sense that the optimal weight is strictly between 0 and 1. Using Monte Carlo experiments we investigate the in-sample performances as well as the short-term and long-term prediction performances of this kind of estimators, in univariate and bivariate settings, and we propose a practical sequential technique for computing an optimal averaging. We find that in-sample and out-of-sample forecast performances are clearly in favor of the class of averaging estimators. For instance, in the univariate setting, with $\rho = 0.99$ and $T = 160$, we find that: *a*) the root mean square error (RMSE) provided by the averaging estimator is more than five times (nine times, respectively) smaller than that associated to the other bias-corrected estimators (OLS estimator, respectively); *b*) the percentage of increase of the one-step ahead and twenty-step ahead root mean square forecast error (RMSFE, normalized by the unfeasible RMSFE corresponding to the true value of ρ) is about four times smaller for the optimal averaging estimator, compared with the bias-corrected ones, and six times smaller than the OLS one.

In order to further corroborate the reliability of the averaging estimators to handle the persistence problem, we provide (in a 3-dimensional setting) an empirical application to the extraction of the expectation term from yields with different residual maturities.

In the second part of our analysis we study the above mentioned persistence-based issues in a more general setting in which the large level of persistence is generated by both the conditional mean and the time-varying conditional variance of the variable of interest (first-order and second-order persistence, say). In particular, we see how the averaging estimators can be used for the estimation of a long-term Value-at-Risk (VaR) in the case of a persistent AR(1)-ARCH(1) model. We find, for instance, that: *i*) the downward bias of the OLS estimator of ρ may strongly rise (around 60%)

when we move from an homoscedastic to an heteroscedastic setting; *ii*) the *VaR* may be multiplied by 3 when the autoregressive parameter moves from 0.9 to 1; *iii*) the averaging estimator provides a relative RMSEs of the *VaR* that may be between five and six times smaller than the RMSE corresponding to the OLS estimator.

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 compares its prediction performances with those of the previously defined "bias-corrected" estimators. Section 5 extends the study to bivariate and trivariate models. More precisely, Section 5.1 provides a Monte Carlo investigation of the reliability of our estimation strategy in a bivariate model, while Section 5.2 shows in a 3-dimensional setting the usefulness of the averaging estimator to extract the expectation term from medium and long term U.S. interest rates. Section 6 considers the problem of estimating long-term *VaR*'s when persistence and conditional heteroscedasticity are present, while Section 7 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 $\widehat{\rho}_T$ of ρ is:

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

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 $\sqrt{T}(\widehat{\rho}_T - \rho)$ is $N(0, 1-\rho^2)$, whereas if $\rho = 1$ the asymptotic distribution of $T(\widehat{\rho}_T - 1)$ 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 $\widehat{\rho}_T$ 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 $\widehat{\rho}_T$ 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 $\widehat{\rho}_T = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2$, the asymptotic behavior of $\widehat{\rho}_T$, 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

$(\widehat{\rho}_T - \rho_T)$ is $\frac{1}{T}$ and the asymptotic distribution of $T(\widehat{\rho}_T - \rho_T)$ 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 that 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 $2[1 - \rho_T]^{-1/2}\, T^{1/2}(\widehat{\rho}_T - \rho_T)$ 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 $[2(1 - \rho_T)]^{-1/2} T^{1/2}(\widehat{\rho}_T - \rho_T)$ 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 $\widehat{\rho}_T$, given by (2), depends only on ρ (and not on μ and σ^2) since, from (1) we can equivalently write:

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

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

(see Figure 1) and $T = 40$ (see Figure 2). These values of T correspond to 40 years of observations with, respectively, quarterly and annual data. We can see the well known increasing left asymmetry of the distributions when 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 $T = 160$.

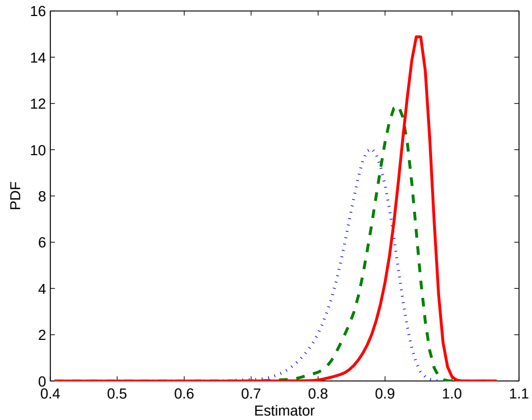


Figure 1:
Probability density function of the OLS estimator $\widehat{\rho}_T$ (sample size $T = 160$): $\rho = 0.91$ (short dashes), $\rho = 0.95$ (dashes) and $\rho = 0.99$ (solid line).

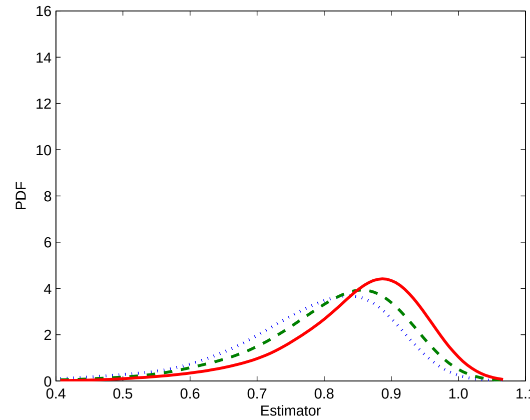


Figure 2:
Probability density function of the OLS estimator $\widehat{\rho}_T$ (sample size $T = 40$): $\rho = 0.91$ (short dashes), $\rho = 0.95$ (dashes) and $\rho = 0.99$ (solid line).

Table 1 gives the skewness and the kurtosis of the $\widehat{\rho}_T$ 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 and 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.

	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 $\widehat{\rho}_T$, computed from 5×10^5 simulations.

Let us now focus on the finite sample bias of $\widehat{\rho}_T$. It is known, since Kendall (1954), that the bias of order $\frac{1}{T}$ is $-\frac{1+3\rho}{T}$, 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^q and $(\rho - 0.025)^q$, for ρ close to 1, are very different for large q. Kendall's formula is however an approximation and it is worth considering a more accurate estimation of the bias function :

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

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 much worse than the one provide by the Kendall's approximation. For $\rho = 0.99$, the exact 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 $\widehat{\rho}_T$ is evaluated as 0.965 whereas the exact expectation, based on the exact bias, is 0.956. The true mean reversion percentage of the prediction at horizon $q = 20$ (five years) is $100(1 - 0.99^{20}) \approx 18\%$, whereas this percentage evaluated at the approximated and the exact expectation of $\widehat{\rho}_T$ are respectively $100(1 - 0.965^{20}) \approx 51\%$ and $100(1 - 0.956^{20}) \approx 59\%$. The situation is even more dramatic for $q = 40$: the true mean reversion percentage is 33%, while this value provided by the approximated and exact expectation are respectively 76% and 83%. So, the consequences of the bias problem may be very severe in terms of predictions.

The bias functions are nonlinear in ρ and, in order to easily work with them, we use an approximation by quadratic spline functions with a knot at $\rho = 0.9$. We obtain excellent fits 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 quadratic spline functions for the expectations $e_T(\rho) = E(\widehat{\rho}_T)$:

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

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

It is also interesting to consider the median bias of $\widehat{\rho}_T$. 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 have 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 $m_T(\rho)$:

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

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

For ease of presentation we will now concentrate on the case $T = 160$ (more detailed results are available upon request from the authors).

3 Bias-corrected estimators

The previous section has shown that very large biases may appear both in the mean and in the median of $\widehat{\rho}_T$, if the true parameter value ρ is close to one. Given our purpose to provide reliable forecasts of persistent variables, and since these biases may have an important impact at the prediction stage, it is natural to introduce various bias-corrected estimators that will be compared, in Section 4, in terms of forecast's precision. This comparison will also include the class of averaging estimators introduced by B. Hansen (2010), and we will verify that the latter are superior both for in-sample and out-of-sample performances.

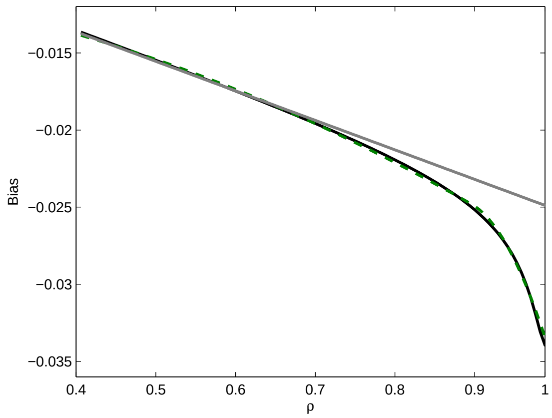


Figure 3:
Bias of the OLS estimator $\widehat{\rho}_T$ (sample size $T = 160$): exact (black solid line), spline approximation (green dashed line) and Kendall's approximation (grey solid line).

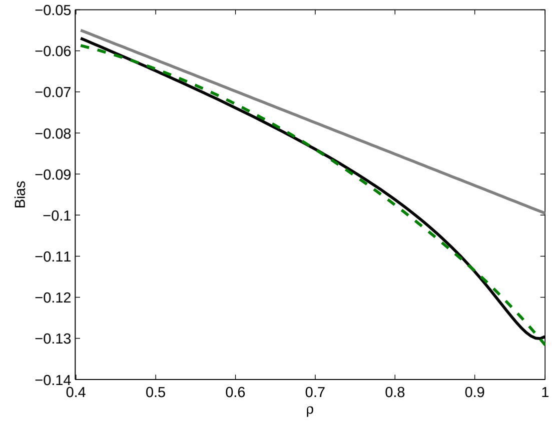


Figure 4:
Bias of the OLS estimator $\widehat{\rho}_T$ (sample size $T = 40$): exact (black solid line), spline approximation (green dashed line) and Kendall's approximation (grey solid line).

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 uncomputable and the method is based on an auxiliary tractable model [see Gourieroux, Monfort and Renault (1993)]. Nevertheless, indirect inference is also useful to remove the finite sample bias in models which are easily estimable. The idea is to replace a given estimator $\widehat{\rho}_T$ by :

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

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

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

and, therefore, $\widehat{\rho}_T^I$ is e_T unbiased and, if e_T was linear, $\widehat{\rho}_T^I$ would be exactly unbiased. In the general case, it can be shown that $\widehat{\rho}_T^I$ is unbiased at order $\frac{1}{T}$ [see Gourieroux and Monfort (1996), Gourier-

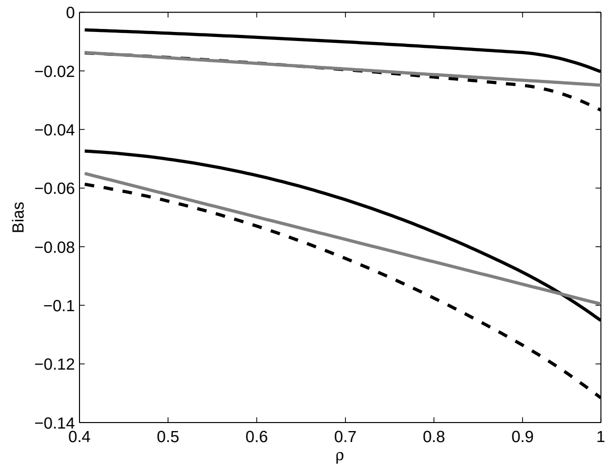


Figure 5:
Mean and Median Bias of the OLS estimator $\widehat{\rho}_T$: Median Bias (black solid lines), Mean Bias (black dashed line) and Kendall's Mean Bias (grey solid line). $T = 160$ (highest curves), $T = 40$ (lowest curves).

oux, Renault and Touzi (2000)]. Moreover, this method has been very successful for removing bias both in a time series [see Phillips and Yu (2005)] and panel data setting [see 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 $\widehat{\rho}_T$, and to replace $\widehat{\rho}_T$ by the OLS estimator based on T pseudo observations drawn in the estimated model. In other words, $b_T(\rho)$ is replaced by $b_T(\widehat{\rho}_T) = e_T(\widehat{\rho}_T) - \widehat{\rho}_T$ and, therefore, the bootstrap bias-corrected estimator is :

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

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

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

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

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

3.3 The "Median-unbiased" estimator

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[4]:

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

where $m_T(\rho)$ is the median function. This estimator, proposed by Andrews (1993) and Rudebusch (1992), is exactly median unbiased if m_T is increasing. Indeed, since m_T^{-1} is increasing, the median of $m_T^{-1}(\widehat{\rho}_T)$ 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 $\widehat{\rho}_T^M$ is ρ, smaller than 1, and an upper truncation above the median does not change it.

[4]See Bauer, Rudebusch and Wu (2011) for an extension of the median-bias correction to VAR models.

In Figures 6 we show the functions $\widehat{\rho}_T^I(\rho), \widehat{\rho}_T^B(\rho), \widehat{\rho}_T^K(\rho)$ and $\widehat{\rho}_T^M(\rho)$ providing the corrections of the OLS estimators, in the range $\rho \in [0.8, 1]$. We see that the indirect inference provides the more important correction, the lowest one being the median-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.

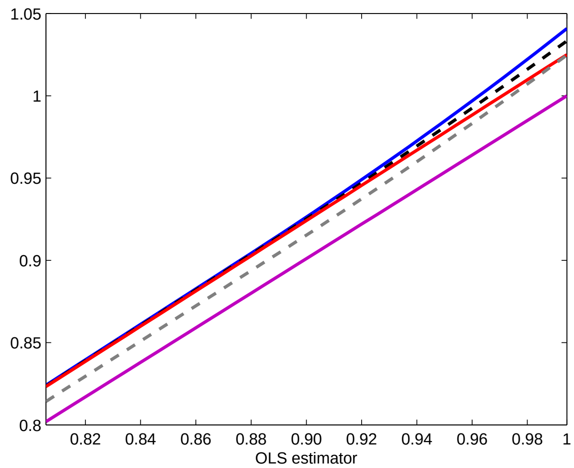


Figure 6:
Bias Corrected Estimators $\widehat{\rho}_T^I$, $\widehat{\rho}_T^B$, $\widehat{\rho}_T^K$ and $\widehat{\rho}_T^M$ providing corrections of the OLS estimators ($T = 160$): Indirect Inference (upper blue solid lines), Bootstrap (black dashed line), Median (grey dashed line), Kendall (red solid line) and OLS (lower purple solid line).

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

Since the main objective of this paper is to find improvements in prediction of persistent time series, we are going to investigate the forecast performances of the estimators introduced in the previous section. Moreover, we will include in the comparison the averaging estimator proposed by B. Hansen (2010). Using "local to unit root asymptotics" techniques, B. Hansen (2010) 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 the AR(1) setup the class of averaging estimators is :

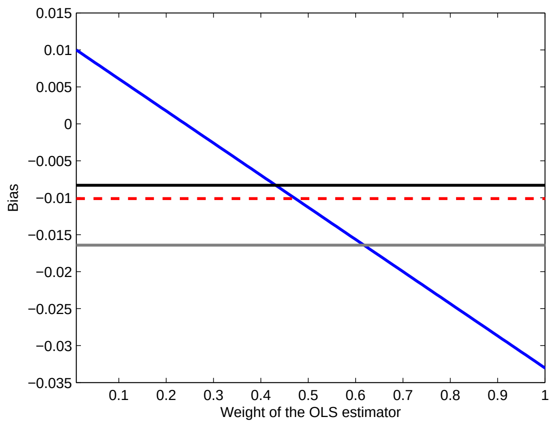
$$\widehat{\rho}_T^A(\lambda) = (1-\lambda) + \lambda \widehat{\rho}_T, \;\; 0 \leq \lambda \leq 1 \,. \tag{11}$$

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

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

Finally, we calculate the ratio of this $RMSFE_i(q)$ to the root mean square forecast error obtained with the true value of ρ. The horizon $q = 20$ would correspond to a five-year horizon for quarterly data.

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 consider the Kendall estimator.



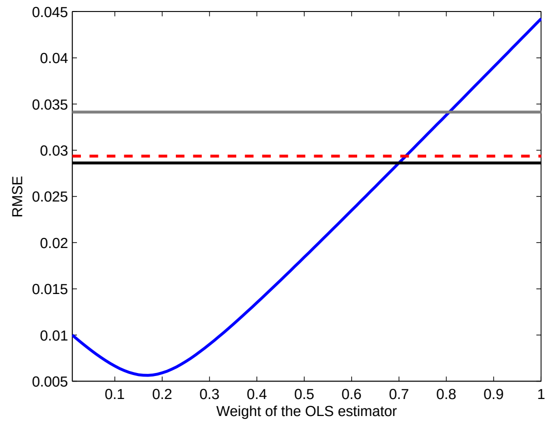


Figure 7:
Bias of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line), when $\rho = 0.99$ and $T = 160$.

Figure 8:
RMSE of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line), when $\rho = 0.99$ and $T = 160$.

Let us first consider the case $\rho = 0.99$ (and $T = 160$; Figures 7 to 10). Among the bias-corrected estimators, the best correction is obtained by $\widehat{\rho}_T^I$, since the -0.034 bias of the OLS is reduced to -0.008. With regard to the $\widehat{\rho}_T^A(\lambda)$ class, the bias is obviously the oblique line between 0.01 and -0.034. The RMSE (Figure 8) is minimal for the $\widehat{\rho}_T^A(\lambda)$ estimator corresponding to $\lambda \simeq 0.15$ and the optimal RMSE is more than five times 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 9 and 10) they are optimal for the $\widehat{\rho}_T^A(\lambda)$ estimator corresponding approximately to $\lambda = 0.25$ for $q = 1$ and $q = 20$. Moreover, for $q = 20$, the percentage of increase of the RMSFE with respect to the unfeasible one corresponding to the true value of ρ, is about four times smaller for the optimal $\widehat{\rho}_T^A(0.25)$ one compared to those obtained from $\widehat{\rho}_T^I, \widehat{\rho}_T^\beta$ and $\widehat{\rho}_T^M$ and six times smaller than the OLS one.

For $\rho = 0.95$ (and $T = 160$; Figures 11 to 14), again, among bias-corrected estimators, the best correction is obtained for $\widehat{\rho}_T^I$ (Figure 11). All bias-corrected estimators dominate in-sample (Figure 12) the OLS ($\lambda = 1$) and the constrained estimator ($\lambda = 0$), but the best in-sample estimator is $\widehat{\rho}^A(\lambda)$ with $\lambda \simeq 0.55$. The best predictions are also obtained for $\widehat{\rho}^A(\lambda)$, with $\lambda \simeq 0.55$ when $q = 1$ (Figure 13), and with $\lambda \simeq 0.8$ when $q = 20$ (Figure 14). The OLS ($\lambda = 1$) is dominated by

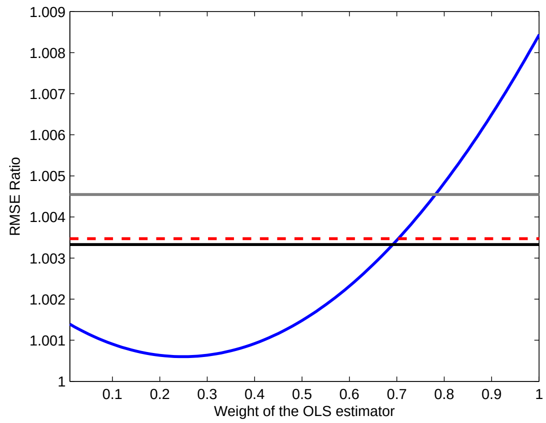


Figure 9:
RMSFE ratios of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line),when $\rho = 0.99$, $T = 160$ and $q = 1$.

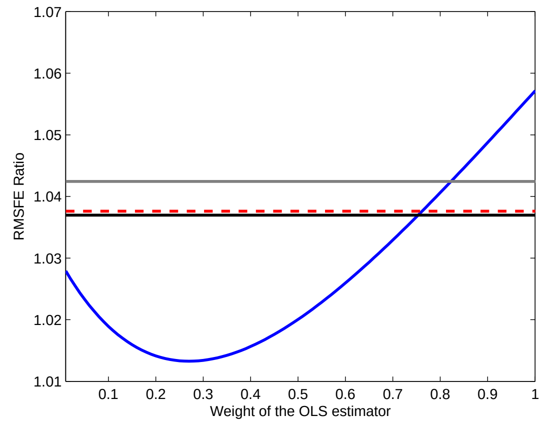


Figure 10:
RMSFE ratios of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line),when $\rho = 0.99$, $T = 160$ and $q = 20$.

the bias-corrected estimators for short-term prediction ($q = 1$), while the converse it true for the long-term forecast ($q = 20$). In any case, the averaging estimator $\widehat{\rho}_T^A(\lambda)$ remains the best solution, with a weight between 0.55 and 0.8, giving, as expected, much more importance to the OLS than when $\rho = 0.99$.

An obvious global conclusion is that, in all the situations, the averaging estimator $\widehat{\rho}_T^A(\lambda)$ is by far the best.

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

- define a sequence of increasing windows $\{1, \ldots, t\}$, with $t \in \{t_0, \ldots, T - q\}$;
- for each t compute the unconstrained estimator $\widehat{\theta}_t^{(u)}$ and the constrained estimator $\widehat{\theta}_t^{(c)}$ of the parameter θ;

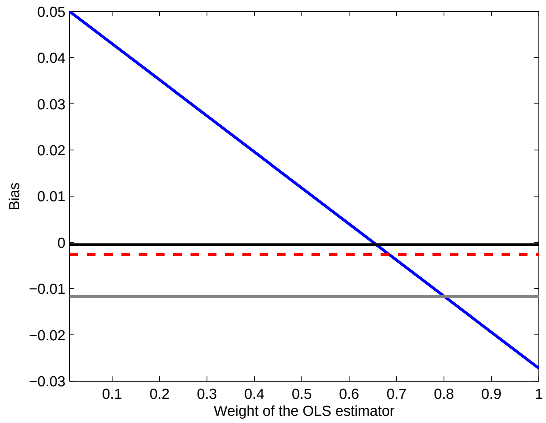


Figure 11:
Bias of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line), when $\rho = 0.95$ and $T = 160$.

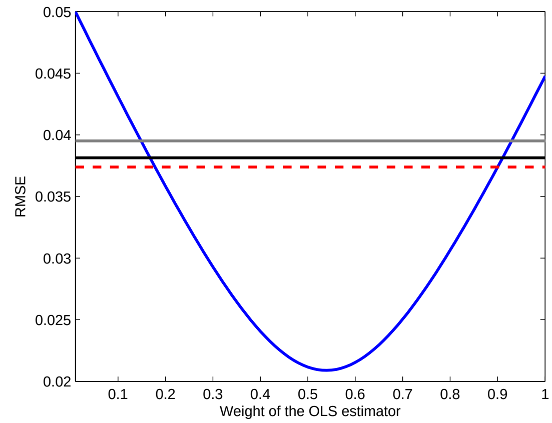


Figure 12:
RMSE of $\widehat{\rho}_T^A$ (blue solid curve), $\widehat{\rho}_T^I$ (black solid line), $\widehat{\rho}_T^B$ (red dashed line) and $\widehat{\rho}_T^M$ (grey solid line), when $\rho = 0.95$ and $T = 160$.

- for each t compute the class of averaging estimators $\widehat{\theta}_t(\lambda) = (1-\lambda)\widehat{\theta}_t^{(c)} + \lambda\widehat{\theta}_t^{(u)}$, the corresponding predictions $\widehat{g}_{t,q}(\lambda)$ of $g(\underline{y_{t+q}})$ and the prediction error $[g(\underline{y_{t+q}}) - \widehat{g}_{t,q}(\lambda)]$;
- compute $\Psi_T(\lambda, q) = \sum_{t=t_0}^{T-q} [g(\underline{y_{t+q}}) - \widehat{g}_{t,q}(\lambda)]^2$;
- calculate $\lambda^*(q) = argmin_{\lambda\in[0,1]} \Psi_T(\lambda, q)$;
- compute $\widehat{\theta}_T(\lambda^*(q))$.

5 Averaging estimators and multivariate dynamics

5.1 A Monte-Carlo investigation of bivariate models

In order to give an example of our averaging estimator strategy in a multivariate setting, let us consider three bivariate data generating processes :

$$y_{1t} = (1-\rho) + \rho y_{1,t-1} + \varepsilon_{1t} \quad (12)$$

$$y_{2t} = 2y_{1t} + \varepsilon_{2t} \quad (13)$$

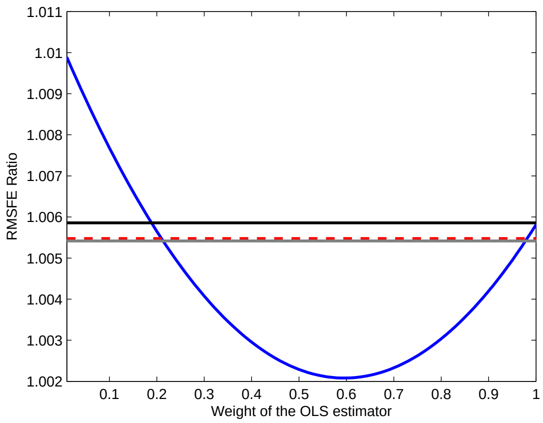


Figure 13:
RMSFE ratios of $\widehat{\rho}_T^I$ (solid line), $\widehat{\rho}_T^B$ (dashes), $\widehat{\rho}_T^A$ (solid curve) and $\widehat{\rho}_T^M$ (short dashes), when $\rho = 0.95$, $T = 160$ and $q = 1$.

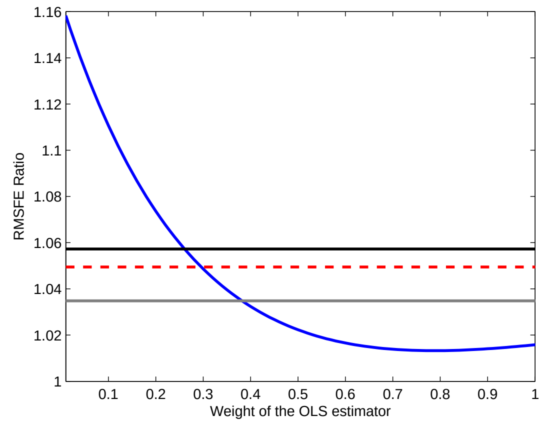


Figure 14:
RMSFE ratios of $\widehat{\rho}_T^I$ (solid line), $\widehat{\rho}_T^B$ (dashes), $\widehat{\rho}_T^A$ (solid curve) and $\widehat{\rho}_T^M$ (short dashes), when $\rho = 0.95$, $T = 160$ and $q = 20$.

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 $I(1)$, $(y_{1t} - 2y_{2t})$ is stationary, and, therefore, $(y_{1t}, y_{2t})'$ is cointegrated.

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

$$y_t = \nu + A y_{t-1} + \eta_t \, , \tag{14}$$

and the unconstrained estimators of ν and A, denoted by $\widehat{\nu}_T^{(u)}$ and $\widehat{A}_T^{(u)}$ 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 $\widehat{\beta}_T$ of β is obtained by regressing y_{1t} on y_{2t}, and the estimators $(\widehat{\mu}_{1T}, \widehat{\alpha}_{1T})$ [resp. $(\widehat{\mu}_{2T}, \widehat{\alpha}_{2T})$] of (μ_1, α_1) [resp. (μ_2, α_2)] are obtained by regressing Δy_{1t} [resp. Δy_{2t}] on $(1, y_{1,t-1} - \widehat{\beta}_T y_{2,t-1})$.

So the constrained estimators of ν and A are :

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

and the class of averaging estimators is :

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

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

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

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

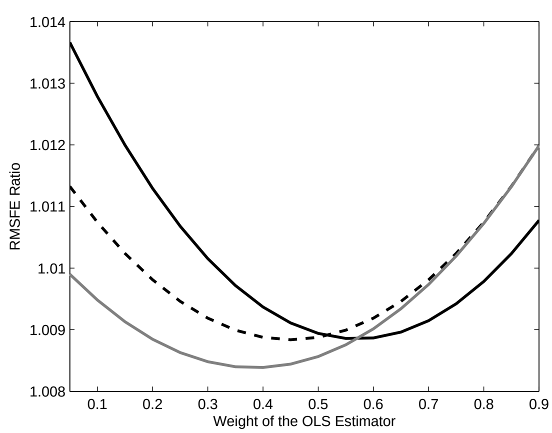


Figure 15:
RMSFE ratio for y_{1t} when $T = 160$ and $q = 1$. $\rho = 0.97$ (black solid line), $\rho = 0.98$ (black dashed line), $\rho = 0.99$ (grey solid line).

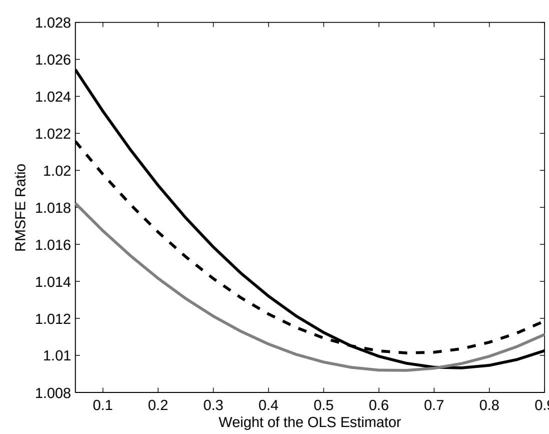


Figure 16:
RMSFE ratio for y_{2t} when $T = 160$ and $q = 1$. $\rho = 0.97$ (black solid line), $\rho = 0.98$ (black dashed line), $\rho = 0.99$ (grey solid line).

Again the number of simulations is 10^5. In Figures 15 and 16 we consider the RMSFE ratios

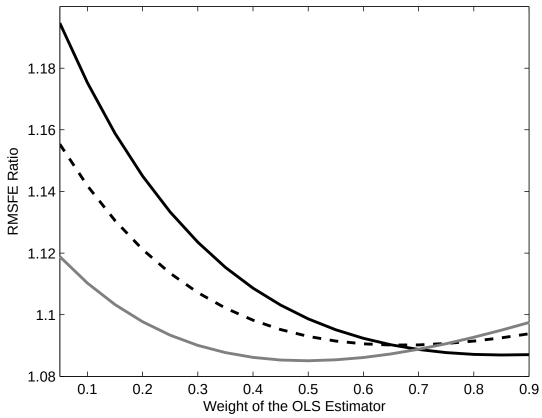


Figure 17:
RMSFE ratio for y_{1t} when $T = 160$ and $q = 20$. $\rho = 0.97$ (black solid line), $\rho = 0.98$ (black dashed line), $\rho = 0.99$ (grey solid line).

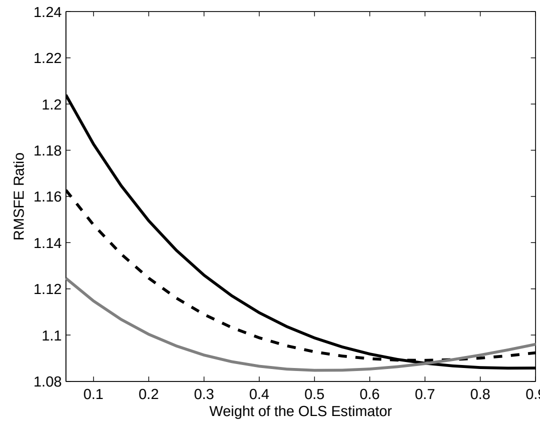


Figure 18:
RMSFE ratio for y_{2t} when $T = 160$ and $q = 20$. $\rho = 0.97$ (black solid line), $\rho = 0.98$ (black dashed line), $\rho = 0.99$ (grey solid line).

for $y_{1,T+q}$ and $y_{2,T+q}$, respectively, in the case $T = 160$ and $q = 1$, with $\rho \in \{0.97, 0.98, 0.99\}$. For $y_{1,T+q}$ 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 $y_{2,T+q}$ these values are, respectively, 0.7, 0.65 and 0.55.

Figures 17 and 18 consider the same situations but for long term predictions ($q = 20$). Here, the optimal value of λ is approximately the same for $y_{1,T+q}$ and $y_{2,T+q}$, 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 $\Psi_T(\lambda)$ (say), based on averaging $\Psi_T(\lambda, q)$ over q, selecting a weighting parameter λ^* independent of the forecasting horizon.

5.2 An application to the extraction of short rate expectations

Up to now we have seen the usefulness of the class of averaging estimators *à la* B. Hansen (2010) reducing RMSEs and RMSFEs (over both short and long horizons) by means of Monte Carlo experiments. The purpose of this section is to apply this methodology to real data and, in particular, to extract short rate expectations (or, equivalently, measures of term premia), over several horizons, by means of a 3-dimensional Gaussian VAR model estimated using 174 quarterly observations of U.S. zero-coupon bond yields and U.S. real GDP, covering the period from 1964:Q1 to 2007:Q2 [see Jardet, Monfort and Pegoraro (2010a) for details]. In Section 5.2.1 we briefly present the so called near-cointegration approach while, in Section 5.2.2, we focus on the extraction of the expectation part from yields with different residual maturities.

5.2.1 The VAR Model

The purpose of the near-cointegration approach of JMP (2010) is to tackle the estimation issues induced by interest rates persistence, and affecting forecast performances, by adopting the class of averaging estimators *à la* B. Hansen (2010) in a multivariate autoregressive setting.

The near-cointegration analysis starts by applying unit root tests and a cointegration analysis to the joint autoregressive dynamics of the short rate (r_t), the long rate (R_t) and the log-GDP (G_t). This econometric procedure leads to a Cointegrated VAR(3) model, or CVAR(3), for $X_t = (r_t, S_t, g_t)'$, the long-term spread $S_t = R_t - r_t$ being the only cointegrating relationship. We obtain in that way a constrained parameter estimates, denoted by $\widehat{\theta}_{T,cvar}$, of the true parameter value $\theta^{(o)} := (\nu^{(o)}, \Phi_1^{(o)}, \Phi_2^{(o)}, \Phi_3^{(o)})$ appearing in the data generating process (DGP):

$$X_t = \nu^{(o)} + \sum_{j=1}^{3} \Phi_j^{(o)} \, X_{t-j} + \eta_t^{(o)} \, , \quad \eta_t^{(o)} \; \sim \; IIN(0, \Omega^{(o)}) \, . \tag{16}$$

The second step consists in estimating $\theta^{(o)}$ by the OLS method applied to the stationary VAR(3) dynamics, and its value is denoted by $\widehat{\theta}_{T,var}$. The last step provides the class of averaging estimators $\theta_T(\lambda)$ given by

$$\theta_T(\lambda) = \lambda \, \widehat{\theta}_{T,var} + (1 - \lambda) \, \widehat{\theta}_{T,cvar} \, , \quad \lambda \, \in [0, 1] \, . \tag{17}$$

The averaging parameter λ in (17), as indicated in Section 4, is selected by minimizing the forecast error of the variable of interest $g(\underline{X_{t+q}})$ (say), over a given forecasting horizon q (say), and using an increasing window for the computation of the constrained and unconstrained estimator of $\theta^{(o)}$.

5.2.2 Averaging Estimations and Extraction of Short Rate Expectations

Given a yield with residual maturity h at date t, denoted by $R_t(h)$, we define its expectation term as $EX_t(h) = -\frac{1}{h}\log \widetilde{B}_t(h)$ with $\widetilde{B}_t(h) = E_t[g(\underline{r_{t,t+h}})] = E_t[\exp(-(r_t + r_{t+1} + ... + r_{t+h-1}))]$. The associated term premium is given by $TP_t(h) = R_t(h) - EX_t(h)$. For a given maturity h, the parameter $\lambda = \lambda(h)$ (say) is selected as the solution of the following problem:

$$\lambda^*(h) = arg \min_{\lambda\in[0,1]} \sum_{t=t_0}^{T-h} [g(\underline{r_{t,t+h}}) - \widehat{g}_{t,h}(\lambda)]^2 \tag{18}$$

where, for each date t and residual maturity h, $g(\underline{r_{t,t+h}}) = \exp(-(r_t+r_{t+1}+...+r_{t+h-1}))$ while $\widehat{g}_{t,h}(\lambda)$ is the VAR model's forecast of $g(\underline{r_{t,t+h}})$ using the parameter $\theta_T(\lambda)$ given by (17). The out-of-sample forecasts are performed during the period 1990:Q1 - 2007:Q2, using an expanding window for the estimation of models VAR(3) and CVAR(3). More precisely, we first compute the estimates $\widehat{\theta}_{t,var}$ and $\widehat{\theta}_{t,cvar}$ over the period 1964:Q1 to 1989:Q4 and we calculate $g(\underline{r_{t,t+h}})$ with t =1989:Q4. Then, at each later point in time t, we recompute $\widehat{\theta}_{t,var}$ and $\widehat{\theta}_{t,cvar}$ taking into account the new observation and, in doing so, we replicate the typical behavior of an investor that incorporates new information over time [see also Favero, Kaminska and Sodestrom (2006)].

	h	AR(1) (Vasicek)	$\lambda^*(h)$	NCVAR(3)	CVAR(3)	VAR(3)	VAR(1)
$g(\underline{r_{t,t+h}})$	4	0.0071	1.0000	0.0064	0.0066	0.0064	0.0075
	8	0.0257	0.7913	0.0228	0.0237	0.0229	0.0261
	12	0.0476	0.5667	0.0423	0.0436	0.0429	0.0469
	20	0.0882	0.1760	0.0736	0.0742	0.0806	0.0848
	32	0.1229	0.2311	0.0991	0.1037	0.1135	0.1196
	40	0.1411	0.2617	0.1012	0.1155	0.1224	0.1406

Table 2: Out-of-sample forecasts of $g(\underline{r_{t,t+h}}) = \exp(-(r_t + r_{t+1} + ... + r_{t+h-1}))$. Table entries are associated RMSFEs. AR(1) (Vasicek) denotes forecasts of $g(\underline{r_{t,t+h}})$ using a Gaussian AR(1) process describing the dynamics of the (one-quarter) short rate. The residual maturities (h) are measured in quarters.

In table 2 we compare, for h ranging from 4 to 40 quarters, the RMFSE obtained from the estimated model with $\lambda^*(h)$ solution of (18) (named Near-Cointegrated VAR(3) or NCVAR(3)), and those obtained by the CVAR(3), VAR(3), VAR(1) and a short rate AR(1) models. We observe that, for $h > 4$, the NCVAR(3) estimation outperforms that of the VAR(3) and CVAR(3) models: there exists a $\lambda^*(h)$, strictly between 0 and 1, such that the average of the estimated parameters in the CVAR(3) and VAR(3) models improves the forecasts of $g(\underline{r_{t,t+h}})$ [similar results are obtained if we consider a rolling window instead of an expanding one, see JMP (2010b)]. Moreover, the NCVAR(3) estimation outperforms the VAR(1) and AR(1) models for any h; in particular, for long maturities, it reduces their out-of-sample RMSFEs of 20-30%.

6 Averaging estimators, volatility and Value-at-Risk

The purpose of this section is to study and to try to handle the same issues as the ones considered above, but in a more general setting where the persistence is generated by both the conditional mean (*first-order persistence*, say) and the time-varying conditional variance (*second-order persistence*, say). More specifically, we will consider the problem of estimation of the popular notion of Value-at-Risk (*VaR*), in particular long-term *VaR*, in the presence of persistence and conditional heteroscedasticity. In order to tackle these issues we will consider that the variable of interest y_t follows an AR(1)-ARCH(1) model of the form:

$$\begin{aligned} y_t &= \rho\, y_{t-1} + \varepsilon_t\,, \;\; -1 < \rho < 1\,, \\ \varepsilon_t &= (1 - a + a\,\varepsilon_{t-1}^2)^{1/2}\, u_t\,, \;\; 0 \leq a \leq 1\,, \end{aligned} \tag{19}$$

where u_t is a standard Gaussian white noise. Note that we have normalized to 1 the unconditional variance of ε_t.

More precisely we will, first, revisit in Section 6.1 the evaluation of the bias of the OLS estimator of ρ in this new setting. Note, however, that we now suppose that the mean of y_t is known and equal to zero and that $\widehat{\rho}_T$ is equal to $\sum_{t=1}^T y_t\, y_{t-1} / \sum_{t=1}^T y_{t-1}^2$. Second, we will evaluate in Section 6.2 the *VaR* at the 0.5% level and horizon $q = 40$ (i.e., ten years for quarterly observations, as frequently adopted in the VaR or portfolio allocation literature; see Campbell and Viceira (2002), Bec and Gollier (2010) and Jondeau and Rockinger (2010)), as a function of ρ and a. Third, we will study in Section 6.3 the potentialities of the averaging estimators of ρ for the estimation of this VaR and, finally, in Section 6.4 we will discuss the feasibility of the optimal averaging estimator.

6.1 The bias of the OLS estimator $\widehat{\rho}_T$ revisited

Let us consider again the case $T = 160$ and let us compute the bias of the OLS estimator $\widehat{\rho}_T$, for $\rho \in (0.4, 1)$ and for $a = 0$, $a = 0.5$ and $a = 0.9$, that is for different degree of persistence provided by the conditional mean or the conditional variance. The case $a = 0$ is the benchmark case considered in Section 2, with μ known and equal to zero [5]. It is clear from figure 19 that conditional heteroscedasticity makes the situation worse since the bias becomes larger, in particular for large values of ρ and a. For example, for $\rho = 0.9$, the bias is -0.026 for $a = 0$ (benchmark case), -0.029 for $a = 0.5$ and -0.042 for $a = 0.9$.

If the variable of interest is a conditional long-term Value-at-Risk, it is relevant to study the behavior of this VaR as a function of ρ and a, that is a function of the two parametric sources of persistence.

6.2 Evaluation of the VaR

We compute by simulation (10^5 replications) the conditional VaR of y_{T+q}, for $q = 40$, $y_T = y_{T-1} = y_{T-2} = 0$ at level 0.5%, as a function of ρ and a. This VaR, denoted $VaR(\rho, a)$ is defined by:

$$\mathbb{P}\left[y_{T+q} < -VaR(\rho, a) \,|\, \underline{y}_T\right] = 0.005\,. \tag{20}$$

As shown on Figure 20, it turns out that, as expected, the VaR is an increasing function of ρ. The $VaR(\rho, a)$ for $a = 0.9$ is larger than the $VaR(\rho, a)$ for $a = 0$, and the $VaR(\rho, a)$ for $a = 0.5$

[5] Observe that this information about μ makes the bias of $\widehat{\rho}_T$ slightly smaller in absolute value when ρ is close to 1.

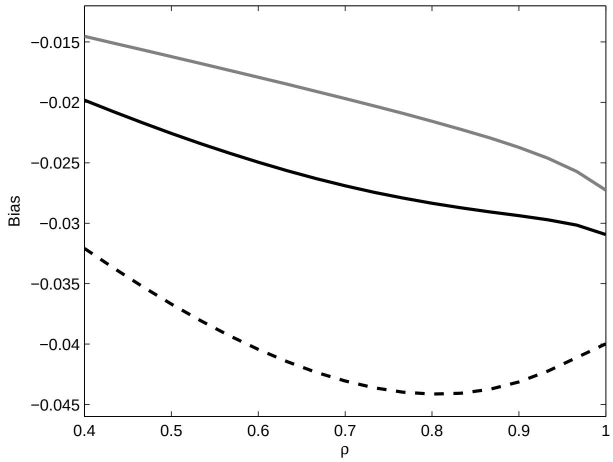


Figure 19:
Bias of the OLS estimator $\widehat{\rho}_T$, for $\rho \in (0.4, 1)$ and for $a = 0$ (grey solid line), $a = 0.5$ (black solid line) and $a = 0.9$ (black dashed line).

(not shown on the graph) is almost identical to that corresponding to $a = 0.9$. We see that the *VaR* is extremely sensitive to variations of a large (close to one) ρ. For instance, for $a = 0$, $VaR(\rho, a)$ is almost multiplied by 3 when ρ moves from 0.9 to 1.

Since $VaR(\rho, a)$, as a function of ρ and for given $a = 0$, $a = 0.5$ and $a = 0.9$, will be useful in the following section, we have fitted it by quadratic splines (with a knot at $\rho = 0.8$). The (excellent) fit of $-VaR(\rho, a)$, with $\rho \in [0.4, 1]$ and $a = 0$, $a = 0.5$ and $a = 0.9$, is given by:

$$\begin{aligned} -VaR(\rho, 0) &= 1.47 - 13.91\,\rho + 9.46\,\rho^2 - 297\,(\rho - 0.8)^2\,\mathbb{1}_{(\rho > 0.8)}\,, \\ -VaR(\rho, 0.5) &= 1.03 - 13.86\,\rho + 8.89\,\rho^2 - 286\,(\rho - 0.8)^2\,\mathbb{1}_{(\rho > 0.8)}\,, \\ -VaR(\rho, 0.9) &= 0.12 - 10.05\,\rho + 5.62\,\rho^2 - 273\,(\rho - 0.8)^2\,\mathbb{1}_{(\rho > 0.8)}\,. \end{aligned} \tag{21}$$

6.3 Potentialities of the averaging estimators

In order to evaluate the potentialities of the averaging estimators, let us consider the following Monte Carlo experiment. We simulate S paths of length $T = 160$ in the AR(1)-ARCH(1) model defined above for given pair of values (ρ, a) and, for each path $s \in \{1, \ldots, S\}$, we compute the

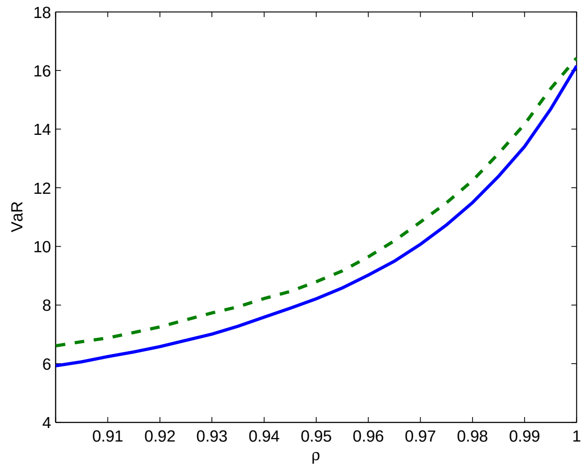


Figure 20:
$VaR(\rho, a)$ at 0.5% level, for $\rho \in (0.9, 1)$, $a = 0$ (blue solid line) and $a = 0.9$ (green dashed line).

OLS estimator of ρ, denoted $\widehat{\rho}_T^{(s)}$. Then, for $\lambda \in [0, 1]$, we compute the averaging estimators $\widehat{\rho}_T^{(s)}(\lambda) = (1 - \lambda) + \lambda\, \widehat{\rho}_T^{(s)}$ and the corresponding *VaR*, $VaR(\widehat{\rho}_T^{(s)}(\lambda), a)$, at horizon $q = 40$. Finally, for $S = 5 \times 10^4$, we compute the Root Mean Square Errors (RMSE):

$$RMSE(\lambda) = \left\{ \frac{1}{S} \sum_{s=1}^{S} \left[VaR(\widehat{\rho}_T^{(s)}(\lambda), a) - VaR(\rho, a) \right]^2 \right\}^{1/2}, \tag{22}$$

for $\rho \in \{0.95, 0.99\}$ and $a \in \{0, 0.5, 0.9\}$. In Figures 21 and 22 we provide the relative RMSE in percentage, that is the RMSE divided by $VaR(\rho, a)$. In Figure 21, corresponding to the case $\rho = 0.99$, we see that the optimal weights of the OLS estimators are small. This is not surprising since the constrained estimator, namely $\rho = 1$, is close to the true value $\rho = 0.99$. Moreover, we see that the optimal relative RMSEs are approximately five times (for $a = 0.5$) and six times (for $a = 0.9$) smaller than the RMSE corresponding to the OLS estimator. This means that the huge estimation uncertainty of 30% when using OLS can be potentially reduced to 5% when using an optimal averaging estimator of ρ. From Figure 22, corresponding to $\rho = 0.95$, we see that now the optimal weights of the OLS estimators are much larger and that the ratios between the optimal relative RMSE and the RMSE corresponding to the OLS estimator are approximately 2/3.

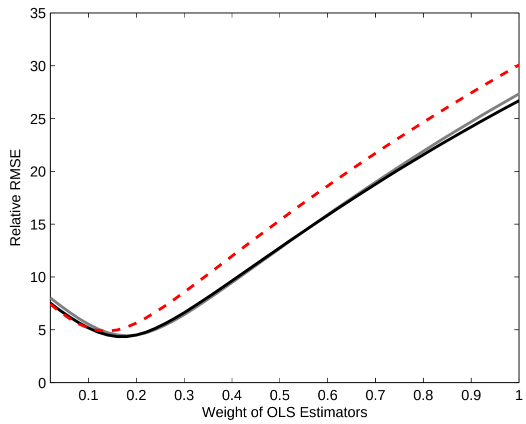


Figure 21:
Relative RMSE of $VaR(\widehat{\rho}_T^{(s)}(\lambda), a)$ for $\rho = 0.99$, $a = 0$ (grey solid line), $a = 0.5$ (black solid line) and $a = 0.9$ (red dashed line).

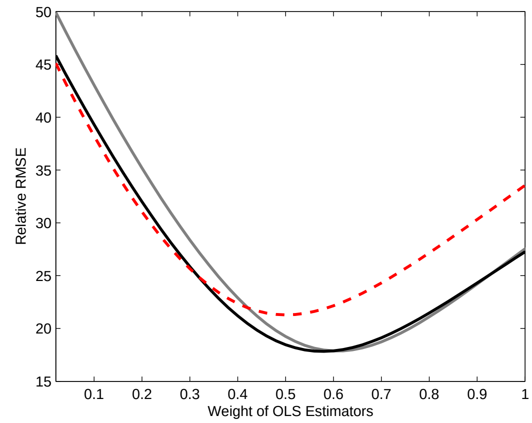


Figure 22:
Relative RMSE of $VaR(\widehat{\rho}_T^{(s)}(\lambda), a)$ for $\rho = 0.95$, $a = 0$ (grey solid line), $a = 0.5$ (black solid line) and $a = 0.9$ (red dashed line).

6.4 Attainability of the optimal averaging estimator

The procedure proposed at the end of Section 4 to estimate the optimal weight λ is not directly applicable in the present case since the quantity we want to approximate, namely $VaR(\rho, a)$, is not observable, contrary to $g(\underline{y_{t+q}})$ which is observed at $t+q$. However, for a given horizon q, it is natural to conjecture that the estimation of the optimal weight obtained for the prediction of y_{t+q} at date t, by the procedure proposed above, will also have good properties for the estimation of the VaR's at the same horizon q.

In order to check this conjecture let us compare the value of λ giving the minimum of $RMSE(\lambda)$ with the one giving the minimal of the following Root Mean Square Forecast Errors (RMSFE):

$$RMSFE(\lambda) = \left\{ \frac{1}{S} \sum_{s=1}^{S} \left[(\widehat{\rho}_T^{(s)})^q \, y_T^{(s)} - y_{T+q}^{(s)} \right]^2 \right\}^{1/2}, \tag{23}$$

for various values of ρ and a. Let us for instance concentrate on the cases $\rho \in \{0.95, 0.99\}$ and $a = 0.5$. In Figures 23 and 24 we consider the various functions $RMSFE(\lambda)$ and $RMSE(\lambda)$, all of them being normalized in the sense that their mean and standard error over λ have been taken equal to 0 and to 1 respectively.

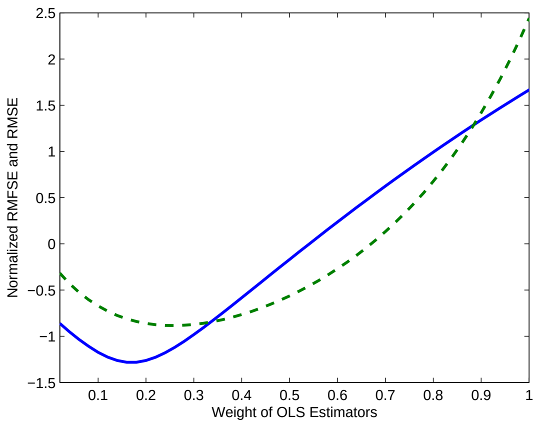


Figure 23:
Normalized $RMSE(\lambda)$ (blue solid line) and $RMSFE(\lambda)$ (green dashed line) of $VaR(\widehat{\rho}_T^{(s)}(\lambda), a)$ for $\rho = 0.99$ and $a = 0.5$.

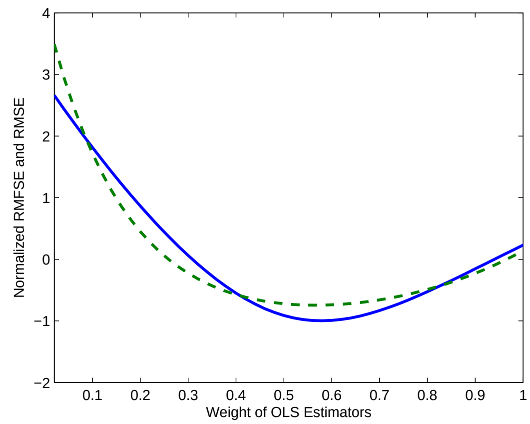


Figure 24:
Normalized $RMSE(\lambda)$ (blue solid line) and $RMSFE(\lambda)$ (green dashed line) of $VaR(\widehat{\rho}_T^{(s)}(\lambda), a)$ for $\rho = 0.95$ and $a = 0.5$.

In the case $\rho = 0.99$ (Figure 23) the minimal $RMSE$ and $RMSFE$ are approximately reached for $\lambda = 0.15$ and $\lambda = 0.20$ respectively. In the case $\rho = 0.95$ (Figure 24) these values are approximately both equal to $\lambda = 0.55$. So, it seems from these simulations that the estimations of the optimal weight based on the prediction of y_{t+q} would be, at least, a good first approximation of the optimal weight for the computation of the VaR at the same horizon. In practice, it would be also useful to implement a Monte Carlo checking, by computing the optimal VaR weight based on simulations from the model in which the parameters are equal to the optimal averaging estimators based on the data.

7 Conclusions

In this paper we have compared the prediction performances of various estimators when high (first-order or second-order) 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, according to our Monte Carlo experiments, all these estimators are dominated by the optimal averaging estimator.

B. Hansen (2010)'s work gives a theoretical explanation of this phenomenon at least for short-term (one-step ahead) prediction and for a simple univariate autoregressive 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 with or without conditional heteroscedasticity.

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.

Bauer, M., Rudebusch, G., Wu, C., (2011): "Unbiased Estimation of Dynamic Term Structure Models", Federal Reserve Bank of San Francisco, Working Paper Series.

Bec, F., and Gollier, C., (2010): "Cyclicality and term structure of Value-at-Risk within a threshold autoregression setup", Toulouse School of Economics, Working Paper.

Campbell, J., Y., and Viceira, L., (2002): "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors", Oxford University Press.

Cochrane, J., (2006): "Financial Markets and the Real Economy", Edward Elgar Publishing Limited.

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.

Jondeau, E., and Rockinger, M., (2010): ”Portfolio Allocation for European Markets with Predictability and Parameter Uncertainty”, HEC Lausanne, Working Paper.

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., Renault, E., (1993): ”Indirect Inference”, Journal of Applied Econometrics, 8, S85-S118.

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

Gourieroux, C., Monfort, A. (2008): ”Quadratic Stochastic Intensity and Prospective Mortality Tables”, Insurance Mathematics and Economics, 43, 174-184.

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., (2010): "Averaging Estimators for Autoregressions with a Near Unit Root", Journal of Econometrics, 158, 142-155.

Jardet C., Monfort A., Pegoraro, F. (2010a): "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth", working paper Banque de France.

Jardet, C., Monfort, A., and Pegoraro, F., (2010b): "Online Appendix for 'No-arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth' ", available at the website `http://www.crest.fr/pageperso/pegoraro/pegoraro.htm`.

Joslin, S., Priebsch, M., and Singleton, K., (2010): "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks", Graduate School of Business, Stanford University, Working Paper.

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

Kozicki, S., and Tinsley, P.A., (2001a): "Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations", Journal of Economic Dynamics and Control, 25, 149-184.

Kozicki, S., and Tinsley, P.A., (2001b): "Shifting Endpoints in the Term Structure of Interest Rates", Journal of Monetary Economics, 47, 613-652.

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.

Rossi, B., (2005): ”Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle”, International Economic Review, 46 (1), 61-92.

Rudebusch, G. (1992): ”Trends and Random Walks in Macroeconomic Time Series: A Re-Examination”, International Economic Review, 33(3), 661-680.

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.