Karim Abadir, Paolo Paruolo

On efficient simulation in dynamic models

2007/9

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

Printed in Italy in February 2008
Università degli Studi dell'Insubria
Via Monte Generoso 71, 21100 Varese, Italy

ON EFFICIENT SIMULATIONS IN DYNAMIC MODELS

Karim M. Abadir and Paolo Paruolo*
Imperial College London and University of Insubria

February 5, 2008

Abstract

Ways of improving the efficiency of Monte-Carlo (MC) techniques are studied for dynamic models. Such models cause the conventional Antithetic Variate (AV) technique to fail, and will be proved to reduce the benefit from using Control Variates with nearly nonstationary series. This paper suggests modifications of the two conventional variance reduction techniques to enhance their efficiency. New classes of AVs are also proposed. Methods of reordering innovations are found to do less well than others which rely on changing some signs in the spirit of the traditional AV. Numerical and analytical calculations are given to investigate the features of the proposed techniques.

Keywords: Dynamic models, Monte-Carlo (MC), Variance Reduction Technique (VRT), Antithetic Variate (AV), Control Variate (CV), Efficiency Gain (EG), Response Surface (RS).

J.E.L. Classification: C15.

*We thank David Hendry for being a continuing source of inspiration for us. We also thank two anonymous referees for their comments.

1 Introduction

Unit roots have become a central theme of modern econometrics. Numerous studies of the subject have used Monte-Carlo (MC) techniques. Yet we do not know how these MC techniques, usually designed for stationary series, fare with nearly nonstationary ones.[1] An important aspect is the control of MC variability, which is shown here to increase with the sample size in nearly nonstationary series. For example, when simulating quantiles of AR statistics, this excess MC variation can be easily shown by using formula (10.29) of Kendall and Stuart (1977, p.252) together with the thin long lower tails that are typical of AR densities.

The study of MC techniques in dynamic models lags behind other –admittedly more pressing– aspects of the subject. This paper will attempt to partially redress this imbalance by presenting ideas on how to improve MC studies of autoregressive series, possibly with roots close to unity. More specifically, Variance Reduction Techniques (VRTs) will be suggested to cope with the problems of simulating dynamic models of this type.

VRTs are methods of combining different estimates obtained from using a single set of generated random numbers more than once. When successful, they reduce MC imprecision as explained in detail by Hendry (1984). Furthermore, unconventional uses for these VRTs can be found, as we will discuss later. In addition to MC, we use response surfaces and asymptotic results to compare the relative performance of various VRTs. Asymptotic methods are employed in this context both in the econometrics and finance literature. Asymptotics for a large number of MC replicates are reported in Paruolo (2002), where VRT are used to increase precision in test-power comparison. Ericsson (1991) employs small-σ asymptotics.[2] Finally, several other asymptotic methods are in use in finance; see e.g. Takahashi and Yoshida (2005) and references therein.

The outline of the paper is as follows. First, the VRTs to be considered are defined in Section 2. The new VRTs of this work will have to be compared to some benchmark. For this reason, existing VRTs will be briefly defined alongside the new ones. Next, Section 3 compares these VRTs numerically and analytically (by response surfaces) in finite samples, placing special emphasis on typical nearly nonstationary conditions without excluding the possibility of stationarity from the study. In the response surface analysis, we show how to bridge the gap in the distribution theory that arises from stationary and nonstationary data by means of a simple function that will be given in (25).

The results show large efficiency gains (e.g. by an average factor of about 20 times) for a few of the VRTs, leading to the ability to conduct faster and more precise MC in the future. This is useful not just for academic econometricians. For example, in financial econometrics, a trader may wish to price an option (which is an expectation) precisely and quickly, in order to make a profitable trade that may otherwise be missed, and our VRTs can be used for this purpose.

Section 4 then describes encompassing formulations for these VRTs, and looks at some practical problems where they may be beneficially used. Some of these applications are nonstandard, including an illustration of how to use the results of earlier sections to devise a method of improving the efficiency of MC work on nearly nonstationary series. Other possible unconventional uses of these VRTs include the derivation of power functions of tests and numerical integration. It is worth stressing that all these benefits come at little programming cost. Typically, only 2 or 3 lines of code are required to program these VRTs, and they can be easily added to subroutines that generate random numbers.

In Section 5, we use large-sample asymptotic results to analyze the variance-reduction

[1] The term "near nonstationarity" is used in this paper to refer to autoregressive roots in the neighbourhood of unity, including a unit root.

[2] For the problem considered in this paper the statistics of interest is scale invariant, and small-σ asymptotics is not a viable option.

factor of various VRTs. We employ functional central limit theorems to study the behavior of relevant statistics for unstable autoregressions; for stable autoregressions we use standard central limit theorems together with covariance calculations. We are thus able to describe the correlation coefficient between antithetic variates as an explicit function of the autoregressive parameter, where this correlation is the key element in the MC variance reduction formula. Moreover, these results allow us to discuss analytically the choice of a rotation parameter in some class of orthogonal antithetic variates.

Finally, concluding observations are made in Section 6.

2 VRTs: new and old

The two most prominent types of VRTs in econometrics will be considered in this work. They are Antithetic Variates (AVs) and Control Variates (CVs). The failing of known forms of AVs in dynamic models has been documented in Hendry (1984). Here, it will be shown analytically that conventional CVs also fail when the variables are nearly nonstationary. For reasons of simplicity, let the dynamic Data Generating Process (DGP) be

$$y_t = \alpha y_{t-1} + \varepsilon_t, \tag{1}$$

where $t = 1, ..., T$, $y_0 = 0$, $\varepsilon_t \sim \mathrm{IN}(0, \sigma^2)$. The reason for choosing this DGP is that it is representative of the problems associated with the nearly nonstationary case. More general ARMA processes where the orders may even be unknown will give rise to similar results when treated as in Said and Dickey (1984), as long as determinsitics are not included. Also, the study's emphasis on near nonstationarity means that the choice of distribution for ε_t in (1) is not crucial to the results, especially when T is not small; see Phillips (1987). So (1) was adopted to keep complexity of the exposition at a minimum.

A DGP with zero intercept and trend was chosen to retain the nonstandard results that arise in near nonstationarity with its associated unconventional problems. If any of these parameters were nonzero, normality of distributions would be restored to the usual statistics associated with (1) at the fast rates of $T^{3/2}$ and $T^{5/2}$, respectively. The choice of DGP (1) should not be seen as restricting the validity of the new VRTs defined below. Their general definition is independent of (1) in spite of being motivated by it, and can be used in a variety of frameworks other than (1).

Ordinary Least Squares (OLS) yields a consistent estimate of α in (1), and is given by

$$\widehat{\alpha} := \sum\nolimits_t y_t y_{t-1} / \sum\nolimits_t y_{t-1}^2, \tag{2}$$

where $t = 1, \ldots, T$ in the summation ($t = 1$ gives a zero term because $y_0 = 0$). VRTs have many uses, which we will illustrate with evaluating $\mathrm{E}(\widehat{\alpha})$ when (α, σ^2) belongs to the parameter space $\{|\alpha| \leq 1, \sigma^2 > 0\}$ and $T \in \mathcal{T} \subset \mathbb{N}$. Here, the only difference with Hendry (1984) is in allowing $|\alpha| = 1$ within this parameter space.

A simple Monte Carlo technique would be to generate $n = 1, \ldots, N$ replications of the series in DGP (1), and calculate N OLS estimates of α in (2). By taking the average of these, one can estimate $\mathrm{E}(\widehat{\alpha})$. This is called the 'crude MC estimator' and it is an unbiased estimator of $\mathrm{E}(\widehat{\alpha})$, with MC variance given by $\mathrm{var}(\widehat{\alpha})/N$. One can improve on the crude MC estimator by using VRTs. VRTs re-use the same set $\{\varepsilon_t\}$ once more, instead of simply using it once as the crude MC estimator.

The computational cost of VRTs is represented by the ratio, κ say, of the computing time required by the re-use of the set $\{\varepsilon_t\}$ relative to the computing time required by an additional replication of the crude MC estimator, which involves generating a new set of $\{\varepsilon_t\}$. The ratio κ depends on the hardware and software solutions used in the implementation; usually one has $\kappa < 1$ and often $\kappa \ll 1$. The benefits of VRTs are the ability to generate an estimator having the same expectation as $\widehat{\alpha}$, but be less variable by a factor of 20 or so, as we will see in the next section; this is equivalent to having increased N by a factor of 20 in the ratio $\mathrm{var}(\widehat{\alpha})/N$. A successful VRT can therefore lead to massive gains in terms of efficiency and speed.

In general, AVs transform the set $\{\varepsilon_t\}$ to an antithetic counterpart $\{\varepsilon_t^-\}$ that is used to generate another set of observations $\{y_t^-\}$ which is also called antithetic. Then, the same estimation method is applied to $\{y_t^-\}$, yielding a second estimate of α, denoted by $\widehat{\alpha}^-$ and called the antithetic estimate. The antithetic transform is chosen so that $\widehat{\alpha}$ and $\widehat{\alpha}^-$ have the same expectation; this implies that the combined estimate

$$\widetilde{\alpha} := \frac{1}{2}\left(\widehat{\alpha} + \widehat{\alpha}^-\right) \tag{3}$$

has also the same expectation of $\widehat{\alpha}$ and $\widehat{\alpha}^-$ with variance

$$\mathrm{var}(\widetilde{\alpha}) = \frac{1}{4}\left(\mathrm{var}(\widehat{\alpha}) + \mathrm{var}\left(\widehat{\alpha}^-\right) + 2\,\mathrm{cov}(\widehat{\alpha}, \widehat{\alpha}^-)\right) \tag{4}$$

that is designed to be lower than the original $\mathrm{var}(\widehat{\alpha})$. Following Hendry (1984), define the associated Efficiency Gain (EG) as

$$\mathrm{EGv} := \frac{\mathrm{var}\left(\widehat{\alpha}\right)}{\mathrm{var}\left(\widetilde{\alpha}\right)}, \tag{5}$$

where v is the name of the VRT associated with $\widetilde{\alpha}$. When $\mathrm{var}(\widehat{\alpha}) = \mathrm{var}(\widehat{\alpha}^-) = \eta$, say, $\mathrm{var}(\widetilde{\alpha})$ simplifies to

$$\mathrm{var}(\widetilde{\alpha}) = \frac{\eta}{2}\left(1+\rho\right), \tag{6}$$

where $\rho := \mathrm{corr}(\widehat{\alpha}, \widehat{\alpha}^-)$. In this case,

$$\mathrm{EGv} = \frac{2}{1+\rho}. \tag{7}$$

The conventional AV uses

$$\{\varepsilon_t^-\} := \{-\varepsilon_t\} \tag{8}$$

in an attempt to induce $\rho < 0$ which, if successful, would lead to EG > 2.

Unfortunately in the case of (1), the conventional AV fails: $\widehat{\alpha}$ and $\widehat{\alpha}^-$ are identical, and the combined estimator is no better than either of its components since $\mathrm{corr}(\widehat{\alpha}, \widehat{\alpha}^-) = 1$ and $\mathrm{var}(\widehat{\alpha}) = \mathrm{var}\left(\widehat{\alpha}^-\right) = \mathrm{var}(\widetilde{\alpha})$. One way out of this impasse is to use another type of estimator for $\widehat{\alpha}^-$. For example, let $\widehat{\alpha}^-$ be based on Instrumental Variable estimation rather than OLS. The two estimates of α would be numerically different but would be positively correlated (especially in large samples), so that the variance of the combined estimator given in (4) would be only marginally smaller than either $\widehat{\alpha}$ or $\widehat{\alpha}^-$. Because the expected efficiency gain is not likely to be large, one has to think of another alternative for generating $\{\varepsilon_t^-\}$ and hence $\widehat{\alpha}^-$.

The other direction that can be pursued in developing a different $\widehat{\alpha}^-$ is to reuse $\{\varepsilon_t\}$ differently. The general idea is to attempt to create a series $\{\varepsilon_t^-\}$ which leads to an $\widehat{\alpha}^-$ that is preferably negatively correlated with $\widehat{\alpha}$. Four general alternative techniques can now be suggested.

The first new AV, denoted henceforth by AV1, is based on transforming the pair ε_i and ε_j $(i \neq j)$ that were generated as $\mathrm{IN}(0, \sigma^2)$ into

$$\varepsilon_i^- := \left(\pm\varepsilon_i \pm \varepsilon_j\right)/\sqrt{2} \quad \text{and} \quad \varepsilon_j^- := \left(\mp\varepsilon_i \pm \varepsilon_j\right)/\sqrt{2}, \tag{9}$$

which will also be $\mathrm{IN}(0, \sigma^2)$. This can be checked by taking expectations of powers of ε in (9). If the estimator of interest is invariant to scale, as is the case here with $\widehat{\alpha}$, one can omit the $\sqrt{2}$ factors to speed up the calculations.

The signs in (9) mean that the new method can generate, for example, ε_i^- as either $(\varepsilon_i + \varepsilon_j)/\sqrt{2}$ or $-(\varepsilon_i + \varepsilon_j)/\sqrt{2}$. For now, only the upper signs from the general definition (9) will be considered, because one wishes to isolate the separate influences of combining innovations and of switching their signs. The sign-switching features of (9) will be temporarily ignored, as they will be considered separately by other explicit VRTs here. The encompassing generality of the formulation of (9) will be returned to later in Section 4.

Definition (9) does not constrain the order of the variates i and j (except that $i \neq j$), though we shall also temporarily ignore this property to isolate the influence of combining two innovations as opposed to reordering them. Successive pairs will be selected so that $j = i + 1$. Again, we shall come back to these features in Section 4 below.

Finally, we assume for simplicity that the sample size T is even. Otherwise, the last value of the antithetic set, ε_T^- , would need to be generated as if it were ε_{T+1}.

The second antithetic variate, AV2, is one way of resampling $\{\varepsilon_t\}$, and it consists of reversing the order of the original i.i.d. series to get

$$\{\varepsilon_t^-\} := \{\varepsilon_{T-t}\}. \tag{10}$$

It has the small disadvantage of requiring all the innovations $\{\varepsilon_t\}$ to be kept in storage, unlike the previous method which only requires storage space for two consecutive innovations at a time.

The third method, AV3, was mentioned but not tested by Hendry and Harrison (1974, p.156). It is based on using

$$\{\varepsilon_t^-\} := \{(-1)^t \varepsilon_t\}, \tag{11}$$

which alters the sign of every other ε_t. This method should do best when α is close to 0 because the values assumed by $\{\varepsilon_t\}$ matter significantly less to the distribution of $\widehat{\alpha}$ as $|\alpha| \to 1$ (see Phillips (1987)), and because $\alpha = 0$ here gives deterministically

$$\widetilde{\alpha} \equiv \frac{1}{2}\left(\widehat{\alpha} + \widehat{\alpha}^-\right) = \frac{1}{2} \sum\nolimits_t \left(\varepsilon_t \varepsilon_{t-1} - \varepsilon_t \varepsilon_{t-1}\right) / \sum\nolimits_t \varepsilon_{t-1}^2 = 0 \tag{12}$$

as the combined estimator, hence providing an infinite variance reduction relative to the crude MC estimator. Note, however, that in this case $\widetilde{\alpha}$ is degenerate at 0, and it cannot be considered as a realization of an estimator (with the same expectation as $\widehat{\alpha}$), which is a nondegenerate random variable. Hence, in the MC simulations concerning AV3, we will discard the case of $\alpha = 0$.

Finally, AV4 is a very intuitive and easily applicable new VRT. It is cheap on both programming and storage cost considerations. It relies on using exactly the same innovations

$$\{\varepsilon_t^-\} := \{\varepsilon_t\} \tag{13}$$

to generate $\{y_t\}$ through the same DGP as before but with parameter(s) of interest of the opposite sign, namely

$$y_t = -\alpha y_{t-1} + \varepsilon_t = \beta y_{t-1} + \varepsilon_t \tag{14}$$

with $\beta := -\alpha$. The new coefficient β is then estimated by OLS as in (2), and the negative of the resulting estimate, $-\widehat{\beta}$, is the antithetic $\widehat{\alpha}^-$ for the generic AV form in (3). When $\alpha = 0$, $\{y_t^-\} = \{y_t\}$ and $\widehat{\alpha}^- \equiv -\widehat{\beta} = -\widehat{\alpha}$, which causes $\widetilde{\alpha}$ to be zero deterministically as in (12); a case which will be discarded in the MC simulations. Such an infinite variance reduction is not expected for any other value of α, and it is clear that, as α moves away from zero, the variance reduction will fall to finite levels both here and for AV3.

One can think of the last two methods as switching the sign of the effect on y_t of every other ε_t to obtain the antithetic set $\{y_t^-\}$. AV3 does it directly by changing the sign of every other ε_t, while AV4 does it indirectly by changing the sign of α through which lags of ε_t affect y_t. DGP (1) can be rewritten as

$$y_t = \sum_{j=0}^{t-1} \alpha^j \varepsilon_{t-j}, \tag{15}$$

where it is obvious that a change in the sign of α affects every other ε_{t-j} term. For autoregressive DGPs like (1), the two methods provide equal results since AV3 gives

$$y_t^- = (-1)^t \sum_{j=0}^{t-1} (-\alpha)^j \varepsilon_{t-j}, \tag{16}$$

while AV4 leads to

$$y_t^- = \sum_{j=0}^{t-1} (-\alpha)^j \varepsilon_{t-j}. \tag{17}$$

Given the respective definitions of $\widehat{\alpha}^-$ for AV3 and AV4, (16) and (17) give the same combined estimator $\widetilde{\alpha}$ in the case of DGP (1). This equality is due to the choice of our dynamic DGP, and does not necessarily hold for all other DGPs. For example, a static DGP where the conventional AV works,

$$y_t = \alpha x_t + \varepsilon_t, \tag{18}$$

and where x_t is not a lagged value of y_t, causes AV3 and AV4 to be different. Static DGPs are not the focus of this study.

One should be careful to provide the proper justification for using AV4 to simulate moments of a certain order. In general, any AV is a statistically valid method of simulating the moment of order k of an econometric estimator η if and only if $\mathrm{E}(\widehat{\eta}^k) = \mathrm{E}(\widehat{\eta}^{k-})$, where $\widehat{\eta}^{k-}$ is the k-th power of the antithetic estimator. For applying sign-switching AVs to DGP (1), this condition reduces to the requirement that $\mathrm{sgn}(\widehat{\eta}^k) \times \mathrm{E}(\widehat{\eta}^k)$ is an even function of η, so that changing the sign of η (directly or indirectly) in the DGP produces an antithetic variate $\widehat{\eta}^{k-}$ with exactly the same expected value as $\widehat{\eta}^k$.

For $k = 1$ and with DGP (1), this condition is violated for AV3 and AV4 when $\alpha = 0$ as we have seen before. However, when $\alpha \neq 0$, moment generating functions (White (1958, 1961), Abadir (1993b) and references therein) show that the distribution of $\widehat{\beta}$ is the mirror image of that of $\widehat{\alpha}$, thus warranting the use of the technique of AV4 to simulate moments of any order or any other density-related properties such as quantiles. In addition, the other AVs considered in this work satisfy the condition for the first two moments. So, the optimal combination of an estimator and its antithetic counterpart is the simple (as opposed to weighted) average given in the generic form (3).

The condition detailed in the previous paragraph is also satisfied by the general ARMA models analyzed by Cryer, Nankervis and Savin (1989), who consider conditions for invariance and mirror image of estimators; see also Haldrup (1996) for mirror image properties in the case of time trends. The condition set here is not the same as the ones in Cryer et al. (1989) because of the different focus of the two papers: theirs considers whole distributions, whereas ours is only concerned with certain specific moments. Two different distributions may have the same mean, thus satisfying our condition for the first order moment, but not the one in Cryer et al. (1989). Their conditions are thus sufficient but not necessary for the present application.

As with AVs, CVs are meant to modify estimators like (2) so as to reduce MC variability. The CVs under consideration in this work will be extensions of Hendry's (1984, p.953), which is denoted here by CV1. For CVs, the modified estimator of α takes the general form

$$\widetilde{\alpha} := \widehat{\alpha} - \frac{c}{h} \sum\nolimits_t \varepsilon_t y_{t-1}, \tag{19}$$

where h is a variant of

$$\mathrm{E}\left(\sum\nolimits_t y_{t-1}^2\right) = \begin{cases} \sigma^2 \frac{T}{2}(T-1) & (|\alpha| = 1), \\ \sigma^2 \frac{T(1-\alpha^2)+\alpha^{2T}-1}{(1-\alpha^2)^2} & (|\alpha| < 1). \end{cases} \tag{20}$$

For CV1, only the stationary case was analyzed by Hendry (1984). The parameter c of (19) was set at 1 there and h was taken to be the asymptotically dominant term of (20) for the stationary case:

$$h = \sigma^2 T/(1-\alpha^2). \tag{21}$$

Extending this definition to take account of near nonstationarity gives CV2 where $c = 1$ and h is given by (20). Finally, CV3 extends Hendry's CV a step further by estimating the minimum-variance version of (19). After generating all of the $\widehat{\alpha}_n$ $(n = 1, ..., N)$ replications of the MC experiment, let h be given by (20) and estimate the unrestricted regression

$$\widehat{\alpha}_n = \text{constant} + \widehat{c}\left(\sum\nolimits_t \varepsilon_t y_{t-1}/h\right)_n + r_n \tag{22}$$

so that

$$\widetilde{\alpha}_n := \widehat{\alpha}_n - \widehat{c}\left(\sum\nolimits_t \varepsilon_t y_{t-1}/h\right)_n = \text{constant} + r_n,$$

where $\{r_n\}$ are the regression residuals and $\{\widetilde{\alpha}_n\}$ are the N replications of the CV-modified estimator. This process gives rise to the optimal (minimum variance) CV as is described in Kleijnen (1974, pp.138-159). Its only disadvantage in comparison with CV1 or CV2 is a minor computational requirement: a simple regression has to be run at the end of the simulation experiment.

The storage requirements for all three CVs are the same if the series of estimators generated in each replication are to be preserved for a study of moments other than just the mean. A last remark should be made about h. Since it does not vary from one replication to another, then its only purpose in CV3 is as a scale factor. It could be dropped from expression (22) for CV3 without affecting the final results in any way. We shall not do so here because it will be interesting to see how close the estimated c is, relative to the value of 1 which is assigned to c in the case of CV2. This analysis need not be repeated in practical MC studies.

The weakness of these CVs as $|\alpha|$ approaches 1 can now be established analytically for the first time by considering (19) and (20). The stabilizing normalization for $\sum_t \varepsilon_t y_{t-1}$ is $1/\sqrt{h}$. Therefore, because of the swift convergence of $\sum_t \varepsilon_t y_{t-1}/h = O_p(1/\sqrt{h})$ to zero as $|\alpha| \to 1$ and $T \to \infty$, CVs will fail under precisely these two conditions. In a way, CVs are paying the price for the fast convergence of $\widehat{\alpha}$ to α in the case of α near the unit circle (e.g. see Evans and Savin (1981)), and little can be done by means of CVs to improve the efficiency of the already super-consistent $\widehat{\alpha}$. Note that the stochastic components of the CVs are essentially $\widehat{\alpha} - \alpha \equiv \sum_t \varepsilon_t y_{t-1} / \sum_t y_{t-1}^2$ and $\sum_t \varepsilon_t y_{t-1}/h$, which differ only with respect to the type of normalization (stochastic vs. deterministic) that is chosen for $\sum_t \varepsilon_t y_{t-1}$.

Variable	Mean	Minimum	Maximum
EGAV1	1.54	1.4	1.7
EGAV2	1.03	1.0	1.2
EGCV2	20.9	1.3	184
EGCV3	21.1	1.3	186
$\widehat{c}$	0.99	0.83	1.20

Table 1: Summary of MC results for the whole design.

3 Finite-sample results

DGP (1) was used with $\varepsilon_t \sim$ IN(0,1) to test the relative efficiency of the various VRTs suggested above. The parameters of the experiment were

$$\begin{aligned} \alpha &= 0.00,\ 0.25,\ 0.50,\ 0.75,\ 0.80,\ 0.85,\ 0.90,\ 0.95,\ 0.99,\ 1.00; \\ T &= 26,\ 50,\ 100,\ 200,\ 400; \qquad N = 10^4; \end{aligned} \tag{23}$$

with α chosen to give more detail on (and more weight to) nearly nonstationary data, and where $T = 26$ was chosen as the first sample size to accommodate AV1 (for convenience only). There is no need to consider negative values of α for the purpose of this study where efficiency comparison is the only concern, because efficiency is independent of the sign of α; e.g. see Hendry (1984) or refer to the discussion of AV4 following (18) in Section 2. Furthermore, considering changes in efficiency (variance) is tantamount to comparing mean squared errors because the biases of the modified estimators are the same as those of the original OLS. This follows from the discussion after (18) for the AVs, and from (19) for the CVs.

The results of the experiment are then summarized in two ways. First, descriptive statistics are reported for EGv as a crude means of comparing VRTs. Then, response surfaces are fitted to each EGv for an analytical explanation of how various VRTs fare as α and T change.

Variable	Mean	Minimum	Maximum
EGAV1	1.53	1.4	1.7
EGAV2	1.01	1.0	1.1
EGCV1	23.2	1.1	185
EGCV2	23.0	1.4	184
EGCV3	23.3	1.4	186
$\widehat{c}$	0.98	0.83	1.20

Table 2: Summary of MC results excluding $|\alpha| = 1$.

Variable	Mean	Minimum	Maximum
EGAV1	1.53	1.4	1.7
EGAV2	1.03	1.0	1.2
EGAV4,3	4.33	2.0	18
EGCV2	15.2	1.3	140
EGCV3	15.4	1.3	142
$\widehat{c}$	0.99	0.83	1.20

Table 3: Summary of MC results excluding $\alpha = 0$.

Tables 1-3 give some summary statistics on EGv as well as on $\widehat{c}$, the OLS estimated value of the control coefficient of CV3 in (22). They cover different ranges of the parameter

α because of the occasional distortions to the simple descriptive statistics introduced by the two extremes 0 and 1 of $|\alpha|$.[3] EGCV1 is absent from Tables 1 and 3 because it excludes $|\alpha| = 1$ by definition. AV4,3 excludes $\alpha = 0$ and so is absent from Tables 1 and 2. On the whole, the following remarks can be made from the raw data (not listed here) and the summary tables:

Remark 1. AV2 is the weakest VRT, while AV1 is the second weakest.[4] They are the only ones to show a slightly better performance for $|\alpha| = 1$ than for $|\alpha| < 1$. All the others deteriorate (albeit from better initial levels) when $|\alpha| \to 1$, with AV1 performing better than CVs and almost as well as AV4,3 when $|\alpha| = 1$. On the other hand, the performance of AV2 is quite poor and declines as T increases, and has a relatively high memory cost. We would tend not to recommend methods that just reorder the innovations if the purpose is variance reduction in dynamic MC. Such methods include the Bootstrap which has nevertheless been successfully put to other uses. □

Remark 2. Sign-switching AVs (AV3 and AV4) are extremely efficient and provide staggering efficiency gains. Their performance peaks at $\alpha = 0$ where EG is infinite, and declines exponentially to stabilize from $|\alpha| \approx 0.75$ onwards. Both are undoubtedly the best VRTs for unit roots where they provide an efficiency gain of 2; but CVs outperform them in the middle range of α, especially as T rises there. The average figure of 2 indicates by (4) and (5) that the two antithetics $\widehat{\alpha}^-$ of AV3 and $\widehat{\alpha}^-$ of AV4 are independent of $\widehat{\alpha}$ when $|\alpha| \approx 1$, a finding that will be established analytically in Section 5. □

Remark 3. CV2 is less volatile than CV1 because the latter performs better as $\alpha \to 0$ but is rather poor at the other extreme ($|\alpha| \to 1$) for understandable reasons (by construction). The reversal in ranking occurs at $|\alpha| \approx 0.8$ for $T < 100$, and at $|\alpha| \approx 0.9$ for $T \geq 100$. The higher $|\alpha|$ in the second case is explained by the smaller asymptotic relative difference between the normalization (h) of the two CVs; compare (20) with (21). As expected from the optimum estimation of c, CV3 is by definition the most efficient of all three CVs for each and every α, T. This property could be useful when N is small. In addition, it is more flexible than CV1 because it can cope with any value of α. A common feature of the CVs is that they suffer a marked deterioration of performance as $|\alpha| \to 1$ and T increases; a result in line with the analysis at the end of Section 2. □

Remark 4. The average value of $\widehat{c}$ (the OLS estimator of the optimum control coefficient) is not statistically different from 1. The estimate of c is closest to 1 when $|\alpha| \approx 1$, CV2 and CV3 becoming almost identical. The gain (at little extra cost) from using CV3 instead of CV2 comes from $|\alpha| \ll 1$ where CV2 is even weaker than CV1; and it increases as T and/or N fall. The estimated $\widehat{c}$ rises stochastically as α and T increase, albeit with increased volatility. This increased volatility is due to the problems of simulating near nonstationarity where values of N need to be larger than usual to achieve a given level of accuracy (e.g. see Lai and Siegmund (1983, last column of their Table 2) for a related problem). This need for high N does not contradict the speedy convergence of unit-root distributions to their asymptotics. It just means that the price paid for faster convergence is increased volatility. To understand this, compare the standardization factors for $(\widehat{\alpha} - \alpha)$ when $|\alpha|$ goes from being a stable to a unit root: multiplying $(\widehat{\alpha} - \alpha)$ by T instead of $\sqrt{T}$ amplifies the variability of the resulting statistic from one replication to the other. The problem becomes very serious in explosive series with $|\alpha| > 1$ where the normalization factor is exponential (proportional to $|\alpha|^T$) *and* depends on the initial observation of the

[3] The response surfaces given below are more sophisticated, and do not share these distortions.

[4] In this section, AV1 refers to the narrow definition where the residual-rearrangement and sign-switching features of the general definition of the method in (9) are temporarily ignored because they are covered by AV2 and AV3/AV4, respectively.

series (y_0). The lesson is that one should be careful in drawing conclusions from simulations of a system of series with significantly varying degrees of nonstationarity, especially if N is fixed at the same level for all the series that are being simulated. $\square$

The remainder of this section will be devoted to summarizing the features of each new VRT by means of a Response Surface (RS). The RS ascertains the average behavior of efficiency gains as α and T change, so as to understand and predict the performance of the VRTs. Moreover as explained in Hendry (1984), a RS reduces the specificity of MC experiments. Letting $p_1 = 1, 2, \ldots$ and $p_2 = 0, 1, \ldots$, the explanatory variables are of the form $|\alpha|^{Tp_1}/T^{p_2/2}$, $\sinh(\alpha^2)/T^{p_1/2}$, and $\cosh(\alpha^2)T^{p_1/2}$, with variables that are insignificant at the 5% level dropped from the RS. In the case of AVs, we also have the limiting EGs (that will be derived in Section 5) as additional explanatory variables; whereas in the case of CVs, we add separate functions of $|\alpha|$ and T.

Hyperbolic functions can be defined as

$$\cosh(z) := \frac{1}{2}(\mathrm{e}^{z} + \mathrm{e}^{-z}) \quad \text{and} \quad \sinh(z) := \frac{1}{2}(\mathrm{e}^{z} - \mathrm{e}^{-z}). \tag{24}$$

Their appearance is in line with some analytical results on autoregressions; see Abadir (1993b). Powers of $|\alpha|^T$ can be thought of as dummies representing unit roots asymptotically and near nonstationarity when T is finite; since:

$$\begin{aligned} |\alpha|^T &= 1 \text{ for } |\alpha| = 1 \\ |\alpha|^T &\approx 1 \text{ for } |\alpha| \approx 1 \text{ and } T < \infty \\ |\alpha|^T &\approx 0 \text{ for } |\alpha| \ll 1, \text{ or for } |\alpha| < 1 \text{ and } T \to \infty. \end{aligned} \tag{25}$$

This takes care of the different nature of distributional results for the cases of $|\alpha| \approx 1$ and $|\alpha| \ll 1$. Such terms explain why there will be no evidence of structural breaks in the RSs when the root is near unity, in spite of the radically different asymptotic theories required by the two extremes 0 and 1 of $|\alpha|$. Powers of $|\alpha|^T$ could be used as weights in analytically solving the problem of finding closed forms for general finite-sample distributions whose range of applicability includes $|\alpha| \neq 1$ as well as $|\alpha| = 1$.[5] There is some evidence that this approach could bear fruit since finite-sample distributions are mixtures of the limiting normal and the distribution under $|\alpha| = 1$, with greater weight assigned to the latter for typical T. For example, compare Phillips' (1977, 1978a, and especially 1978b) finite sample approximations for the distributions when $|\alpha| < 1$ with the formulae in Abadir (1993a, 1995), or see the numerical results of Evans and Savin (1981). Abadir (1995) tackles this problem from a different angle (an asymptotic one): unit root distribution functions provide the general formulae that are encompassing generalizations of the standard normal which arises in the limit for the Studentized t statistic when $|\alpha| \neq 1$. The $|\alpha|^T$ term can also be interpreted as an exponential proxy for Hendry's (1984, p.965) notion of "effective sample size", extended to the case when $|\alpha| \geq 1$. Hendry's effective sample size, $T(1-\alpha^2)$, is only valid for $|\alpha| < 1$. For a different generalization of this term to include $|\alpha| \neq 1$, see Abadir (1993c). The disadvantage of the latter generalization is that it is discontinuous at $|\alpha| = 1$ and is useful only in asymptotic analysis, unlike the continuum provided by $|\alpha|^T$ throughout the range of α and T.

The RS of EGAV1 is

$$\begin{aligned} &\log(\mathrm{EGAV1}) \\ &= -\log(5+\alpha^2) + \underset{\substack{(0.005)\\ [0.004]}}{2.07} + \underset{\substack{(0.02)\\ [0.02]}}{0.21}\, |\alpha|^T - \underset{\substack{(0.19)\\ [0.19]}}{0.82}\, |\alpha|^T/\sqrt{T} + \underset{\substack{(0.07)\\ [0.06]}}{0.36}\, \sinh(\alpha^2)/\sqrt{T} \end{aligned} \tag{26}$$

[5] See Abadir (1992) for an asymptotic encompassing formula for the distributions of all statistics when $|\alpha| = 1$.

$$\bar{R}^2 = 0.83, \qquad \text{S} = 0.0218, \qquad \text{CH}(4, 42) = 1.77,$$

where:
(.) = conventional standard errors,
[.] = White's (1980) heteroskedasticity-consistent standard errors,
$\bar{R}^2$ = the coefficient of determination after a degrees-of-freedom adjustment,
S = residual standard error,
CH(a, b) = Chow's (1960) test for structural break at $|\alpha| = 1$, distributed as F(a, b).

The first term of the RS is its asymptotic value when $|\alpha| < 1$, up to an additive constant. It is not estimated: the coefficient is restricted in accordance with the asymptotics of Section 5.2 (notice that $\log 8 = 2.08$). There is no evidence of heteroskedasticity, as is seen from the small difference between the two types of standard errors. Furthermore, the 5% critical value for F$(4, 42)$ is 2.59, so there is no break at $|\alpha| = 1$. The fit is good, but it can be improved if one were to include more variables at the cost of interpretability and parsimony. Here, we can see that increasing $|\alpha|^T$ improves the efficiency gain (the gain increases with α, but declines with T especially when α is small), which was noted in Remark 1 above. The coefficient of the last term works in the same direction, while the impact of $|\alpha|^T / \sqrt{T}$ is smaller than the other terms because of the effect of a large T.

AV2 could be summarized by a RS similar to (26), but there is little point in doing so here because of the breakdown of this method when $T > 100$, and because of its relative cost; see (10) and the ensuing discussion.

The promising features of AV3 and AV4 can be described by the following RS, where the first term is again from the asymptotics of Section 5.2:

$$\log(\text{EGAV4,3}) = \log(1 + \alpha^{-2}) - \underset{\substack{(0.0051)\\ [0.0049]}}{0.0170} + \underset{\substack{(0.03)\\ [0.03]}}{0.28} \cosh(\alpha^2)/\sqrt{T} \tag{27}$$

$$\bar{R}^2 = 0.9998, \qquad \text{S} = 0.0163, \qquad \text{CH}(2, 46) = 1.17,$$

where the five data points pertaining to $\alpha = 0$ have been replaced by five others that were generated under $|\alpha| = 0.1$, because $\alpha \to 0$ leads to an infinite EG which creates a precision problem for the regression. Notice that the figures for $\bar{R}^2$ are somewhat inflated by the large variation in the left-hand side, especially as $\alpha \to 0$. However, there is no break in this relation as $|\alpha| \to 1$, since the critical 5% value of F$(2, 46)$ is 3.20. Together, the asymptotics of Section 5.2 and hyperbolic functions are seen here to dramatically explain a lot of what happens as α changes.

We now turn to CVs. The RS for all three are quite similar; to save space, only the RS of CV3 (which is superior to the other two) will be reported.

$$\begin{aligned}
&\log(\text{EGCV3}) \\
&= \underset{\substack{(0.54)\\ [0.97]}}{10.63} - \underset{\substack{(6.05)\\ [5.75]}}{41.27} /\sqrt{T} - \underset{\substack{(0.37)\\ [0.64]}}{6.79} \cosh(\alpha^2) + \underset{\substack{(5.69)\\ [4.64]}}{33.47} \cosh(\alpha^2)/\sqrt{T} - \underset{\substack{(2.18)\\ [1.15]}}{7.68} \sinh(\alpha^2)/\sqrt{T} \\
&+ \underset{\substack{(0.030)\\ [0.031]}}{0.17} (1 - |\alpha|)\log T - \underset{\substack{(0.11)\\ [0.08]}}{0.72} \left(1 - |\alpha|^T\right) + \underset{\substack{(0.0065)\\ [0.0097]}}{0.079} \left(1 - |\alpha|^T\right)\sqrt{T}
\end{aligned} \tag{28}$$

$$\bar{R}^2 = 0.99, \qquad \text{S} = 0.107, \qquad \text{CH}(5, 42) = 1.19.$$

Note that all the terms in the second line are zero when $\alpha = 1$. Note also that for $\alpha = 1$ and $T \to \infty$, one finds that log(EGCV3) is predicted to converge to the finite number

$10.63 - 6.79\cosh(1) = 0.15$, i.e. (28) predicts that EGCV3 converges to 1.16. On the other hand, when $|\alpha| < 1$ (stationary case), the elements on the second line diverge as $T \to \infty$ and $\log(\text{EGCV3})/\sqrt{T} \to 0.079$, i.e. EGCV3 tends to ∞.

Here also, the fit is good and there is no evidence of either structural breaks at $|\alpha| = 1$ or heteroskedasticity. Though still small, S is higher in (28) than in any other RS presented in this work, due to the more volatile performance of CVs relative to the AVs considered above. The fit of (28) appears to be worse for α very cose to 1, possibly due to the different behavior for $|\alpha| = 1$ and $|\alpha| < 1$ as $T \to \infty$ discussed above; see also Section 5.

4 Uses and extensions of the VRTs

Caution should be exercised when employing VRTs in a dynamic context. There is a marked change in efficiency gains as $|\alpha|$ and T vary jointly. Contrary to what classical asymptotic theory suggests, doubling the sample size will have effects that are dependent on the *level* of T instead of just halving the standard error of the modified estimator $\widetilde{\alpha}$. The RS of the previous section can be used to avoid such pitfalls by predicting how VRTs fare as $|\alpha|$ and T change.

For increased efficiency gains when $\alpha \neq 0$, one may combine more than one of the VRTs described above, depending on their correlation and on the specific problem at hand (e.g. magnitudes of α and T). The best method of doing so was detailed by Davidson and MacKinnon (1992) who show that the optimal combination of VRTs will be in the form of a regression run in the same spirit as (22). In order for the researcher not to waste valuable time generating VRTs that are not useful for the problem at hand, a priori selection of VRTs should be made according to the criteria given earlier. These VRTs should then be the inputs of the aforementioned regression.

Combinations of VRTs of different *types* are likely to do best, when using the method of Davidson and MacKinnon (1992). This is an interesting area for future research. Given their relatively strong performance in dynamic DGPs, CV3 and one of AV3 or AV4 are the most prominent candidates. Moreover, because AV3 and AV4 have a good performance also for unit roots while CV3 does well for intermediate persistence, combining the two types of VRTs gives a smoother performance as persistence changes. Also, the exceptionally good performance of AV3 and AV4 when T is small should remedy the weakness of CV3 in small samples. In the more specific event of a nearly nonstationary DGP, the narrowly interpreted AV1 –which is the only VRT (apart from the erratic AV2) to improve as $|\alpha|$ gets closer to 1– may also be added to the combination to yield even more efficiency gains.

The AVs presented earlier are special cases of what we can call an orthogonal-transform AV. In particular, AV1-AV3 are all special cases involving premultiplication of $\boldsymbol{\varepsilon} := (\varepsilon_1, ..., \varepsilon_T)'$ by an orthogonal matrix, hence $\boldsymbol{\varepsilon}$ and its transformation have the same first two moments. These orthogonal matrices were, respectively:

1. $\boldsymbol{R} := \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2, \dots)$, where $\boldsymbol{Q}_i := \boldsymbol{A}_{p_i}\boldsymbol{B}_{\omega_i}$ $(i = 1, \dots, T/2)$ with

   $$\boldsymbol{A}_p := \begin{pmatrix} (-1)^p & 0 \\ 0 & 1 \end{pmatrix}, \qquad p \in \{0,1\}; \qquad \boldsymbol{B}_\omega := \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix}, \qquad \omega \in (0, 2\pi).$$

   The 4 possible AV1s are obtained setting $p_i = p \in \{0,1\}$ and $\omega_i = \omega = \{\frac{1}{4}\pi, \frac{5}{4}\pi\}$.[6] The matrix $\boldsymbol{A}_p$ changes sign to the first innovation, while $\boldsymbol{B}_\omega$ describes a rotation with angle ω. If $\boldsymbol{Q}_i = \boldsymbol{Q}_j$ for all i, j, then $\boldsymbol{R} = \boldsymbol{I}_{T/2} \otimes \boldsymbol{Q}_1$. This class contains many variants; a simple set of k distinct $\boldsymbol{R}$ matrices, denoted by $\boldsymbol{R}_1, \dots, \boldsymbol{R}_k$, can for instance be obtained by fixing $p_i = p \in \{0,1\}$ and choosing angles $\omega_1 < \omega_2 < \cdots < \omega_k$ in $(0, 2\pi)$.

[6] One could also vary p, ω with i, but we will not do so here.

2. $\boldsymbol{S} := (\boldsymbol{e}_T, \boldsymbol{e}_{T-1}, \ldots, \boldsymbol{e}_1)$, where $\boldsymbol{e}_i$ is the i-th column of the identity matrix $\boldsymbol{I}_T$. More generally, $\boldsymbol{S}$ can be chosen as any permutation of the columns of $\boldsymbol{I}_T$; i.e. a permutation matrix having exactly one element equal to 1 in any row (equivalent to sampling from $\boldsymbol{\varepsilon}$ without replacement). A simple set of k distinct $\boldsymbol{S}$ matrices, denoted $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k$, can for instance be obtained by choosing k different permutations of the ordered set $\{1, 2, \ldots T\}$.

3. $\boldsymbol{T} := \operatorname{diag}(-1, 1, -1, \ldots)$, where the typical element of the diagonal is $(-1)^i$. This class contains many variants; a simple set of k distinct $\boldsymbol{T}$ matrices, denoted $\boldsymbol{T}_1, \ldots, \boldsymbol{T}_k$, can for instance be obtained by setting $\boldsymbol{T}_i := \operatorname{diag}(-\boldsymbol{I}_i, \boldsymbol{I}_i, -\boldsymbol{I}_i, \ldots)$, where the last block is possibly truncated so as to fit in a $T \times T$ matrix. (This corresponds to generating the AV $\{(-1)^{\kappa(t)}\varepsilon_t\}$, with $\kappa(t) = \lfloor \frac{t}{i} + 1 \rfloor$). It is also possible to switch the sign randomly, but independently of $\boldsymbol{\varepsilon}$, rather than do it deterministically.

Although VRTs were only mentioned so far in connection with estimating the moments of econometric estimators, it is possible to use them on a wider scale. For example, Rothery (1982) reduced the uncertainty in estimating the MC power of a test by using the known power of another as the control variate in a manner similar to CV1 and CV2 where c is fixed at 1.[7] Rothery's technique could be extended to allow for the estimation of $\widehat{c}$ as in CV3. Another example is in Durbin and Koopman (1997, espec. pp. 674-676), where independence properties of the normal distribution are exploited to generate a new class of AVs. CVs based on Taylor series are also given there.

VRTs can also be used to improve the accuracy of numerical integration routines (but see Fang, Wang and Bentler (1994) for an alternative approach). Using MC in numerical integration has become part of the Bayesian tool kit for analyzing posterior densities since the seminal paper by Kloek and van Dijk (1978). Bayesian and other integration intensive applications stand to benefit from using the new VRTs.

Another illustration of the wide and not necessarily standard applicability of VRTs can be obtained by using the antithetic variates $\widehat{\alpha}^-$ as if they were actual estimates $\widehat{\alpha}$ from another replication, thus saving the time taken to generate half of the replications. This procedure can be used in any simulation work involving $\widehat{\alpha}$, such as estimating the distributions of *any* statistic associated with $\widehat{\alpha}$, under a null or under an alternative (e.g. power).

The dependence of $\widehat{\alpha}^-$ and $\widehat{\alpha}$ can be exploited as discussed above, see Davidson and MacKinnon (1992). When $\widehat{\alpha}^-$ and $\widehat{\alpha}$ happen to be independent, one can use them simply as different replicates; this is what happens for AV3 and AV4 in the case of near nonstationarity, for all values of T, since EGAV4,3 ≈ 2 by RS (27) or by Remark 2 on the results at the beginning of Section 3; see also the following Section 5 for an analytic proof.

5 Large-sample analytics

In this section, we study the asymptotic distribution of $(\widehat{\alpha}, \widehat{\alpha}^-)$ when $T \to \infty$. We hence obtain the limiting EGs for the new AVs, see the discussion at the end of Section 2. A similar analysis is also applied to study the limit behavior of EGs for CVs. We discuss AVs in Sections 5.1 and 5.2 for the nonstationary and the stationary cases, respectively; we treat CVs in Section 5.3.

We start by introducting notation for AVs. Denote the initial sequence $\{\varepsilon_t\}$ as $\{\varepsilon_{1,t}\} := \{\varepsilon_t\}$, and we let $\{\varepsilon_{2,t}\} := \{\varepsilon_t^-\}$ indicate the AV innovations. Conformably, we indicate by $\{y_{1,t}\}$ and $\{y_{2,t}\}$ the corresponding values of y_t generated by the DGP (1). In order to emphasize the dependence on T, we indicate by $\widehat{\alpha}_{1,T}$ the crude MC estimate of α based on

[7] This method assumes that the distribution of the test used as CV can be calculated explicitly; see Paruolo (2002) on the influence of estimation of critical values in comparing test powers by simulation.

$\{y_{1,t}\}$ and by $\widehat{\alpha}_{2,T}$ the one obtaind from $\{y_{2,t}\}$. The combined estimator is indicated by $\widetilde{\alpha}_T := \frac{1}{2}\left(\widehat{\alpha}_{1,T} + \widehat{\alpha}_{2,T}\right)$, where $\mathrm{E}\left(\widehat{\alpha}_{1,T}\right) = \mathrm{E}\left(\widetilde{\alpha}_T\right) = \mathrm{E}\left(\widehat{\alpha}_{2,T}\right)$.

In the light of the discussion of the last section, we extend the definition of AV1 to include all transformation of the type $\boldsymbol{B}_\omega$, excluding the sign-swithing effect associated with $\boldsymbol{A}_1$.[8] Specifically, recall that T is even, and let $t = 1, \ldots, T$, and $m = 1, \ldots, T/2$. For AV1 we fix a given angle $\omega \in (0, 2\pi)$, and we employ the following pairs of definitions for t odd ($t = 2m - 1$) and t even ($t = 2m$):

$$\begin{aligned}\varepsilon_{2,2m-1} &:= \cos(\omega)\,\varepsilon_{1,2m-1} + \sin(\omega)\,\varepsilon_{1,2m} \\ \varepsilon_{2,2m} &:= -\sin(\omega)\,\varepsilon_{1,2m-1} + \cos(\omega)\,\varepsilon_{1,2m}.\end{aligned}$$

Moreover, recall that in the present notation $\varepsilon_{2,t} := \varepsilon_{1,T-t}$ for AV2 and $\varepsilon_{2,t} := (-1)^t \varepsilon_{1,t}$ for AV3. AV4 is defined differently, but gives the same AV $\widehat{\alpha}_{2,T}$ as AV3, so we do not distinguish AV3 from AV4 in this section, and we refer to them as AV4,3.

We let $\widehat{\boldsymbol{\theta}}_T := (\widehat{\alpha}_{1,T}, \widehat{\alpha}_{2,T})'$ be the 2×1 vector of α estimates, so that $\widetilde{\alpha}_T = \frac{1}{2}\boldsymbol{\imath}'\widehat{\boldsymbol{\theta}}_T$ where $\boldsymbol{\imath} := (1,1)'$ and consider limits as $T \to \infty$. Because of the consistency of estimators $\widehat{\alpha}_{i,T}$, i.e. $\operatorname{plim}\widehat{\alpha}_{i,T} = \alpha$, also $\operatorname{plim}\widetilde{\alpha}_T = \alpha$. Moreover if $\widehat{\boldsymbol{\theta}}_T$ satisfies some limit theorem of the type $\boldsymbol{a}_T := T^q(\widehat{\boldsymbol{\theta}}_T - \alpha\boldsymbol{\imath}) \stackrel{w}{\to} \boldsymbol{a}_\infty = O_p(1)$ (where $q = 1$ for $\alpha = 1$ and $q = \frac{1}{2}$ for $|\alpha| < 1$), then also $T^q\left(\widetilde{\alpha}_T - \alpha\right) \stackrel{w}{\to} \frac{1}{2}\boldsymbol{\imath}'\boldsymbol{a}_\infty = O_p(1)$, and the limit distribution of $\widetilde{\alpha}_T$ can be deduced from the one of $\widehat{\boldsymbol{\theta}}_T$. Hence, one can use the limit random vector $\boldsymbol{a}_\infty$ as an approximation to the distribution of $\widehat{\boldsymbol{\theta}}_T$ for finite T.

Let $\boldsymbol{m}_T := \mathrm{E}\left(\boldsymbol{a}_T\right)$, $\boldsymbol{V}_T := \operatorname{var}\left(\boldsymbol{a}_T\right)$, and note that $\operatorname{var}(\widehat{\boldsymbol{\theta}}_T) = T^{-2q}\boldsymbol{V}_T$. Substituting in (6) one finds

$$\operatorname{var}(\widetilde{\alpha}_T) = \frac{1}{4T^{2q}}\boldsymbol{\imath}'\boldsymbol{V}_T\boldsymbol{\imath}. \tag{29}$$

Let also $\boldsymbol{\Sigma} := \operatorname{var}\left(\boldsymbol{a}_\infty\right)$. Under regularity conditions, $\boldsymbol{V}_T$ converges to $\boldsymbol{\Sigma}$ as $\boldsymbol{a}_T \stackrel{w}{\to} \boldsymbol{a}_\infty$. If these conditions are met, then one can use $\boldsymbol{\Sigma}$ as guidance for $\boldsymbol{V}_T$ in (29).

Each AVj implies a different $\boldsymbol{V}_T$ and $\boldsymbol{\Sigma}$. We label $\boldsymbol{\Sigma}_j$ the limiting variance matrix for AVj and indicate by ρ_j the implied correlation coefficient. One can discuss which AV methods is bound to give the highest variance reduction in (29) for large values of T by comparing

$$\boldsymbol{\imath}'\boldsymbol{\Sigma}_j\boldsymbol{\imath} \tag{30}$$

for $j = 1$, 2, and 4,3. This comparison can be based on values of the correlation ρ_j in the case of equal variances, see (7). In particular we indicate the limit value of EGv as

$$\mathrm{EGAV}j_\infty := \frac{2}{1 + \rho_j}. \tag{31}$$

5.1 AVs, nonstationary case

Let $\alpha = 1$ and define the normalized partial sums for $i = 1$, 2

$$s_i(u) := \sigma^{-1}T^{-1/2}\sum_{t=1}^{\lfloor Tu \rfloor}\varepsilon_{it} = \sigma^{-1}T^{-1/2}y_{i,\lfloor Tu \rfloor},$$

[8] The effect of switching the sign is already investigated through the study of AV4,3.

and $\boldsymbol{s}(u) := (s_1(u), s_2(u))'$. In this case $q = 1$ in the normalization $\boldsymbol{a}_T := T^q(\widehat{\boldsymbol{\theta}}_T - \alpha\boldsymbol{\imath})$, where

$$
\begin{aligned}
a_{iT} := T(\widehat{\alpha}_{iT} - 1) &= \frac{\sum_t s_i\left(\frac{t-1}{T}\right)\left(s_i\left(\frac{t}{T}\right) - s_i\left(\frac{t-1}{T}\right)\right)}{\frac{1}{T}\sum_t s_i\left(\frac{t-1}{T}\right)^2} \\
&= \frac{\frac{1}{2}\left(s_i(1)^2 - \sum_t\left(s_i\left(\frac{t}{T}\right) - s_i\left(\frac{t-1}{T}\right)\right)^2\right)}{\frac{1}{T}\sum_t s_i\left(\frac{t-1}{T}\right)^2},
\end{aligned}
$$

which is a functional of $s_i(u)$ for u in $U_T := \{\frac{t}{T}, t = 2, ..., T\}$. One has $a_{iT} = \psi(s_i) + o_p(1)$, where the functional ψ is defined as

$$
\psi(x) := \frac{\frac{1}{2}\left(x^2(1) - 1\right)}{\int_0^1 x^2(\tau)\,\mathrm{d}\tau}
$$

for any function $x(\tau)$ in the space $D[0,1]$ of cadlag functions on [0,1]. We notice that ψ is a continuous functional in the sup metric, provided the denominator is different from 0 a.s., which holds in the case of x equal to Brownian motions (BMs). Note that for the partial sums s_i, one has

$$
\psi(s_i) = \frac{\frac{1}{2}\left(s_i(1)^2 - 1\right)}{\frac{1}{T}\sum_t s_i\left(\frac{t-1}{T}\right)^2}.
$$

We also note that $\psi(x)$ is an even functional, in the sense that, given the function $x := x(\tau)$ and its reflection across the time axis $-x := -x(\tau)$, then $\psi(-x) = \psi(x)$, because $x(\tau)^2 = (-x(\tau))^2$, $0 \leq \tau \leq 1$, both in the numerator and denominator of ψ.

We wish to characterize the dependence between a_{1T} and a_{2T} for large T. In order to do this we study the weak limits of partial sums $s_1(u)$ and $s_2(u)$ in the following lemma. We denote by $W(u)$ a vector Brownian motion with variance $\boldsymbol{V}$, i.e. with $W(u) \sim \mathrm{N}(\boldsymbol{0}, u\boldsymbol{V})$ and by $B(u)$ a standard vector Brownian motion, i.e. $W(u)$ with $\boldsymbol{V} = \boldsymbol{I}_2$.

Lemma 1. *When $\alpha = 1$, the following holds, for various AVj.*

1. *For AV1, $\boldsymbol{s}(u) \xrightarrow{w} W(u)$ where $W(u)$ is a Brownian motion with*

$$
\boldsymbol{V} = \begin{pmatrix} 1 & \cos(\omega) \\ \cos(\omega) & 1 \end{pmatrix}.
$$

2. *For AV2,*

$$
\begin{pmatrix} s_1(u) \\ s_2(u) \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \int_0^u \mathrm{d}B_1(\tau) \\ \int_{1-u}^1 \mathrm{d}B_1(\tau) \end{pmatrix} = \begin{pmatrix} B_1(u) \\ B_1(1) - B_1(1-u) \end{pmatrix},
$$

   i.e. both $s_1(u)$ and $s_2(u)$ converge to a univariate standard Brownian motion, with the property that the second one, $B_1(1) - B_1(1-u)$, is the Brownian motion obtained by reversing the time of the first one, $B_1(u)$.

3. *For AV4,3,*

$$
\boldsymbol{s}(u) \xrightarrow{w} B(u)
$$

   where $B(u)$ is a standard vector Brownian motion.

Proof. Cases AV1 and AV4,3: we apply a Cramér-Wold device. Consider real λ_i, $i = 1, 2$, $\boldsymbol{\lambda} := (\lambda_1, \lambda_2)'$ and define

$$
s_3(u) := \boldsymbol{\lambda}'\boldsymbol{s}(u) = \sum_{i=1}^{2} \lambda_i s_i(u) = \sigma^{-1}T^{-1/2}\sum_{t=1}^{\lfloor Tu \rfloor} c_t \varepsilon_{1t},
$$

where for AV1 one finds $c_{2m-1} = \lambda_1 + \lambda_2 (\cos(\omega) - \sin(\omega))$, $c_{2m} = \lambda_1 + \lambda_2 (\cos(\omega) + \sin(\omega))$, $m = 1, \dots, T/2$, while for AV4,3 $c_t = \lambda_1 + (-1)^t \lambda_2$. Applying the univariate FCLT for martingale differences, see Brown (1971), one sees that

$$s_3(u) \xrightarrow{w} W_3(u) \qquad u = [0, 1]$$

with $W_3(u)$ a Brownian motion with variance $v := \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} c_t^2$. For AV1,

$$\begin{aligned} v &= \frac{1}{2}\left((\lambda_1 + \lambda_2 (\cos(\omega) - \sin(\omega)))^2 + (\lambda_1 + \lambda_2 (\cos(\omega) + \sin(\omega)))^2 \right) \\ &= \lambda_1^2 + 2(\cos(\omega)) \lambda_1 \lambda_2 + \lambda_2^2 = \boldsymbol{\lambda}' \boldsymbol{V} \boldsymbol{\lambda} \end{aligned}$$

which implies the expression of $\boldsymbol{V}$ given in the statement of case 1. For AV4,3

$$v = \lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2 \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} (-1)^t = \lambda_1^2 + \lambda_2^2 = \boldsymbol{\lambda}' \boldsymbol{I}_2 \boldsymbol{\lambda},$$

where we have used the fact that for integer p one has

$$\sum_{i=1}^{2p} (-1)^i = 0. \tag{32}$$

This shows convergence of the finite dimensional distributions of $\boldsymbol{s}(u)$ to $B(u)$. Tightness follows as in Brown (1971) Theorem 3.

Case AV2 follows as a direct application of time reversal of random walks and the associated limit BM. □

Remark 5. AV1 generates correlated BMs. The correlation can be chosen to be positive or negative by choice of the angle ω. For instance $\omega = \pi/4$ gives a correlation of $\sqrt{2}/2$ while $\omega = 3\pi/4$ gives a correlation of $-\sqrt{2}/2$. In order to maximize correlations one can choose a small number δ and set $\omega = \delta$ for highly positive and $\omega = \pi - \delta$ for highly negative correlations.

Note, however, that the correlations generated by choice of ω may not help variance reduction for the functional ψ, because it is even. Note in fact, that if two univariate BMs W_1 and W_2 have correlation ρ, then W_1 and $-W_2$ have correlation $-\rho$. However, $\psi(W_2) = \psi(-W_2)$, so that generating BMs with positive or negative correlations will have the same effect on $\widetilde{\alpha}_T$ in the limit.

For the choice $\omega = \pi/4$ used in previous sections, we have estimated ρ_1 by simulating $\psi(s_1)$ and $\psi(s_2)$ for $T = 1000$, 5000, 10000 and $N = 10000$, and calculating the correlation between MC replications. We have obtained estimates of ρ_1 equal to 0.24, 0.23, 0.24 for the 3 values of T. We can hence infer that the large T value of ρ_1 is about 0.24. This gives $\text{EGAV1}_\infty = 2/1.24 = 1.61$, while the RS of Section 3 gave essentially the same because $\exp(-\log(6) + 2.07 + 0.21) = 1.63$. □

Remark 6. AV2 generates time-reversed BMs, B_1 and $B_1^{\hookleftarrow}$ say. Note that the functional $\psi(x)$ depends on the last value of the argument function x, which is equal for both BM, and on the area under the squared path of x, which is not equal for B_1 and $B_1^{\hookleftarrow}$. In fact by a simple change of variable $t = 1 - \tau$, one has

$$\begin{aligned} \int_0^1 B_1^{\hookleftarrow}(\tau)^2 \, \mathrm{d}\tau &= \int_0^1 (B_1(1) - B_1(1-\tau))^2 \, \mathrm{d}\tau = \int_0^1 (B_1(1) - B_1(t))^2 \, \mathrm{d}t \\ &= \int_0^1 B_1(t)^2 \, \mathrm{d}t + B_1(1)^2 - 2B_1(1) \int_0^1 B_1(t) \, \mathrm{d}t. \end{aligned}$$

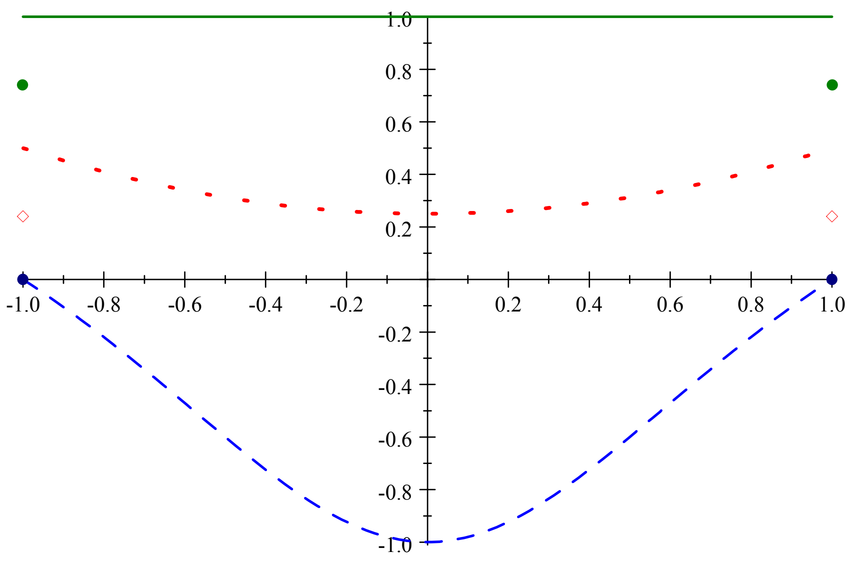

Figure 1: AV correlations as function of α: ρ_1 (dotted line and diamonds, $\omega = \frac{1}{4}\pi$), ρ_2 (solid line and filled dots) and ρ_3 (dashed line and filled dots).

Hence one expects AV2 replicates $\widehat{\alpha}_{1,T}$, $\widehat{\alpha}_{2,T}$ to be correlated.

Also for AV2, we have estimated ρ_2 by simulating $\psi(s_1)$ and $\psi(s_2)$ for $T = 1000$, 5000, 10000 and $N = 10000$, and calculating the correlation between MC replications. We have obtained estimates of ρ_2 equal to 0.74 for all 3 values of T. We can hence infer that the large T value of ρ_2 is about 0.74. This gives $\mathrm{EGAV2}_\infty = 2/1.74 = 1.15$. □

Remark 7. AV4,3 generates independent BMs B_1 and B_2. Hence also $\psi(B_1)$ and $\psi(B_2)$ will be independent, and hence ρ_3 in (31) is null. This gives $\mathrm{EGAV4{,}3}_\infty = 2$; while the RS of Section 3 gave $\exp(\log(2) - 0.0170) = 1.97$.

The same remarks given above apply to AV4,3 when $(-1)^t$ is substituted by some other centered periodic function of t, e.g. like $(-1)^{\kappa(t)}$, $\kappa(t) = \left\lfloor \frac{t}{i} + 1 \right\rfloor$. In this case $2p$ needs to be replaced by ip in (32), where i is the period of the periodic function.

One can enquire how fast this independence is attained for AV4,3 in simulations with finite T. To this end, (32) above shows that $s_1\left(\frac{t}{T}\right)$ and $s_2\left(\frac{t}{T}\right)$ are independent under Gaussian ε_{1t} also for finite T, when t is even. When t is odd, $\mathrm{E}\left(s_1\left(\frac{t}{T}\right) s_2\left(\frac{t}{T}\right)\right) = \frac{1}{T}$. □

5.2 AVs, stationary case

Let $|\alpha| < 1$, and set $q = \frac{1}{2}$ in the normalization of $\boldsymbol{a}_T := T^{1/2}(\widehat{\boldsymbol{\theta}}_T - \alpha\boldsymbol{\imath})$, where

$$a_{iT} := T^{1/2}\left(\widehat{\alpha}_{iT} - \alpha\right) := \frac{N_{iT}}{D_{iT}} := \frac{T^{-1/2}\sum_t y_{i,t-1}\varepsilon_{i,t}}{T^{-1}\sum_t y_{i,t-1}^2}.$$

In the case of $|\alpha| < 1$, y_t is stationary and ergodic, i.e.

$$D_{iT} := T^{-1}\sum_t y_{i,t-1}^2 \overset{a.s.}{\to} D_i := \mathrm{E}\left(y_{i,t}^2\right) = \frac{\sigma^2}{1-\alpha^2}.$$

It is hence natural to consider the approximation $a_{iT} = b_{iT} + o_p(1)$ with

$$b_{iT} := \frac{N_{iT}}{D_i} := \frac{T^{-1/2}\sum_t y_{i,t-1}\varepsilon_{i,t}}{\mathrm{E}\left(y_{i,t}^2\right)}.$$

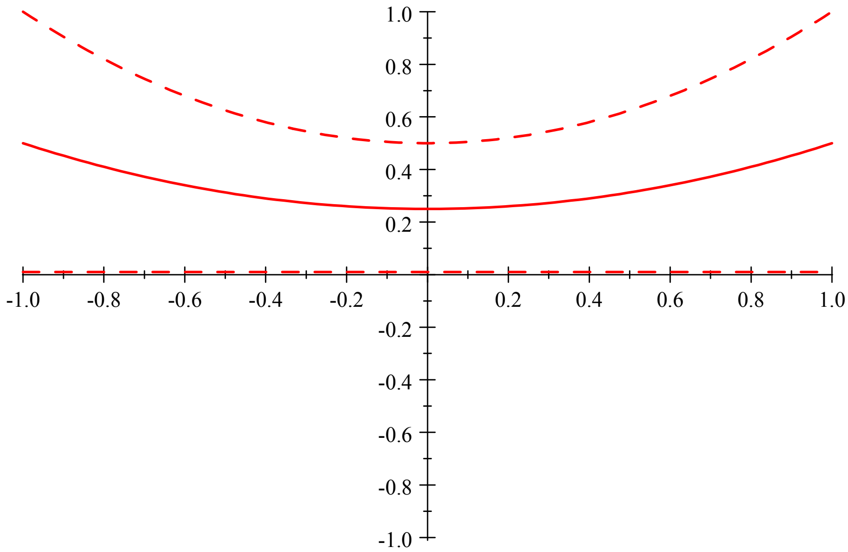

Figure 2: Influence of ω on ρ_1 as a function of α, $|\alpha| < 1$: upper dashed line $\omega = 0$, solid line $\omega = \frac{1}{4}\pi$, lower dashed line $\omega = \frac{1}{2}\pi$.

Note that $\mathrm{E}(b_{i,T}) = 0$ and

$$\mathrm{var}\left(b_{i,T}\right) = \frac{\sigma^2}{TD_i^2} \sum\nolimits_t \mathrm{E}\left(y_{i,t-1}^2\right) \to \frac{\sigma^2}{D_i} = 1 - \alpha^2.$$

The following lemma summarizes results in the stationary case.

Lemma 2. *When $|\alpha| < 1$, $\boldsymbol{b}_T := (b_{1,T}, b_{2,T})'$ is $O_p(1)$ for $T \to \infty$, and*

$$\mathrm{var}\left(\boldsymbol{b}_T\right) \to \boldsymbol{\Sigma} := \left(1 - \alpha^2\right) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where the values of the correlation term ρ are given below for the various AVj, using the notation ρ_j for AVj:

$$\text{for } AV1,\ \rho_1 = \frac{1+\alpha^2}{2} \cos^2 \omega \text{ and } EGAV1_\infty = \frac{4}{2 + (1+\alpha^2)\cos^2 \omega};$$

$$\text{for } AV2,\ \rho_2 = 1 \text{ and } EGAV2_\infty = 1;$$

$$\text{for } AV4,3,\ \rho_3 = -\frac{1-\alpha^2}{1+\alpha^2} \text{ and } EGAV4,3_\infty = 1 + \frac{1}{\alpha^2}.$$

Proof. The proof of this lemma is given in the Appendix. □

The correlation terms ρ_j are continuous functions in α for $-1 < \alpha < 1$, but not all ρ_j are continuous as $|\alpha| \to 1$; see Fig.1. It is likely that these could be bridged if a DGP with local-to-unity parameter is adopted; e.g. see Phillips (1987) for a definition of such a DGP and its relation to ours. For the current setup, we can note the following:

Remark 8. The $AV1$ correlation ρ_1 is always positive, and hence implies not-so-big variance reductions. It tends to $\cos^2 \omega$ when $|\alpha| \to 1$ and to $\frac{1}{2}\cos^2 \omega$ for $\alpha = 0$. For $\omega = \pi/4$, for instance $\rho_1 \to \frac{1}{2}$ for $|\alpha| \to 1$ and $\rho_1 = \frac{1}{4}$ for $\alpha = 0$. The correlation ρ_1 is discontinuous at $|\alpha| = 1$. From Remark 5 above, $\rho_1 = 0.24$ when $\alpha = 1$, $\omega = \pi/4$. Hence, $\lim_{\alpha \to 1} \rho_1 = \frac{1}{2} \neq 0.24$ and EG improves when $|\alpha|$ reaches 1. □

Remark 9. The choice of ω influences the variance reduction properties of $AV1$. When $\omega \to 0$, the performance is worst, because $\cos^2 \omega = 1$; for $\omega = \pi/2$ the performance is best, because in this case $\rho_1 = 0$ for all α. This is shown graphically in Fig. 2 for $-1 < \alpha < 1$.

The case $\omega = \pi/2$ implies that the antithetic ε_{2t} is formed by interchanging consecutive ε_{1t} and changing the sign for odd t, i.e. $\varepsilon_{2,2m} = -\varepsilon_{1,2m-1}$, $\varepsilon_{2,2m-1} = \varepsilon_{1,2m}$. Comparing this to the effect of $AV4,3$, which just changes the sign for odd t, we see that interchanging consecutive ε_{1t} has an adverse effect on variance reduction. □

Remark 10. The $AV2$ correlation ρ_2 equals 1 for all $|\alpha| < 1$. Hence it gives no variance reduction. The correlation ρ_2 is discontinuous at $|\alpha| = 1$. From Remark 6 above, $\rho_2 = 0.74$ when $\alpha = 1$. Hence, $\lim_{\alpha \to 1} \rho_2 = 1 \neq 0.74$ and EG improves when $|\alpha|$ reaches 1. □

Remark 11. The $AV4,3$ correlation ρ_3 equals 0 when $|\alpha| \to 1$ and equals -1 for $\alpha = 0$. It gives negative correlations for all $-1 < \alpha < 1$, and hence implies bigger variance reductions than $AV1$. The correlation ρ_3 is continuous at $\alpha = 1$. In fact $\lim_{\alpha \to 1} \rho_3 = 0$ where $\psi(s_1)$ and $\psi(s_2)$ are also independent thanks to Lemma 1, which implies $\rho_3 = 0$ when $\alpha = 1$. $AV4,3$ has the best performance in terms of EGv, and also presents a continuous behavior as $|\alpha| \to 1$. □

5.3 CVs

We next study the EG of control variates for large T. We consider the case of CV3 as a representative case; minor modifications apply for other CVs. Let $\widehat{\alpha} - \alpha = N_T/D_T$, where $N_T := T^{-1/2} \sum_t y_{t-1}\varepsilon_t$ and $D_T := T^{-1} \sum_t y_{t-1}^2$ with a notation similar to the one in Section 5.2. By properties of least squares, the fact that $\mathrm{E}(N_T) = 0$, and using the discussion in Davidson and MacKinnon (1992) bottom of page 206, one finds that

$$\mathrm{EGCV3}_T = \frac{1}{1 - \xi_T^2}, \quad \text{where } \xi_T := \mathrm{corr}\left(\frac{N_T}{D_T}, \frac{N_T}{T^{-1/2}h}\right) = \mathrm{corr}\left(\frac{N_T}{D_T}, N_T\right). \tag{33}$$

Note that, being defined as a correlation coefficient, ξ_T is scale-invariant in both arguments; both can then be scaled appropriately and independently as T increases.

Equation (33) applies both for $\alpha = 1$ and for $|\alpha| < 1$. Under regularity conditions, when T grows large and $|\alpha| < 1$, $D_T \overset{a.s.}{\to} D$, a constant and hence ξ_T converges to $\mathrm{corr}\,(N_T/D, N_T) = \mathrm{corr}\,(N_T, N_T) = 1$. This implies that $\mathrm{EGCV3}_T \to \infty$ for $|\alpha| < 1$, in accordance with the prediction of (28).

When $\alpha = 1$ instead, both N_T/D_T and N_T – when appropriately scaled – have the non-degenerate weak limits $\frac{1}{2}\left(B_1^2(1) - 1\right) / \int_0^1 B_1^2(\tau)\mathrm{d}\tau$ and $\frac{1}{2}\left(B_1^2(1) - 1\right)$ respectively, where B_1 is a standard Brownian motion, see Section 5.1. Hence under regularity conditions ξ_T converges to

$$\xi := \mathrm{corr}\left(\frac{\frac{1}{2}\left(B_1^2(1) - 1\right)}{\int_0^1 B_1^2(\tau)\mathrm{d}\tau}, \frac{1}{2}\left(B_1^2(1) - 1\right)\right).$$

We have simulated ξ discretizing the Brownian motion as a random walk with T steps for $T = \{10^3, 10^4, 10^5, 5 \times 10^5\}$, using 10^5 replications; we obtained the following values for ξ: 0.5080, 0.5076, 0.5036, 0.4972. Regressing these values of ξ on a constant, $T^{-1/2}$, and T^{-1}, we get the prediction of $\xi = 0.4969$ for $T = \infty$. On the basis on this estimate, one finds

$$\log \mathrm{EGCV3}_\infty = \log \frac{1}{1 - \xi^2} = \log \frac{1}{1 - 0.4969^2} = 0.28.$$

This prediction is similar to the prediction of formula (28). A test that $\gamma_1 + \gamma_3 \cosh(1) = 0.28$ in regression (28), where γ_1 and γ_3 indicate the contant and the coefficient of $\cosh(\alpha^2)$, gave an F(1,47) statistics of 1.70 with p-value of 0.198; hence the prediction of (28) and of (33) are not statistically different for $\alpha = 1$, $T = \infty$.

6 Conclusions

This paper has investigated the effectiveness of a few VRTs in the context of dynamic DGPs. The best performers were new VRTs, with some providing staggering efficiency gains and thus potential for large time and/or precision gains. In general, the most efficient of all was AV4, a new VRT that relies on generating antithetic parameter estimates by inverting the sign of these parameters in the DGP. The resulting antithetic variate was always nonpositively correlated to the crude MC estimates, thereby reducing the variance of the combined estimator by large factors of 2 or more. Equally good but for the slower generation of estimators, AV3 was overall next best to AV4. AV3 relied on Hendry and Harrison's (1974) untried suggestion of changing the sign of every other residual.

This and many other methods that reformulate innovations were shown to be special cases of the new encompassing orthogonal-transform AV, which has a simple and convenient general formulation. Another special case of it was AV2 which applied some of the Jackknife/Bootstrap philosophy to variance reduction in dynamic MC. It seemed the least promising VRT for autoregressive series. Clearly however, this does not preclude the successful use of the Jackknife/Bootstrap philosophy in other areas and towards other ends.

Finally, the performance of CVs –as represented by the optimum CV3– was encouraging the closer α was to 0, in which case it improved with T. Large efficiency gains of 15 times were quite common, though AV3 and AV4 outperformed CVs whenever $|\alpha|$ was in the neighborhood of 0 or 1, particularly when T was not large.

The invariance of the performance of AV3 and AV4 with respect to T, and their independence from $\widehat{\alpha}$ when $|\alpha| \approx 1$ meant that they could be used for the fast generation of additional nearly nonstationary data for any sample size, thus reducing simulation times. This is a new and hitherto unknown general function that some VRTs may now serve in simulation under certain conditions.

The benefit derived from this unconventional application of VRTs was to allow the possibility of saving a significant amount of time in any future MC study of ARMA series with one autoregressive root near unity and stable/invertible remaining roots. The exact method for doing so was described in the latter part of Section 4, and it shows that simulating nearly nonstationary series is not as uniquely problematic as it might have seemed earlier: in spite of the need for a larger number of replications to counteract the significant increase in MC volatility as $|\alpha|$ and T increase, AV3 and AV4 could be used unconventionally to compensate for such a requirement when $|\alpha| \approx 1$. Other VRTs can be used to fulfill such a function for cases when $|\alpha| > 1$. AV1 is one candidate, given its improved performance for α^2 close to 1.

In addition to summarizing the results, the response surfaces used in the paper had the following benefits. First, they allowed for a smaller number of replications than would otherwise be needed because they reduce specificity. Second, they predict the range of beneficial application of a VRT by calculating the magnitude of the efficiency gain. Third, by comparing fitted and actual values, they detect features like increased MC volatility as $|\alpha|$ and T grow. The implication of this latter result for simulating Wiener processes is particularly important. These processes are typically generated by discrete random walks ($\alpha = 1$) with a large number of observations ($T \to \infty$). These are precisely the two ingredients that will increase MC variability, and one needs to be aware that a number of replications that is larger than usual is needed in that context. More generally, normalized Ornstein-Uhlenbeck processes are approximated by DGP (1), and their accuracy can now be controlled better.

It can be shown analytically that the results of this paper are applicable to vector autoregressions and error-correction mechanisms. For technical details of the necessary matrix transformations, see Abadir, Hadri and Tzavalis (1999). The magnitude of efficiency gains will depend on a mixture of the eigenvalues of an autoregressive matrix. Stable systems will behave like a stable AR, purely nonstationary systems like a random walk. Cointegrated systems will produce linear combinations that depend primarily on the extent of a rank-deficiency parameter.

References

Abadir, K.M. (1992), A distribution generating equation for unit-root statistics, Oxford Bulletin of Economics and Statistics, 54, 305-323.

Abadir, K.M. (1993a), The limiting distribution of the autocorrelation coefficient under a unit root, Annals of Statistics, 21, 1058-1070.

Abadir, K.M. (1993b), OLS bias in a nonstationary autoregression, Econometric Theory, 9, 81-93.

Abadir, K.M. (1993c), On the asymptotic power of unit root tests, Econometric Theory, 9, 187-219.

Abadir, K.M. (1995), The limiting distribution of the t ratio under a unit root, Econometric Theory, 11, 775-793.

Abadir, K.M., K. Hadri and E. Tzavalis (1999), The influence of VAR dimensions on estimator biases, Econometrica, 67, 163-181.

Brown B.M. (1971), Martingale central limit theorems, Annals of Mathematical Statistics, 42, 59-66.

Chow, G.C. (1960), Tests of equality between sets of coefficients in two linear regressions, Econometrica, 28, 591-605.

Cryer, J.D., J.C. Nankervis and N.E. Savin (1989), Mirror-image and invariant distributions in ARMA models, Econometric Theory, 5, 36-52.

Davidson, R. and J.G. MacKinnon (1992), Regression-based methods for using control variates in Monte Carlo experiments, Journal of Econometrics, 54, 203-222.

Durbin, J. and S.J. Koopman (1997), Monte Carlo maximum likelihood estimation for non-Gaussian state space models, Biometrika, 84, 669-684.

Ericsson, N.R. (1991), Monte Carlo methodology and the finite sample properties of instrumental variables statistics for testing nested and non-nested hypotheses, Econometrica, 59, 1249-77.

Evans, G.B.A. and N.E. Savin (1981), Testing for unit roots: 1, Econometrica, 49, 753-779.

Fang, K.-T., Y. Wang and P.M. Bentler (1994), Some applications of number-theoretic methods in statistics, Statistical Science, 9, 416-428.

Haldrup, N. (1996), Mirror image distributions and the Dickey-Fuller regression with a maintained trend, Journal of Econometrics, 72, 301-312.

Hendry, D.F. (1984), Monte Carlo experimentation in econometrics, in: Z. Griliches and M.D.Intriligator (eds), Handbook of Econometrics, vol.2, Amsterdam: North-Holland.

Hendry, D.F. and R.W. Harrison (1974), Monte Carlo methodology and the small sample behaviour of ordinary and two-stage least squares, Journal of Econometrics, 2, 151-174.

Kendall, M. and A. Stuart (1977), The advanced theory of statistics, vol.1. 4th ed. London: Charles Griffin & Co.

Kleijnen, J.P.C. (1974), Statistical techniques in simulation, part I, New York: Marcel Dekker.

Kloek, T. and H.K. van Dijk (1978), Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, Econometrica, 46, 1-19.

Lai, T.L. and D. Siegmund (1983), Fixed accuracy estimation of an autoregressive parameter, Annals of Statistics, 11, 478-485.

Paruolo, P. (2002), On Monte Carlo estimation of relative power, Econometrics Journal, 5, 65-75.

Phillips, P.C.B. (1977), Approximations to some finite sample distributions associated with a first-order stochastic difference equation, Econometrica, 45, 463-485.

Phillips, P.C.B. (1978a), Edgeworth and saddlepoint approximations in the first-order noncircular autoregression, Biometrika, 65, 91-98.

Phillips, P.C.B. (1978b), A note on the saddlepoint approximation in the first order noncircular autoregression, Cowles Foundation Discussion Paper No.487, Yale University.

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

Rothery, P. (1982), The use of control variates in Monte Carlo estimation of power, Applied Statistics, 31, 125-129.

Said, S.E. and D.A. Dickey (1984), Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika, 71, 599-607.

Takahashi, A. and N. Yoshida (2005), Monte Carlo simulation with asymptotic method, Journal of Japanese Statistical Society, 35, 171-203.

White, H. (1980), A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica, 48, 817-838.

White, J.S. (1958), The limiting distribution of the serial correlation coefficient in the explosive case, Annals of Mathematical Statistics, 29, 1188-1197.

White, J.S. (1961), Asymptotic expansions for the mean and variance of the serial correlation coefficient, Biometrika, 48, 85-94.

Appendix

Proof of Lemma 2. For all AVs, we observe that $b_{1,T}$ and $b_{2,T}$ satisfy the central limit theorem for martingale differences, Brown (1971) Theorem 2; specifically $b_{1,T} \xrightarrow{w} \mathrm{N}(0, 1-\alpha^2)$. This implies that $\boldsymbol{b}_T := (b_{1,T}, b_{2,T})'$ is $O_p(1)$. In order to calculate $\mathrm{var}\,(\boldsymbol{b}_T)$, we note that $\mathrm{var}\,(b_{i,T}) \to 1-\alpha^2$ and that

$$\mathrm{cov}(b_{1,T}, b_{2,T}) = \frac{1}{T}\frac{A}{D_1 D_2} \tag{34}$$

where $A = \sum_{s=2}^{T}\sum_{t=2}^{T} a_{st}$ and $a_{st} := \mathrm{E}\left(y_{1,s-1}\varepsilon_{1,s}y_{2,t-1}\varepsilon_{2,t}\right)$.

We compute A below in different subsections for each AV. Let $\mathrm{E}_t\left(\cdot\right)$ indicate expectations conditional on $\mathcal{F}_t := \sigma\{\varepsilon_{1,t-j}, j \geq 0\}$. We also define $\gamma_{12}\left(t\right) := \mathrm{E}\left(y_{1,t}y_{2,t}\right)$ and $\gamma_{12} := \lim_{t\to\infty}\gamma_{12}\left(t\right)$. In the following we only need that $\varepsilon_{1,t}$ has 0 third moment; this is implied by simmetry and existence of 3-rd absolute moments, as for the case of Gaussian $\varepsilon_{1,t}$.

AV1

Let $a := d := \cos\omega$, $b := -c = \sin\omega$. We note that $\varepsilon_{2,t}$ depends on $\varepsilon_{1,t-1}$, $\varepsilon_{1,t}$, for t even and $\varepsilon_{1,t}$, $\varepsilon_{1,t+1}$ for t odd; hence $y_{2,t}$ is $\mathcal{F}_t$-measurable for t even and $\mathcal{F}_{t+1}$-measurable for t odd. More precisely for $t = 2m-1$ (t odd) one has

$$\varepsilon_{2,t} = a\varepsilon_{1,t} + b\varepsilon_{1,t+1}, \qquad y_{2,t} = a\varepsilon_{1,t} + b\varepsilon_{1,t+1} + \alpha y_{2,t-1} \tag{35}$$

where $y_{2,t-1} \in \mathcal{F}_{t-1}$, whereas for $t = 2m$ (t even)

$$\varepsilon_{2,t} = c\varepsilon_{1,t-1} + d\varepsilon_{1,t}, \qquad y_{2,t} = g\varepsilon_{1,t} + h\varepsilon_{1,t-1} + \alpha^2 y_{2,t-2} \tag{36}$$

where $g := d + \alpha b$, $h := c + \alpha a$ and $y_{2,t-2}$ is $\mathcal{F}_{t-2}$-measurable. Sometimes in the following the observation $y_{2,t}$ is $\mathcal{F}_t$-measurable for t even suffices.

We first prove that $a_{st} = 0$ for $t + 2 \leq s$ and $s \leq t - 2$, i.e.

$$A = \sum_{t=2}^{T} \sum_{s=t-1}^{t+1} a_{st} =: A_1 + A_2 + O(1). \tag{37}$$

where $A_1 := \sum_{m=1}^{T/2-1} (a_{2m-2,2m-1} + a_{2m-1,2m-1} + a_{2m,2m-1})$ includes the terms for t odd and $A_2 := \sum_{m=1}^{T/2-2} (a_{2m-1,2m} + a_{2m-1,2m} + a_{2m+1,2m})$ the one for t even. In fact for $t+1 \leq s-1$ one has $y_{2,t-1}$is $\mathcal{F}_t$-measurable $\varepsilon_{2,t}$ is $\mathcal{F}_{t+1}$-measurable, and $\mathcal{F}_{t+1} \subseteq \mathcal{F}_{s-1}$

$$a_{st} = \mathrm{E}\left(y_{1,s-1}\varepsilon_{1,s}y_{2,t-1}\varepsilon_{2,t}\right) = \mathrm{E}\left(\mathrm{E}_{s-1}\left(\varepsilon_{1,s}\right)y_{1,s-1}y_{2,t-1}\varepsilon_{2,t}\right) = \mathrm{E}(0) = 0$$

For $s = t - 2$ one has instead

$$a_{st} = \mathrm{E}\left(y_{1,s-1}\varepsilon_{1,s}y_{2,t-1}\varepsilon_{2,t}\right) = \mathrm{E}\left(y_{1,t-3}\varepsilon_{1,t-2}\,\mathrm{E}_{t-2}\left(y_{2,t-1}\varepsilon_{2,t}\right)\right) = 0$$

because $\mathrm{E}_{t-2}\left(y_{2,t-1}\varepsilon_{2,t}\right) = 0$; this follows from the fact that $\varepsilon_{2,t}$ is uncorrelated by construction with $y_{2,t-1}$ given the past, and because $\varepsilon_{2,t}$ (which depends on $\varepsilon_{1,t-1}$, $\varepsilon_{1,t}$, $\varepsilon_{1,t+1}$) is unaffected by conditioning on $\mathcal{F}_{t-2}$. The same argument holds when setting $s < t - 2$.

We next consider a_{st} for $t-1 \leq s \leq t+1$, distinguishing the cases of t odd or even. For $s = t - 1$ and t even one finds

$$a_{st} = \mathrm{E}\left(y_{1,t-2}\varepsilon_{1,t-1}\left(a\varepsilon_{1,t-1} + b\varepsilon_{1,t} + \alpha y_{2,t-2}\right)\left(c\varepsilon_{1,t-1} + d\varepsilon_{1,t}\right)\right) = \alpha c\sigma^2\gamma_{12}\left(t-2\right).$$

Here we have used the fact that $\varepsilon_{1,t}$ has 0 third moment. For $s = t - 1$ and t odd one finds $a_{st} = \mathrm{E}\left(y_{1,t-2}\varepsilon_{1,t-1}y_{2,t-1}\left(a\varepsilon_{1,t} + b\varepsilon_{1,t+1}\right)\right) = 0$ by conditioning on $\mathcal{F}_{t-1}$.

For $s = t$ and t even one finds

$$a_{st} = \sigma^4\left(ac + db\right) + \sigma^4\alpha^2 d\gamma_{12}\left(t-2\right) = \sigma^4\alpha^2 d\gamma_{12}\left(t-2\right)$$

because $ac + db = 0$, see (35), (36). Again here we have used the fact that $\varepsilon_{1,t}$ has 0 third moment. For $s = t$ and t odd one finds $a_{st} = a\sigma^2\gamma_{12}\left(t-1\right)$.

For $s = t + 1$ and t even, one has $a_{st} = 0$, while for $s = t + 1$ and t odd one finds $a_{st} = \alpha b\sigma^2\gamma_{12}\left(t-1\right)$. Summarizing:

$$\begin{aligned}
A_1 &= \sum_{m=1}^{T/2-1}\left(a_{2m-1,2m-1} + a_{2m,2m-1}\right) = \sum_{m=1}^{T/2-1}\left(a + \alpha b\right)\sigma^2\gamma_{12}\left(2m-2\right), \\
A_2 &= \sum_{m=1}^{T/2-2}\left(a_{2m-1,2m} + a_{2m,2m}\right) = \sum_{m=1}^{T/2-2}\left(\alpha c + \alpha^2 d\right)\sigma^2\gamma_{12}\left(2m-2\right), \\
A &= \left(1+\alpha^2\right)\cos\left(\omega\right)\sigma^2\sum_{m=1}^{T/2-2}\gamma_{12}\left(2m-2\right) + O(1),
\end{aligned}$$

where $a + \alpha b + \alpha c + \alpha^2 d = \left(1+\alpha^2\right)\cos\omega$ and we note that the covariances $\gamma_{12}\left(t\right)$ are summed for t even.

In order to calculate $\gamma_{12}\left(t\right)$ for t even, write $y_{2,t} = \alpha^2 y_{2,t-2} + \eta_t$ where $\eta_t := g\varepsilon_{1,t} + h\varepsilon_{1,t-1}$, $g := d + \alpha b$, $h := c + \alpha a$ see (36). Next set $t = 2m$, write $y_{2,2m} = \alpha^2 y_{2,2(m-1)} + \eta_{2m}$, and solve the recursions in m to find

$$y_{2,2m} = \sum_{i=0}^{m-1}\alpha^{2i}\left(g\varepsilon_{1,2(m-i)} + h\varepsilon_{1,2(m-i)-1}\right).$$

Represent $y_{1,t}$ in a similar way as

$$y_{1,2m} = \sum_{i=0}^{m-1} \alpha^{2i} \left(\varepsilon_{1,2(m-i)} + \alpha \varepsilon_{1,2(m-i)-1}\right).$$

Hence, because $b + c = 0$,

$$\gamma_{12}(2m) = \sum_{i=0}^{m-1} \mathrm{E}\left(\alpha^{2i}\left(\varepsilon_{1,2(m-i)} + \alpha\varepsilon_{1,2(m-i)-1}\right)\right)\left(\alpha^{2i}\left(g\varepsilon_{1,2(m-i)} + h\varepsilon_{1,2(m-i)-1}\right)\right) =$$

$$= \sigma^2 g \sum_{i=0}^{m-1} \alpha^{4i} + \sigma^2 \alpha h \sum_{i=0}^{m-1} \alpha^{4i} = \sigma^2 (g + \alpha h) \frac{1-\alpha^{4m}}{1-\alpha^4} = \sigma^2 \cos\omega \frac{1-\alpha^{4m}}{1-\alpha^2}$$

Moreover for $T \to \infty$

$$\sum_{m=0}^{T/2-2} \gamma_{12}(2m) = \frac{\sigma^2\cos\omega}{1-\alpha^2}\left(\frac{T}{2} - 2 - \frac{1-\alpha^{2T-4}}{1-\alpha^4}\right) = \frac{T}{2}\frac{\sigma^2\cos\omega}{1-\alpha^2} + O(1) = \frac{T}{2}\gamma_{12} + O(1)$$

where $\gamma_{12} := \left(1-\alpha^2\right)^{-1}\sigma^2\cos\omega = \lim_{m\to\infty}\gamma_{12}(2m)$. Hence

$$T^{-1}A = \frac{1}{2}\left(1+\alpha^2\right)\left(1-\alpha^2\right)^{-1}\cos^2(\omega)\,\sigma^4 + O(1)$$

and substituting into (34)

$$\mathrm{cov}(b_{1,T}, b_{2,T}) = \frac{A}{TD_1D_2} \to \frac{\cos^2(\omega)}{2}\left(1+\alpha^2\right)\left(1-\alpha^2\right).$$

AV2

Recall that for AV2 one has $\varepsilon_{2,t} = \varepsilon_{1,T-t}$; the following representations hold:

$$y_{1,s-1} = \sum_{j=0}^{s-2} \alpha^j \varepsilon_{1,s-1-j}, \qquad y_{2,t-1} = \sum_{i=0}^{t-2} \alpha^i \varepsilon_{2,t-1-i} = \sum_{i=0}^{t-2} \alpha^i \varepsilon_{1,T-t+1+i}$$

and hence

$$a_{st} := \mathrm{E}\left(y_{1,s-1}\varepsilon_{1,s}y_{2,t-1}\varepsilon_{2,t}\right) = \sum_{i=0}^{t-2}\sum_{j=0}^{s-2} \alpha^{i+j}\,\mathrm{E}\left(\varepsilon_{1,s-1-j}\varepsilon_{1,s}\varepsilon_{1,T-t+1+i}\varepsilon_{1,T-t}\right). \tag{38}$$

Indicate subscripts in the last term in (38) as follows $n_1 := s-1-j$, $n_2 := s$, $n_3 := T-t+1+i$, $n_4 := T-t$. In order for the expectation on the r.h.s. of (38) to be nonzero, n_1, ..., n_4 must be equal in pairs. Note in fact that they cannot be all equal because of the presence of at least 1 lag between the terms that originate from t, $t-1$, and s, $s-1$. One has 2 cases: case 1, with $n_1 = n_3$, $n_2 = n_4$, and case 2, with $n_1 = n_4$, $n_2 = n_3$.

In case 1 one has $s-1-j = T-t+1+i$ and $s = T-t$ implies $i = -(j+2)$ which is outside the range of $i = 0$, ..., $t-1$. This is because if $s = T-t$, then $y_{1,s-1}$ involves the past while $y_{2,t-1}$ involves the future, with no overlap. Hence case 1 gives 0 contribution to a_{st}.

Consider next case 2, with $s-1-j = T-t$ and $s = T-t+1+i$. This is equivalent to $i = j$ and $s-T+t-1 = i$. Consider a given fixed t; then there are as many values of s as $i = 0$, ..., $t-2$ with $s \geq T-t+1$ for which $a_{st} = \alpha^{2i}\sigma^4$. Hence

$$T^{-1}A = \frac{\sigma^4}{T}\sum_{t=2}^{T}\sum_{i=0}^{t-2}\alpha^{2i} = \frac{\sigma^4}{T}\sum_{t=2}^{T}\frac{1-\alpha^{2(t-1)}}{1-\alpha^2} = \frac{\sigma^4}{1-\alpha^2} + o(1)$$

and thus

$$\mathrm{cov}(b_{1,T}, b_{2,T}) = \frac{1}{T}\frac{A}{D_1D_2} \to \frac{\sigma^4}{1-\alpha^2}\frac{\left(1-\alpha^2\right)^2}{\sigma^4} = 1-\alpha^2.$$

AV4,3

We proceed as for AV1, recalling that for AV4,3 one has $\varepsilon_{2,t} = (-1)^t \varepsilon_{1,t}$. In this case all terms

$$a_{st} := \mathrm{E}\left(y_{1,s-1}\varepsilon_{1,s}y_{2,t-1}\varepsilon_{2,t}\right)$$

are equal to 0 for $s < t$ by conditioning on $\mathcal{F}_{t-1}$. Simmetrically for $t < s$ one has $a_{st} = 0$ by conditioning on $\mathcal{F}_{s-1}$. Hence one is left with

$$a_t := a_{tt} = \mathrm{E}\left(y_{1,t-1}\varepsilon_{1,t}y_{2,t-1}(-1)^t \varepsilon_{1,t}\right) = (-1)^t \sigma^2 \,\mathrm{E}\left(y_{1,t-1}y_{2,t-1}\right) = (-1)^t \sigma^2 \gamma_{12}(t-1).$$

Moreover, one finds for $t = 2p$ (t even)

$$\begin{aligned}\gamma_{12}(t) = \mathrm{E}(y_{1,t}y_{2,t}) &= \mathrm{E}\left(\sum_{i=0}^{t-1} \alpha^i \varepsilon_{1,t-i}\right)\left(\sum_{i=0}^{t-1} \alpha^i (-1)^{t-i} \varepsilon_{1,t-i}\right) = \\ &= \sum_{i=0}^{t-1} \alpha^{2i}(-1)^{t-i}\,\mathrm{E}(\varepsilon_{1,t-i}^2) = (-1)^t \sigma^2 \sum_{i=0}^{t-1} \alpha^{2i}(-1)^i = \frac{1-\alpha^{2t}}{1+\alpha^2}\sigma^2,\end{aligned}$$

where we have used the following fact, listing first the even and then the odd terms,

$$\sum_{i=0}^{t-1} \alpha^{2i}(-1)^i = \sum_{m=0}^{p-1}\left(\alpha^{2\cdot 2m} - \alpha^{2\cdot(2m+1)}\right) = \left(1-\alpha^2\right)\frac{1-\alpha^{4p}}{1-\alpha^4} = \frac{1-\alpha^{4p}}{1+\alpha^2}.$$

When t is odd we use recursions and the previous expression to find

$$\begin{aligned}\gamma_{12}(t) &= \mathrm{E}(y_{1,t}y_{2,t}) = \mathrm{E}\left(\left(\alpha y_{1,t-1} + \varepsilon_{1,t}\right)\left(\alpha y_{2,t-1} + (-1)^t \varepsilon_{1,t}\right)\right) \\ &= \alpha^2\,\mathrm{E}\left(y_{1,t-1}y_{2,t-1}\right) - \sigma^2 = \left(\frac{1-\alpha^{2t-2}}{1+\alpha^2}\alpha^2 - 1\right)\sigma^2 \\ &= \frac{\alpha^2 - \alpha^{2t} - 1 - \alpha^2}{1+\alpha^2}\sigma^2 = -\frac{1+\alpha^{2t}}{1+\alpha^2}\sigma^2.\end{aligned}$$

Hence substituting in a_t for t even ($t-1$ odd) one finds $a_t = \sigma^2\gamma_{12}(t-1) = -\sigma^4\frac{1+\alpha^{2t}}{1+\alpha^2}$, while for t odd ($t-1$ even) one has $a_t = -\sigma^2\gamma_{12}(t-1) = -\sigma^4\frac{1-\alpha^{2t}}{1+\alpha^2}$. Hence, recalling that T is even

$$A = \sum_{t=2}^{T} a_t = \sum_{m=2}^{T/2} a_{2m-1} + \sum_{m=1}^{T/2} a_{2m} = -\frac{T}{2}\frac{\sigma^4}{1+\alpha^2} - \frac{T}{2}\frac{\sigma^4}{1+\alpha^2} + O(1).$$

Hence $T^{-1}A \to -\sigma^4\left(1+\alpha^2\right)^{-1}$ and therefore

$$\mathrm{cov}(b_{1,T}, b_{2,T}) = \frac{1}{T}\frac{A}{D_1 D_2} \to -\sigma^4 \frac{1}{1+\alpha^2}\frac{\left(1-\alpha^2\right)^2}{\sigma^4} = -\frac{1-\alpha^2}{1+\alpha^2}\left(1-\alpha^2\right).$$