International Journal of Forecasting 36 (2020) 761–780

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Bias corrections for exponentially transformed forecasts: Are they worth the effort?

Matei Demetrescu [a], Vasyl Golosnoy [b], Anna Titova [a,*]

[a] Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40, D-24118 Kiel, Germany
[b] Faculty of Management and Economics, Ruhr-Universität Bochum, Universitätsstr. 150, D-44801, Bochum, Germany

ARTICLE INFO

Keywords:
Bias correction
Linear autoregression
Linex forecast
Log transformation
Volatility forecast

ABSTRACT

In many economic applications, it is convenient to model and forecast a variable of interest in logs rather than in levels. However, the reverse transformation from log forecasts to levels introduces a bias. This paper compares different bias correction methods for such transformations of log series which follow a linear process with various types of error distributions. Based on Monte Carlo simulations and an empirical study of realized volatilities, we find no choice of correction method that is uniformly best. We recommend the use of the variance-based correction, either by itself or as part of a hybrid procedure where one first decides (using a pretest) whether the log series is highly persistent or not, and then proceeds either without bias correction (high persistence) or with bias correction (low persistence).

1. Introduction

Taking logs is used widely in applied (time series) econometrics for linearizing relations or stabilizing variances. It has become a standard transformation for time series in numerous economic and financial applications; see among others Andersen, Bollerslev, and Huang (2011), Bauer and Vorkink (2011), Brechmann, Heiden, and Okhrin (2018), Golosnoy, Okhrin, and Schmid (2012), Hautsch (2012), Lütkepohl and Xu (2012), Mayr and Ulbricht (2015), or Gribisch (2018). Indeed, models in logs often turn out to be better suited for both estimation and forecasting. Then, the forecast in logs should be transformed back in order to predict the original variable of interest. However, exponential transformations introduce a point forecast bias into the procedure, as has been emphasized already by Granger and Newbold (1976).

From a practical point of view, the relevant question is how such bias should be dealt with in finite samples for various types of distributions for log-model errors. Of course, ignoring the bias and simply transforming the forecast in logs through the exponential function is one possible course of action, albeit a naïve one, since ignoring the bias may lead to substantial losses in forecasting precision (cf. Lütkepohl & Xu, 2012; Proietti & Lütkepohl, 2013). At the other end of the spectrum of possibilities, one finds bootstrap-based corrections (cf. Thombs & Schucany, 1990); however, these are computationally demanding and not always easy to implement. In regard to interval forecasts, note that the required quantiles remain unbiased after any monotone transformation, so that bias correction for exponentially transformed point forecasts is indeed the relevant problem at hand.

This paper compares several bias correction procedures which are of high practical applicability (cf. Stock & Watson, 2012, pp. 314–315). Concretely, we consider a popular method that exploits the residual variance for the bias correction, as well as one that relies on computing the sample mean of exponentially-transformed log-model residuals. Whereas the variance-based correction is optimal for normally distributed innovations, the mean-based correction only requires the existence of the relevant expectation. We also examine a semiparametric

* Corresponding author.
E-mail addresses: mdeme@stat-econ.uni-kiel.de (M. Demetrescu), vasyl.golosnoy@rub.de (V. Golosnoy), atitova@stat-econ.uni-kiel.de (A. Titova).

https://doi.org/10.1016/j.ijforecast.2019.09.001
0169-2070/

approach based on the estimation of the model in logs under the Linex loss (Varian, 1975), which we show to provide asymptotically unbiased forecasts of the original (untransformed) series, so that no correction is necessary in this case. However, the Linex-based approach exploits a non-linear estimation procedure that could cause losses in estimation efficiency compared to maximum likelihood estimation; such efficiency losses may impact the behavior seriously in finite samples.

We study settings with linear data generating processes in logs, with model errors following different types of distributions. The interest here lies in producing one-step-ahead forecasts of the original variable, but our analysis could be applied easily to the task of predicting at longer horizons. We compare the effectiveness of the above-mentioned bias correction methods with that of the naïve approach without any adjustment. The forecasting performances of different correction methods have been already studied in several settings, e.g. for a family of data generating processes with Markov switching (cf. Patton & Timmermann, 2007). We extend this strand of the literature by focusing specifically on error distributions that exhibit deviations from normality, which are of high empirical relevance. In particular, we study the effects of skew-normal (Azzalini, 1985), mixture normal (Everitt & Hand, 1981; McLachlan & Peel, 2004), contaminated normal (Seidel, 2011) and t-distributed innovations (cf. Tarami & Pourahmadi, 2003). Since we investigate several autoregressive and ARMA models with different degrees of persistence, our setup covers a broad selection of models that are relevant in practice.

We find, first, that the variance-based correction appears to be the preferable approach in smaller samples, even for a range of deviations from normality; the expectation-based correction of the residual exponent is a close competitor. Second, despite being attractive from a theoretical point of view, the Linex-based approach that requires no specific correction shows losses in estimation efficiency. It appears to be dominated by two of the above-mentioned alternatives in terms of the forecasting loss functions considered, but becomes competitive with an increasing sample size as the estimation error diminishes. Third, a naïve prediction without bias correction is found to be suitable for highly persistent AR processes in logs with the AR(1) coefficient ≥ 0.9. This perhaps surprising finding might be due to difficulties with variance estimation for bias correction factors when the process approaches the unit root. For this reason we propose a hybrid approach where we first test for persistence by means of the augmented Dickey–Fuller (ADF) test for a unit root, and then choose the appropriate course of action: in the case of non-rejection of the unit root null hypothesis, one should apply no bias correction, whereas in case of rejection one should rely on e.g. the variance-based correction method. Our Monte Carlo simulation study finds the performance of this hybrid approach to be satisfactory.

This paper focuses on taking logs, which is the non-linear transformation used most frequently in practice. A more general Box–Cox transformation is of high relevance for selected applications; see e.g. Taylor (2017) and the literature cited therein. However, the cost of this more flexible case is that one cannot easily obtain corrections for the induced prediction bias. We address the bias correction for the Box–Cox procedure in Appendix B, where we show that it is not at all a trivial task, meaning that one should rather rely on the bootstrap-based corrections provided there.

In summary, the variance-based correction performs well even for cases where the normality assumption is violated, but the Linex-based method providing an unbiased forecast, or even the naïve no-correction approach, may outperform the variance-based method for specific cases. Our practical recommendation is to use the hybrid procedure described above. These findings are supported by the empirical results of using a log heterogeneous autoregressive (HAR) model (Corsi, Audrino, & Renò, 2012) for the purpose of predicting the daily realized volatility for highly liquid U.S. stocks, from both a statistical and an economic point of view.

The remainder of the paper is structured as follows. Section 2 discusses the necessity of bias correction, provides a summary of the established methods that are suitable for this purpose, and discusses the approach based on estimation under Linex loss. An extensive simulation study covering various types of linear processes and error distributions is presented in Section 3, whereas Section 4 contrasts the behaviors of the different bias corrections in an empirical application. Section 5 concludes the paper, while the Appendix collects some technical arguments.

2. Problem setting and bias correction techniques

2.1. The model

Let the strictly positive original (untransformed) process of interest be given as y_t, and assume that its log-transformation $x_t = \log(y_t)$ follows a stationary AR(p) process with *iid* innovations ϵ_t:

$$y_t = \exp(x_t), \qquad \rho(L)x_t = \mu + \epsilon_t, \qquad \epsilon_t \sim iid\,(0, \sigma^2), \quad (1)$$

where $\rho(L) = 1 - \sum_{j=1}^{p} \rho_j L^j$ is an invertible lag polynomial of order p. This setting could be generalized directly for ARMA(p, q) time series models, since linear processes may be approximated by means of AR(p) processes with a sufficiently large order p (see Berk, 1974; Bhansali, 1978), and even some forms of nonstationarity (Demetrescu & Hassler, 2016). We investigate both AR(1) and ARMA(1, 1) settings in the Monte Carlo simulation study in Section 3.

We are interested in one-step-ahead mean squared error (MSE) optimal forecasts of y_{T+1} given $y_T, y_{T-1}, \ldots$, and hence search for the conditional expectation of the examined series:

$$y_T(1) = \mathrm{E}[y_{T+1}|\mathcal{F}_T],$$
$$\mathcal{F}_T = \sigma\{y_T, y_{T-1}, \ldots, y_1\} = \sigma\{x_T, x_{T-1}, \ldots, x_1\}$$

with $\sigma\{a_1, a_2, \ldots\}$ denoting the σ-algebra generated by a_1, a_2, etc. Denote by $x_T(1) = \mathrm{E}[x_{T+1}|\mathcal{F}_T]$ the one-step-ahead MSE-optimal forecast of the log series x_t, so that $x_T(1) = \mu + (1 - \rho(L))x_{T+1}$.

Note that $E[y_{T+1}]$ must be finite in order for an MSE-optimal forecast to exist. We therefore require the distribution of ϵ_t to have thin tails. We take ϵ_t to have thin tails if $E[|\epsilon_t|^k] \leq Ca^k$ for some $C > 0$, $a > 1$ and all $k \in \mathbb{R}^+$, which implies, for example, that $E[\exp(\epsilon_t)] < \infty$. Moreover, given the stable finite-order autoregressive model assumed, an application of Minkowski's inequality shows that the moments of x_t satisfy the same conditions as those of ϵ_t, and hence, x_t has thin tails as well.

The representation of Eq. (1) in logs is often quite useful for modeling and estimation purposes. However, the interest in this paper is in producing the out-of-sample one-step-ahead MSE-optimal forecast $y_T(1) = E[\exp(x_{T+1})|\mathcal{F}_T]$ of the original variable y_{T+1}, which is then given as

$$y_T(1) = \exp\Big(E[x_{T+1}|\mathcal{F}_T]\Big)\ E\Big[\exp(\epsilon_{T+1})\Big]$$
$$= \exp\Big(x_T(1)\Big)\ E\Big[\exp(\epsilon_{T+1})\Big], \tag{2}$$

because of the representation

$$x_{T+1} = \mu + \sum_{j=1}^{p} \rho_j x_{T+1-j} + \epsilon_{T+1} = x_T(1) + \epsilon_{T+1} \quad \text{with} \quad E(\epsilon_{T+1}) = 0.$$

As in general $E[\exp(\epsilon_{T+1})] > 1$ due to Jensen's inequality, a naïve (uncorrected) forecast is

$$y_T(1) = \exp\Big(x_T(1)\Big). \tag{3}$$

Clearly, the naïve forecast in Eq. (3) is not MSE-optimal, and has a *downward* bias given by

$$E\Big[y_T(1) - y_{T+1}|\mathcal{F}_T\Big] = \exp\Big(x_T(1)\Big)\Big(1 - E[\exp(\epsilon_{T+1})]\Big).$$

The magnitude of the bias depends on the unknown distribution of the shocks ϵ_t, but also on the conditional expectation of x_t. In practice, one should estimate these forecast functions by plugging in consistent estimators $\hat{\mu}$ and $\hat{\rho}_j$, leading to $\hat{x}_T(1)$, so that the issue of the forecasting bias could be addressed subsequently. The sample $x_1, \ldots, x_T$ is used for estimating the model parameters in Eq. (1) and any of the bias correction terms.

To summarize, the MSE of the forecast $y_T(1)$ has three main sources: the bias, the volatility of shocks, and the model estimation error. Note that the estimation error could play a substantial role, since it is not negligible in smaller samples. Hence, the lack of efficiency in parameter estimates may become an issue when the sample is not large enough. For example, for Gaussian errors ϵ_t, a least squares (LS) estimation is maximum likelihood (ML), and thus is asymptotically efficient, while estimation under the Linex loss could be advantageous under a skewed error distribution, as it may be approximately proportional to the logarithm of the errors' density for a suitable skewness.

2.2. Variance-based bias corrections

When the model innovations ϵ_t are iid normally distributed, the optimal forecast $y_T(1)$ can be obtained as an explicit function of the error variance (Granger & Newbold, 1976):

$$y_T(1) = E[y_{T+1}|\mathcal{F}_T] = \exp\Big(x_T(1)\Big)\ E\Big[\exp(\epsilon_{T+1})\Big]$$
$$= \exp\Big(x_T(1)\Big) \cdot \exp\Big(\frac{1}{2}\sigma^2\Big).$$

Then, a feasible variance-corrected forecast is given by

$$\hat{y}_T(1) = \exp\Big(\hat{x}_T(1)\Big)\ \exp\Big(\frac{1}{2}\hat{\sigma}^2\Big), \tag{4}$$

where $\hat{\sigma}^2$ denotes a consistent estimator of the error variance σ^2 and $\hat{x}_T(1)$ is the estimated forecast from the log model in Eq. (1). For large values of T, where the estimation noise is negligible, this correction is exact for normally distributed model innovations. In practice, though, pronounced empirical deviations from normality are rather frequent. For this reason, we now examine bias corrections, which place fewer restrictions on the distribution of ϵ_t.

2.3. Mean-based bias correction

One could estimate the expectation $E\Big[\exp(\epsilon_{T+1})\Big]$ in Eq. (2) directly from the sample, e.g. as the sample average of transformed residuals,

$$\widehat{E}\Big[\exp(\epsilon_{T+1})\Big] = \frac{1}{T}\sum_{t=1}^{T} \exp(\hat{\epsilon}_t), \tag{5}$$

where $\hat{\epsilon}_t$ are the in-sample model residuals. Since we assumed thin-tailed innovations, the expectation is finite. The mean-corrected forecast would then be

$$\hat{y}_T(1) = \exp\Big(\hat{x}_T(1)\Big) \cdot \widehat{E}\Big[\exp(\epsilon_{T+1})\Big]. \tag{6}$$

Of course, more robust estimates of the central tendency (e.g. the median or truncated mean) could be applied here as well. However, our Monte Carlo simulations — not presented here but available on request — showed that using either the robust mean or robust variance estimator (concretely, 1% trimmed estimators to eliminate outliers) makes almost no difference in the results. Either way, we resort to residuals in order to compute an estimate of $E[\exp(\epsilon_{T+1})]$ in practice, so that the estimation error plays a role in this step as well.

2.4. Forecasts based on the Linex loss

The variance-based and mean-based bias corrections considered above are two-step procedures, where one estimates the AR model in Eq. (1) in the first step and computes the bias correction factor in the second step. We now consider a distribution-free approach which enables unbiased forecasts to be obtained in a single step for exponentially transformed values.

We obtain such a single step forecast by letting

$$m_{t+1} = \log E\left[y_{t+1}|\mathcal{F}_t\right],$$

such that $\exp(m_{t+1}) = E\left[y_{t+1}|\mathcal{F}_t\right] = E\left[\exp(x_{t+1})|\mathcal{F}_t\right]$, or, equivalently,

$$E\Big[e^{x_{t+1}-m_{t+1}} - 1\Big|\, x_t, x_{t-1}, \ldots\Big] = 0. \tag{7}$$

Note that this equality holds irrespective of the distribution of forecast errors.

Then, rather than predicting x_{T+1} and correcting the bias introduced by a non-linear transformation of $x_T(1)$, the idea is to estimate the conditional quantity m_{T+1} by imposing the moment condition in Eq. (7) directly. The latter is simply a transformed version of the required MSE-optimal forecast for y_t, as we have $e^{m_{t+1}} = \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right]$ for all t by definition.

Note that since the forecast m_{T+1} is not the conditional expectation of x_{T+1} given $\mathcal{F}_T$, it delivers a biased prediction of the log-transformed variables x_{T+1}. However, this bias in the log series forecasts is such that the exponent transformation to the original variable of interest y_{T+1} provides an unbiased forecast. This differs from conventional procedures that consist of first generating unbiased forecasts for log series, then correcting for bias in order to predict original variables.

The generalized method of moments (GMM) estimation is a natural choice for imposing the condition Eq. (7); for our case, a particular selection of instruments leads to the following estimator with a nice interpretation. Namely, we obtain from Eq. (1) that

$$m_{t+1} = \mu + \sum_{j=1}^{p} \rho_j x_{t+1-j} + \log\left(\mathrm{E}\left[\exp(\epsilon_{T+1})\right]\right)$$
$$:= \tilde{\mu} + \sum_{j=1}^{p} \rho_j x_{t+1-j},$$

and consider for $\theta = (\tilde{\mu}, \rho_1, \ldots, \rho_p)'$ the vector of moment conditions

$$\mathrm{E}\left[\left(e^{x_{t+1}-m_{t+1}} - 1\right)\frac{\partial m_{t+1}}{\partial \theta}\right] = 0.$$

For $\mathcal{L}(u) = e^u - u - 1$, these are the first-order conditions for the optimization problem

$$\min_{\theta} \mathrm{E}\left[\mathcal{L}\left(x_{t+1} - m_{t+1}(\theta)\right)\right],$$

where $\mathcal{L}(u)$ is recognized to be the linear-exponential (Linex) loss function introduced by Varian (1975), with parameters $a_1 = a_2 = 1$ in $\mathcal{L}_{a_1,a_2}(u) = e^{a_1 u} - a_2 u - 1$.[1]

Hence, we may estimate the model in logs under the Linex loss instead of using least squares by minimizing the average empirical loss

$$\hat{\theta} = \arg\min \sum_{t=p+1}^{T} \mathcal{L}\left(x_{t+1} - m_{t+1}(\theta)\right),$$

and computing

$$\hat{y}_T(1) = \exp\left(\hat{m}_{T+1}\right) = \exp\left(m_{T+1}(\hat{\theta})\right).$$

We establish the consistency of this forecasting procedure — in the sense that $\hat{y}_T(1)$ converges to the conditional expectation of y_{T+1} — by means of standard extremum estimator theory; the details are provided in Appendix. In particular, Appendix A shows the Linex-based estimators of $\rho_1, \ldots, \rho_p$ to be consistent; note that the intercept estimated via Linex is asymptotically biased, as it converges a.s. to $\tilde{\mu}$, where $\tilde{\mu} \neq \mu$ in general. As was argued above, it is precisely this feature that delivers the desired $\mathrm{E}\left[\exp\left(x_{T+1}\right)\right]$ in the limit, in particular because of the asymptotic bias of $\hat{\mu}$. This guarantees the MSE-optimality of the forecasts of y_t for large T values. In addition, Appendix B discusses the application of bias corrections for the non-linear Box–Cox procedure, which is a generalization of the log transformation.

2.5. Forecasts at higher horizons

Let us now briefly touch on the issue of h-step-ahead forecasts. We note that the decomposition analogous to Eq. (2) holds,

$$y_T(h) = \exp\left(\mathrm{E}[x_{T+h}|\mathcal{F}_T]\right)\,\mathrm{E}\left[\exp(\epsilon_{T+h}^{h})\right]$$
$$= \exp\left(x_T(1)\right)\,\mathrm{E}\left[\exp(\epsilon_{T+h}^{h})\right],$$

where $\mathrm{E}[x_{T+h}|\mathcal{F}_T]$ is the MSE-optimal h-step-ahead forecast of x_{T+h} and ϵ_{T+h}^{h} is the h-step-ahead forecast error. The optimal forecast $\mathrm{E}[x_{T+h}|\mathcal{F}_T]$ may be obtained via either direct or iterated forecasting, as usual; importantly, the "exponentiation bias" depends only on the properties of the forecast errors. Therefore, the variance- and mean-based corrections apply immediately, since only moments of the h-step-ahead forecast errors or their exponential transformation are required. The Linex-based correction is quite easy to compute for direct h-step-ahead forecasts, since the moment conditions for estimation under the Linex loss become $\mathrm{E}\left[\left(e^{x_{t+h}-m_{t+h}} - 1\right)\frac{\partial m_{t+1}}{\partial \theta}\right] = 0$, where $m_{t+h} = \log \mathrm{E}[y_{t+h}|\mathcal{F}_t]$, such that $x_{t+h} - m_{t+h}$ are the (biased) h-step-ahead Linex-based forecast errors.

Hence, there are not many methodological differences between making one-step-ahead forecasts, which is in our focus in what follows, and h-step-ahead forecasting. While we expect that the relative performances of the bias correction methods will be in favor of variance-based correction at longer horizons due to the central limit theorem-based arguments, some differences may occur. We leave such an investigation for further work.

3. Monte Carlo analysis

3.1. The baseline model

We now examine how the shape of the error distribution affects the MSE of the forecasts under consideration in finite samples. In order to contrast various forms of bias correction, we concentrate on both a simple AR(1) and a more sophisticated ARMA(1, 1) process for the log-transformed values, because of their immense practical importance. The stationary AR(1) model is given by

$$x_t = \mu + \rho x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid(0, 1/4), \qquad |\rho| < 1,$$

and we experiment with $\rho \in \{0.2, 0.5, 0.9\}$. We set the error variance equal to 1/4, to be in line with our empirical study; see Table 1.

[1] The existence of the expectation $E[\mathcal{L}_{a_1,a_2}(x_{t+1} - m)]$ is easy to establish given the thin tails of x_t and the linear-exponential shape of $\mathcal{L}_{a_1,a_2}$.

Table 1
Parameter estimates (st. errors) and descriptive statistics of residuals for the full sample log-HAR model in Eq. (9).

Company	β_0	$\beta^{(d)}$	$\beta^{(w)}$	$\beta^{(m)}$	$\beta^{(q)}$	R^2	$\widehat{\sigma}^2$	$\widehat{\xi}$	$\widehat{\kappa}$
S&P 500	−0.325 (0.114)	0.452 (0.017)	0.378 (0.028)	0.050 (0.037)	0.087 (0.029)	0.74	0.302	0.202	3.996
American Express	0.003 (0.010)	0.407 (0.017)	0.372 (0.029)	0.068 (0.041)	0.136 (0.031)	0.80	0.298	0.197	4.046
Microsoft	0.004 (0.010)	0.372 (0.017)	0.356 (0.031)	0.068 (0.045)	0.164 (0.037)	0.62	0.306	0.130	4.145
Exxon Mobil	0.001 (0.009)	0.437 (0.017)	0.400 (0.028)	0.013 (0.037)	0.107 (0.030)	0.68	0.246	0.237	4.229

In addition, we specify the stationary ARMA(1, 1) model as

$$x_t = \mu + \rho x_{t-1} + \phi \epsilon_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid(0, 1/4),$$

where we use the values $\rho = 0.8$ and $\phi = -0.5$. This choice of ARMA(1, 1) parameters leads to an autocorrelation function that is similar to that of the HAR model of Corsi (2009), which is investigated in the empirical study in Section 5. For both AR(1) and ARMA(1, 1), we set $\mu = 0$ without loss of generality, but estimate it from the data.

We are interested in forecasting $y_{T+1} = \exp(x_{T+1})$ given the information set $\mathcal{F}_T$. In the case of normally distributed innovations ϵ_t, the variance-based bias correction would be optimal. In the analysis that follows, we investigate different types of innovation distributions and compare the forecasting losses from the competing bias correction methods.

All simulations are performed in R (R-Core-Team, 2014). The estimation of the models in logs is conducted based on samples of sizes $T \in \{200, 500, 1000\}$ with 10^4 Monte Carlo replications. We also conducted some simulations with 10^5 repetitions, but the form and, apparently, non-smoothness of the resulting plots remained essentially the same. The most plausible reason for this behavior is a rather high variance of the observed MSE of forecasts of *exponentially* transformed series.

3.2. Distribution of innovations

We next consider four different types of deviations from normality, labeled Cases I–IV. Case I assumes that ϵ_t follows a standardized skew normal distribution (Azzalini, 1985), which is of vast importance in the current literature (Bondon, 2009; Sharafi & Nematollahi, 2016). A skew normal distributed (SND) random variable u_t is characterized by three parameters (ξ, ω, β) such that its mean and variance are given, with $\delta = \beta/\sqrt{1+\beta^2}$, by

$$\mathrm{E}[u_t] = \xi + \omega \cdot \delta\sqrt{2/\pi} \quad \text{and} \quad \mathrm{Var}[u_t] = \omega^2(1 - 2\delta^2/\pi).$$

We set $\xi = 0$ and $\omega = 1$ and compute the innovations calibrated to a zero mean and variance equal to 1/4 for various values of the skewness parameter β:

$$\epsilon_t = \frac{u_t - \mathrm{E}[u_t]}{2\sqrt{\mathrm{Var}[u_t]}}.$$

Case II assumes that ϵ_t follows a symmetric normal mixture distribution (NMD). This is an another popular deviation from normality (cf. McLachlan & Peel, 2004). NMD random variables u_t are given as

$$u_t \sim \begin{cases} \mathcal{N}(0, \sigma_1^2), & \text{if } B_t = 1, \text{ i.e. with probability } \pi, \\ \mathcal{N}(0, \sigma_2^2), & \text{if } B_t = 0, \text{ i.e. with probability } 1-\pi, \end{cases}$$

with an *iid* mixture variable B_t drawn from a Bernoulli distribution with the success probability $\pi \in (0, 1)$. Thus, the mixture distribution is characterized by three parameters $(\sigma_1^2, \sigma_2^2, \pi)$, with the variance $\mathrm{Var}[u_t] = \pi\sigma_1^2 + (1-\pi)\sigma_2^2$. We set the mixture probability $\pi = 1/2$, $\sigma_1^2 = 1$ and vary only the second variance σ_2^2. We model innovations as

$$\epsilon_t = \frac{u_t}{2\sqrt{\pi\sigma_1^2 + (1-\pi)\sigma_2^2}}.$$

The standard normal is included implicitly in Case I with $\beta = 0$, and Case II with $\sigma_2^2 = 1/4$.

Case III assumes that the innovations follow a contaminated normal distribution, which allows for higher kurtosis values (cf. Seidel, 2011). This case is rather similar to the Case II specification. The difference lies in the mixture probability, which is now set to $\pi = 0.95$, and the second variance is of larger magnitude. Again, we use innovations ϵ_t that are calibrated to a zero mean and variance of 1/4 for model estimation and evaluation.

Next, Case IV assumes that u_t follows a central t-distribution with $v \in [5, 30]$ degrees of freedom. The adjusted errors ϵ_t are obtained by $\epsilon_t = u_t \cdot 0.5\,(v/(v-2))^{-1/2}$. This choice is designed as a robustness check, since the t-distribution has fat tails, and therefore y_t would not have a finite expectation.

Finally, we investigate Cases I, II, and IV for the ARMA(1, 1) model with $\rho = 0.8$ and $\phi = -0.5$ for different sample sizes T; note that Case III is omitted because its results are very similar to those in Case II.

3.3. Methods for bias correction

The following methods are considered for making one-step-ahead forecasts of y_{T+1}.

1. Naïve forecast ignoring bias corrections, $\hat{y}_T(1) = \exp(\hat{x}_T(1))$.
2. Variance-based correction with $\hat{y}_T(1) = \exp(\hat{x}_T(1)) \exp\left(\frac{1}{2}\hat{\sigma}^2\right)$, where the variance σ^2 is estimated from sample residuals $\hat{\epsilon}_T, \ldots, \hat{\epsilon}_1$ as $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T} \hat{\epsilon}_t^2$.
3. Mean-based bias correction with $\hat{y}_T(1) = \exp(\hat{x}_T(1)) \cdot (1/T)\sum_{t=1}^{T} \exp(\hat{\epsilon}_t)$.

4. Linex-based forecast with $\hat{y}_T(1) = \exp(\hat{m}_{T+1})$.
5. Untransformed forecasts as per (Mayr & Ulbricht, 2015); i.e., fitting either AR(1) or ARMA(1, 1) models directly to the untransformed variables y_t.
6. An average forecast that builds on the mean-based and variance-based approaches, as well as on the hybrid strategy described below.

The plug-in estimates for the autoregressive parameters were obtained by means of OLS. As for the estimation of the Linex-based forecast function, we make use of a Newton–Raphson non-linear optimization algorithm (see e.g. Hamilton, 1994, p. 138) which has as starting values the OLS estimates obtained for the other three corrections.

For the AR(1) model, Figs. 1–3, corresponding to Cases I, II and IV respectively, plot the log MSE differences of the naïve, mean-based, average, Linex-based, and untransformed forecasts to the baseline variance-based forecast correction, which is optimal in the case of normal innovations. Moreover, Fig. 4 shows the log MSE differences for the ARMA(1, 1) model. As the results for Case III are very similar to those for Case II, we decide to skip the Case III plots due to space considerations. To improve the visualization, any methods which are dominated completely by the variance-based correction are excluded from the relevant plots.

3.4. Monte Carlo results

For all cases, we consider both small samples with $T = 200$ and large samples with $T = 1000$. The results for Case I with the skew-normal distribution of innovations are shown in Fig. 1, where we plot the log MSE differences depending on the value of the skewness parameter β. The evidence obtained is quite similar for all $T = 200$ and $T = 1000$, but varies with respect to the autocorrelation parameter value ρ.

For weak and medium autocorrelations $\rho = 0.2$ and $\rho = 0.5$, the Linex-based correction is the best for a pronounced negative skewness with parameter values $\beta < -2$, followed closely by mean-based and average-based corrections. For positive skewness, however, the variance-based correction appears to be mostly appropriate. This may be explained by the way in which the parameters are estimated: for negative skewness, the Linex loss function mimics the negative log-density of innovations and, hence, the estimation under Linex is more efficient than OLS. Both the naïve and untransformed forecasts are much worse than the other procedures.

For strong autocorrelation $\rho = 0.9$, the mean-, variance-, and average-based forecasts are close to each other. The Linex is slightly better for negative skewness values and slightly worse for positive. For $T = 200$, the naïve forecast is to some extent worse than the alternatives; however, it appears to be the worst for the large sample size $T = 1000$. Again, this is most likely to be due to the nature of the estimation error.

The log differences of MSEs in Case II with the normal mixture are presented in Fig. 2 for different values of σ_2^2. For AR(1) parameters $\rho = 0.2$ and $\rho = 0.5$, the variance correction method appears to be the best one for $T = 200$, whereas for $T = 1000$ the mean- and average-based corrections are quite close to it. The Linex-based forecast is dominated for $T = 200$ but approaches the variance-based alternative for $T = 1000$. Again, the naïve uncorrected forecast appears to be reasonable for strong autocorrelation $\rho = 0.9$. The results for Case III with the contaminated distribution are quantitatively similar to those for Case II, so we do not show them here.

Next, in Case IV with t-distributed innovations, shown in Fig. 3, we observe that the variance-based and mean-based forecasts are the best for $\rho = 0.2$ and $\rho = 0.5$. As the Linex-based MSE appears to be unstable numerically, we do not recommend it for correction in the case of t-distributed innovations and do not report Linex-based results for Case IV. As earlier, the advantages of the naïve forecast decrease as the sample size T increases, but the naïve uncorrected forecast should be used for $\rho = 0.9$ and $T = 200$.

The log MSE ratios for the ARMA(1, 1) model specification are shown in Fig. 4. The graphs in the top line correspond to Case I, those in the medium line to Case II, and those in the bottom line to Case IV. For Case I, the Linex-based correction should be used for negative skewness and avoided for positive skewness values. In Cases II and IV, the Linex is dominated by variance-based corrections. The naïve and untransformed forecasts are worse than the variance-based forecasts for all settings, whereas the mean-based and average corrections are mostly close to the variance-based benchmark. Thus, the results for AR(1) and ARMA(1, 1) models are quite similar.

Finally, for Case IV with t-distributed innovations for the small estimation window $T = 200$, we investigate the performances of naïve forecasts with respect to the AR(1) coefficient ρ. The corresponding plots for $\rho \in \{0.75, 0.8, 0.85, 0.9, 0.95, 0.98\}$ are shown in Fig. 5. As can be observed, the naïve uncorrected forecast is very close to the variance-based correction for $\rho = 0.9$, and gets much better for $\rho = 0.95$ and $\rho = 0.98$. Hence, we conclude that an increased persistence in the time series behavior favors the possible usage of naïve predictors.

Hence, our major findings for weak and medium autocorrelation coefficients $\rho \in \{0.2, 0.5\}$ are as follows. In Case I, negative skewness β is in favor of the Linex-based method, whereas for positive β values this method becomes unstable and variance-based correction is preferable. In Cases II and III, variance-based correction is slightly better than mean-based correction in the case of the normal mixture distribution. In Case IV, variance correction is suitable for t-distributed innovations. Higher values of the autoregressive parameter $|\rho|$ lead to more instability for all bias correction methods studied, so that no bias correction appears to be preferable for $\rho \geq 0.9$ and $T = 200$, for example.

Thus, the evidence that we have found suggests that the degree of autoregressive persistence is of great importance for the practical decision of whether or not to perform bias correction. For this reason, we propose a hybrid procedure that can take the degree of persistence into account. Concretely, we first conduct the augmented

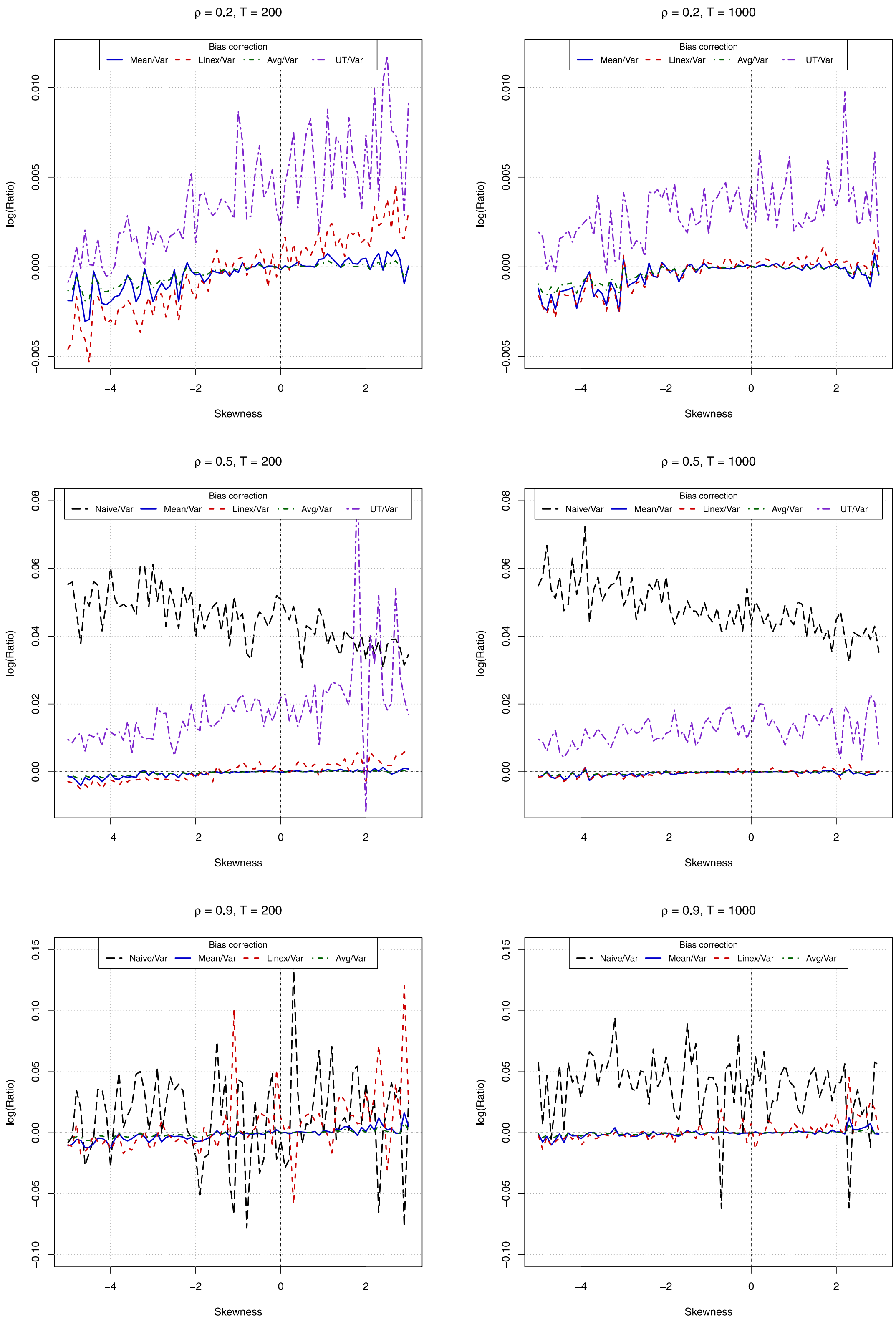

Fig. 1. Log MSE ratios for $\rho \in \{0.2, 0.5, 0.9\}$ from top to bottom, $T \in \{200, 1000\}$ from left to right, with the skewness parameter $\beta \in [-5,\ 3]$ and with $\beta = 0$ for a symmetric distribution.

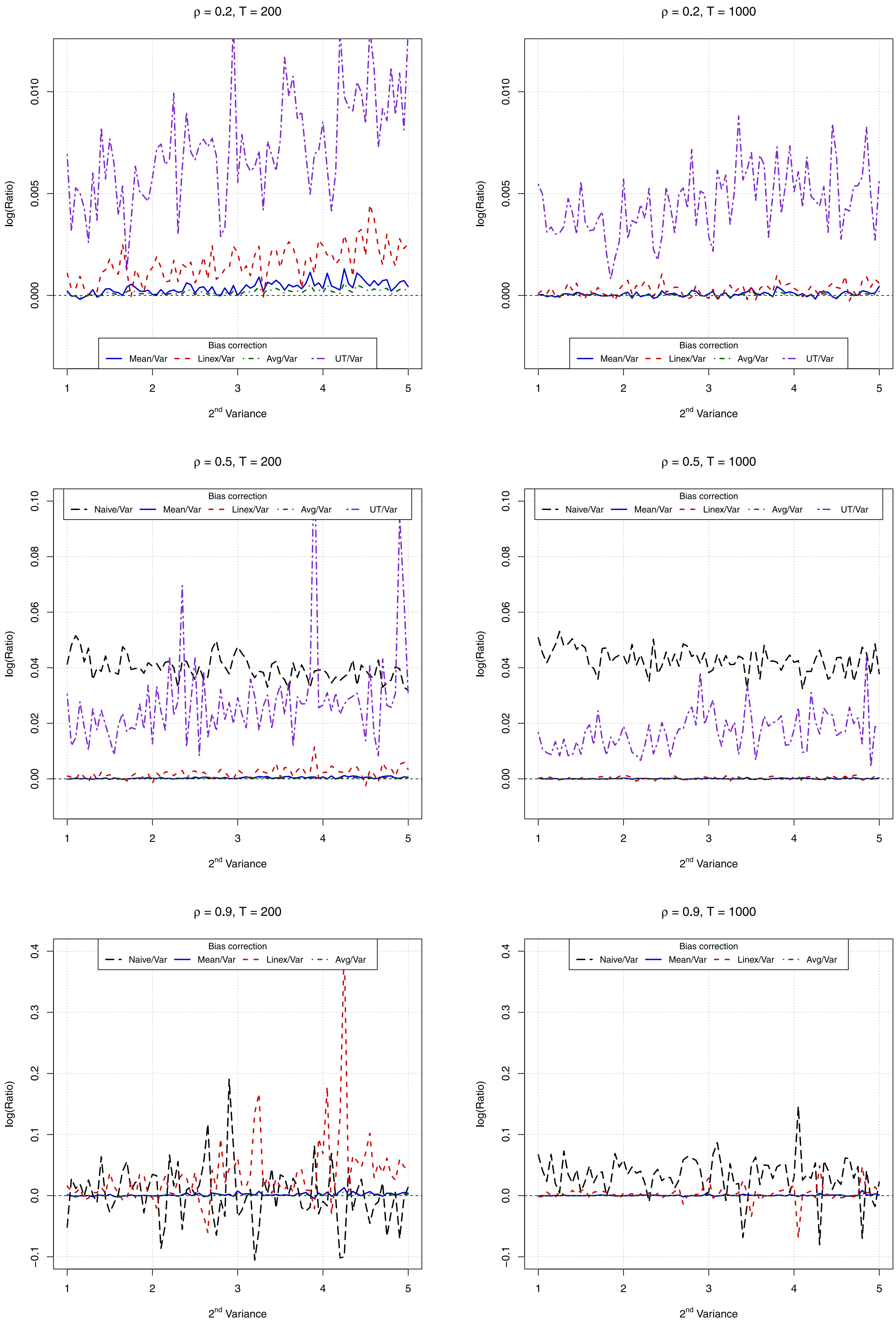

Fig. 2. Log MSE ratios for $\rho \in \{0.2, 0.5, 0.9\}$ from top to bottom, $T \in \{200, 1000\}$ from left to right; mixture of $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, \sigma_2^2)$ with probability $\pi = 0.5$ and $\sigma_2^2 \in [1, 5]$.

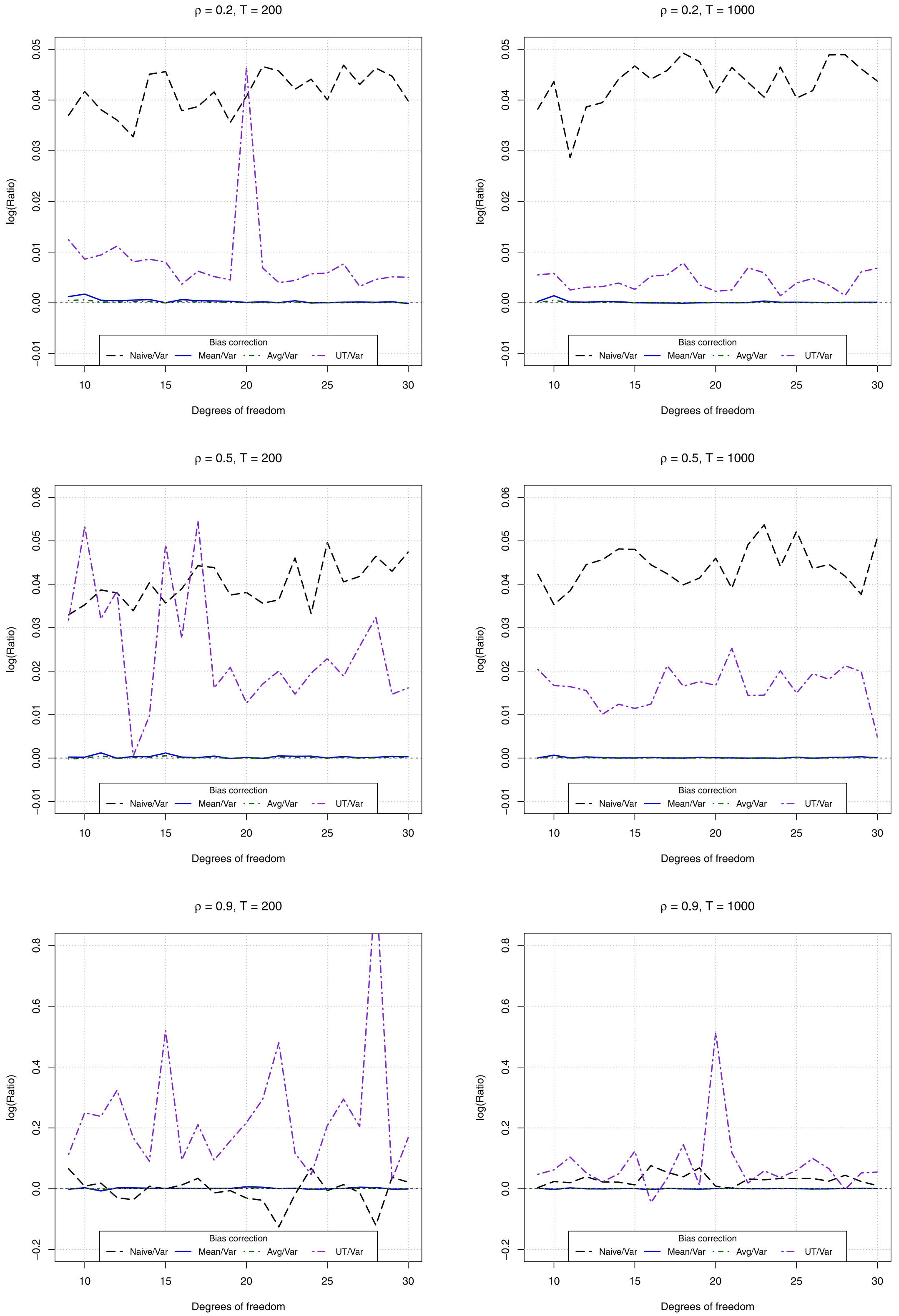

Fig. 3. Log MSE ratios for $\rho \in \{0.2, 0.5, 0.9\}$ from top to bottom, $T \in \{200, 1000\}$ from left to right; t-distribution with $\upsilon \in [5,\ 30]$ degrees of freedom.

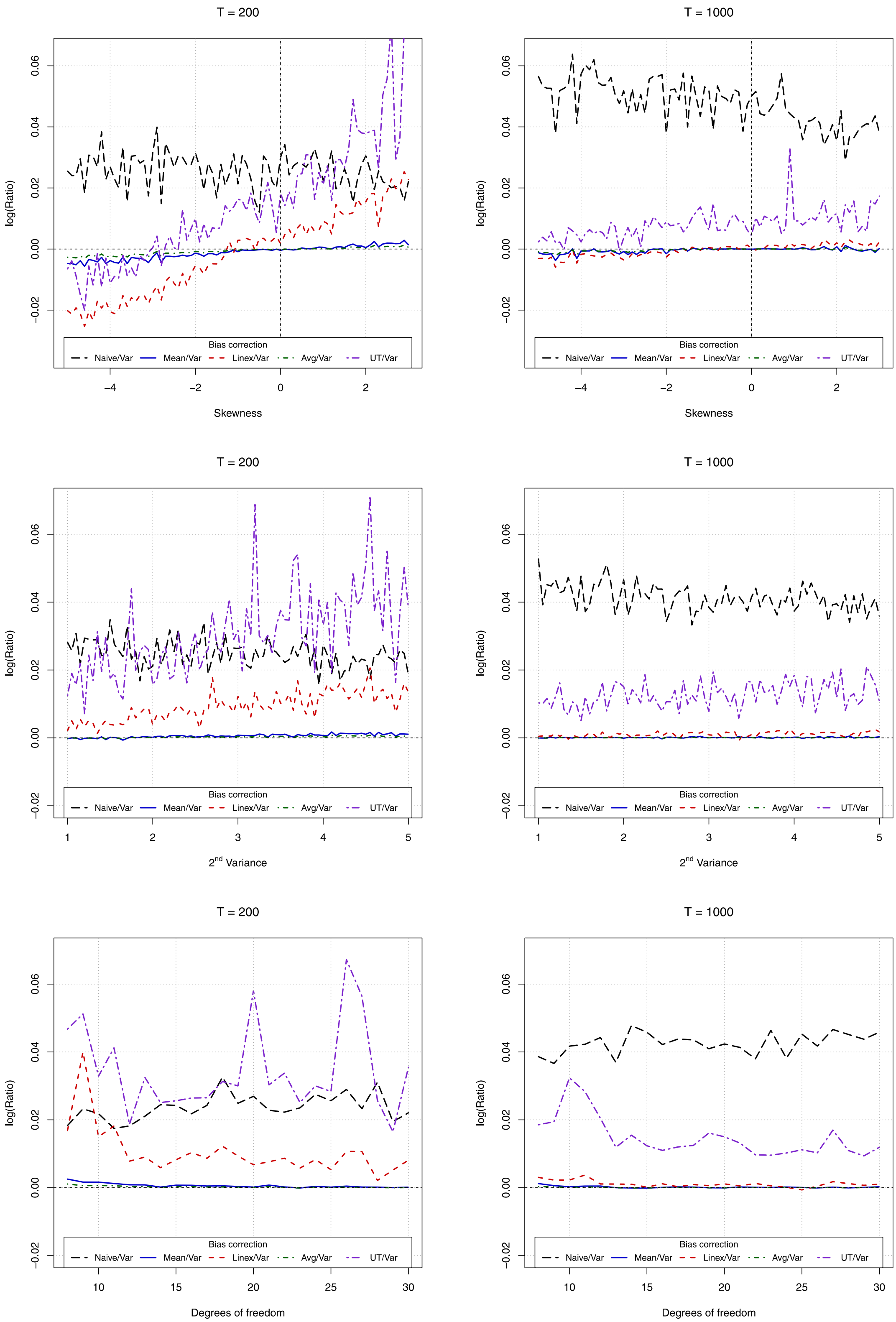

Fig. 4. Log MSE ratios for ARMA(1, 1) with skewed normal, normal mixture and Student-t innovations from top to bottom, $T \in \{200, 1000\}$ from left to right.

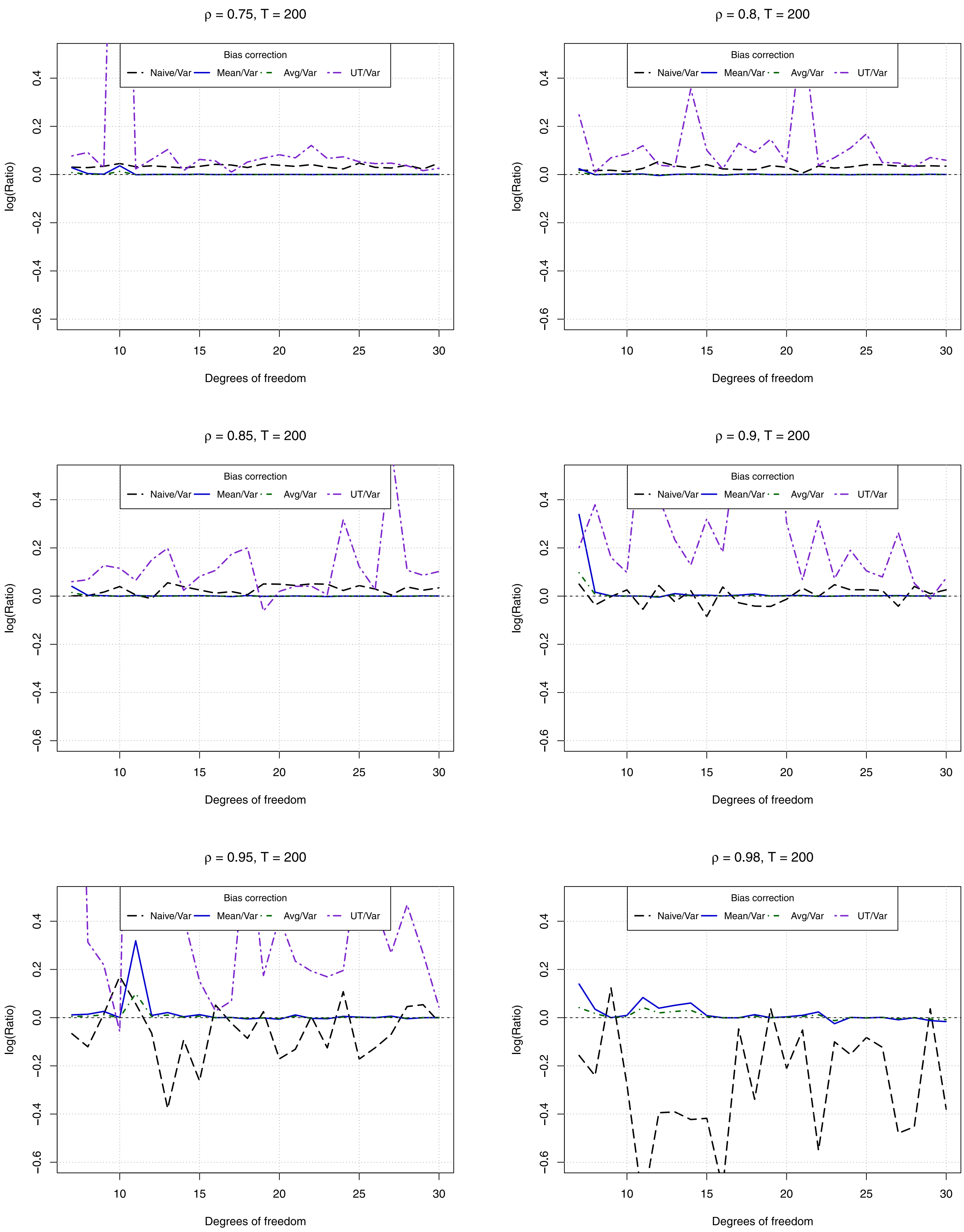

Fig. 5. Log MSE ratios for AR(1) with Student-t innovations and $T = 200$ and increasing persistence.

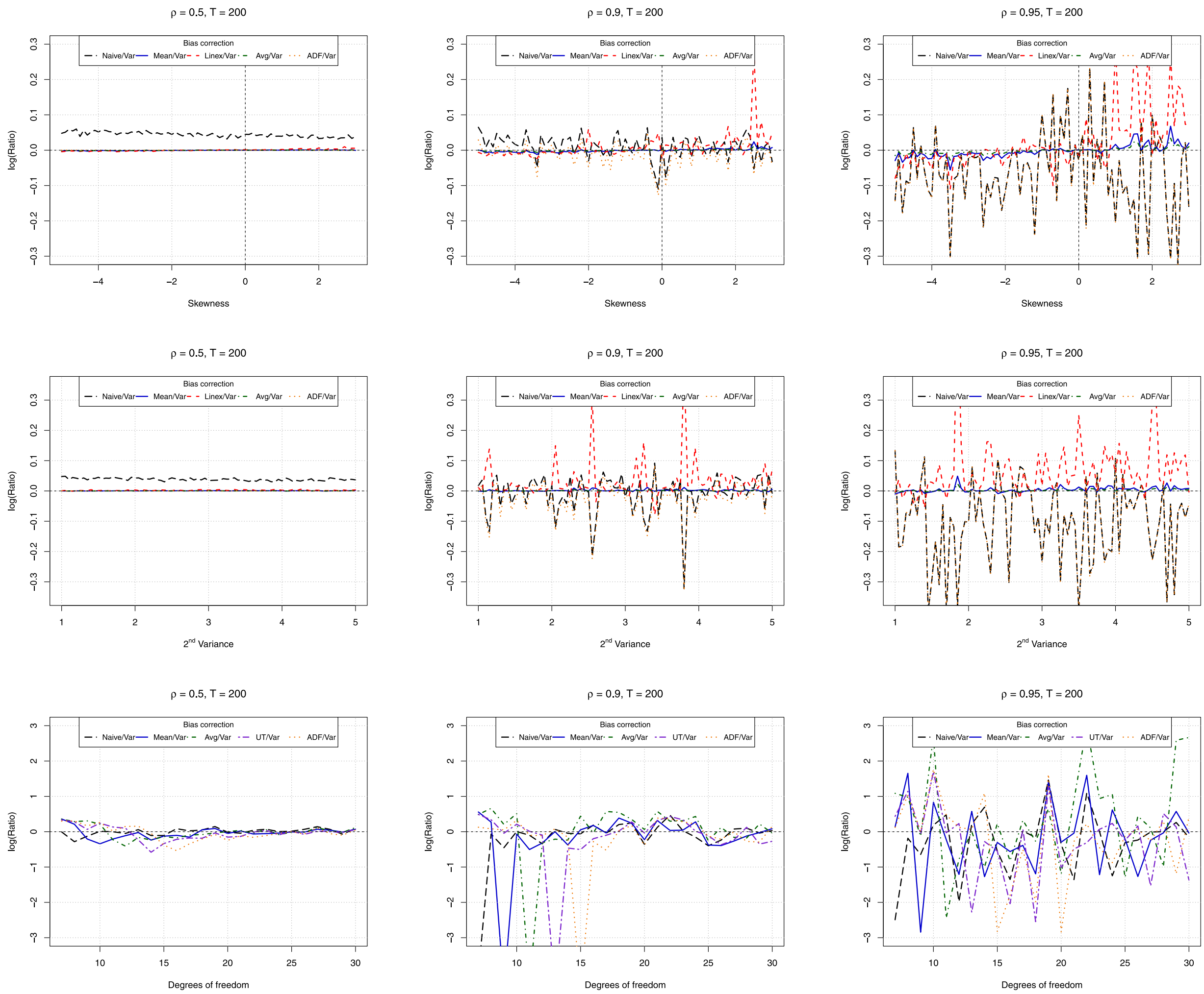

Fig. 6. The effect of the hybrid ADF-test-based procedure: Log MSE ratios for AR(1) with skewed normal, normal mixture and Student-*t* innovations (from top to bottom) and $T = 200$.

Dickey–Fuller (ADF) test for the null hypothesis of a unit root. In the case of rejection, we use one of the bias-correction methods, but in the case of non-rejection, we do not perform any bias correction.

Fig. 6 presents the behavior of the hybrid procedure based on the ADF test at the 5% level, with an intercept but no trend, for the DGPs with increasing levels of persistence, $\rho \in \{0.5, 0.9, 0.95\}$. For $\rho = 0.5$, the hypothesis of a unit root is always rejected, and hence we do not observe any deviations from the variance-based correction method. For $\rho = 0.9$, the null hypothesis is rejected in some cases, which results in the choice of the naïve approach. Finally, the unit root hypothesis is never rejected for $\rho = 0.95$, which makes the log ratio coincide completely with that of the naïve approach. These findings hold for the simulations in Cases I and II. However, the impact of the ADF pre-testing strategy appears to be ambiguous for the heavy-tailed Student-*t* distribution. Thus, we can state that choosing a bias-correction method based on a pre-test can improve the forecasting performance for highly persistent processes.

4. Empirical illustration

The availability of intraday data allows us to estimate the true daily volatility σ_t^2 consistently using its realized measure y_t (cf. Andersen, Bollerslev, & Diebold, 2007), which serves as the time series of interest y_t in our study. It is recognized widely that volatility timing has a pronounced economic effect (cf. Fleming, Kirby, & Ostdiek, 2001, 2003), meaning that a proper volatility forecast is of considerable relevance for investment decisions. We focus on the autoregressive model for realized volatility in logs in order to make forecasts of y_{t+1} conditional on the information set $\mathcal{F}_t$. For this purpose, we contrast naïve uncorrected forecasts with those from the variance-based, mean-based, average of variance- and mean-based, and Linex-based methods for the bias correction with the aim of one-step-ahead prediction of the daily realized volatility.

4.1. HAR model for daily realized volatility

The heterogeneous autoregressive (HAR) model of Corsi (2009) is quite successful for modeling and forecasting the daily realized volatility. We assess the complex autoregressive structure of the process y_t by exploiting the HAR model that includes daily, weekly, monthly, and quarterly components (cf. Andersen et al., 2011):

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_t^{(w)} + \alpha_3 y_t^{(m)} + \alpha_4 y_t^{(q)} + \varepsilon_{t+1}, \qquad (8)$$

with $y_t^{(w)} = (1/5) \cdot \sum_{i=0}^{4} y_{t-i}$, $y_t^{(m)} = (1/22) \cdot \sum_{i=0}^{21} y_{t-i}$, and $y_t^{(q)} = (1/65) \cdot \sum_{i=0}^{64} y_{t-i}$. Here, the lag orders 5, 22, and 65 are the average numbers of weekly, monthly, and quarterly trading days, respectively. The HAR model (Eq. (8)) for daily volatility prediction based directly on the non-transformed realized volatility measures stands for the untransformed approach.

A considerable disadvantage of the specification in Eq. (8) is that the symmetry assumption for the distribution of ε_t is obviously violated due to pronounced impact discrepancies of positive and negative volatility shocks (cf. Tsay, 2010). For this reason, a log transformation $x_t = \log y_t$ is applied commonly for the purpose of modeling (Andersen et al., 2007), since it levels out the asymmetries in the innovations. Then, the corresponding HAR model in logs (cf. Corsi et al., 2012; Golosnoy, Hamid, & Okhrin, 2014) is:

$$x_{t+1} = \beta_0 + \beta_1 x_t + \beta_2 x_t^{(w)} + \beta_3 x_t^{(m)} + \beta_4 x_t^{(q)} + \epsilon_{t+1}, \qquad (9)$$

where $x_t^{(\cdot)}$ is defined analogously to $y_t^{(\cdot)}$.

As it holds that $\exp[\mathrm{E}(x_{t+1}|\mathcal{F}_t)] \neq \mathrm{E}[\exp(x_{t+1})|\mathcal{F}_t]$, one requires a bias correction for the volatility forecasts. Hence, we estimate the model in Eq. (9) and make a forecast of y_{t+1} given the information set $\mathcal{F}_t$ by applying various types of bias corrections. Note that the non-normality of the innovations ϵ_{t+1} in log volatility processes is well-documented (cf. Lanne, 2006).

4.2. Data and descriptive statistics

We investigate daily realized volatilities for the S&P500 index and three highly liquid US stocks, namely American Express, Exxon Mobil, and Microsoft, which represent different sectors of the US economy. The realized volatility series of S&P500 is obtained from the Oxford-Man Library, whereas the daily volatility series for individual stocks are computed from 1 min intraday returns taken from QuantQuote.com as realized kernel measures with the Parzen kernel (Barndorff-Nielsen, Hansen, Lunde, & Shephard, 2011).

Our time series cover the period ranging from December 31, 2001, until December 31, 2014, resulting in 3255 daily realized variances for each asset. The time series considered are depicted in Fig. 7 such that both calm and turmoil periods in U.S. financial markets are observed during the period under investigation.

We investigate the properties of residuals from the log-HAR model in Eq. (9) by first estimating this model using the OLS based on the full-sample information. The parameter estimates are given in Table 1, where we also provide the estimates of the residual variance, skewness, and kurtosis. All of the regressor coefficients are significantly larger than zero, which supports the selected HAR specification. The estimated models for all series considered show no unit root behavior, as $\sum_{i=1}^{4} \hat{\beta}_i \approx 0.95$ in a reasonably large sample. The R^2 measures for all assets are quite high, between 0.6 and 0.8. For all four series, the residuals appear to be right-skewed and exhibit a kurtosis of around four, i.e., an excess kurtosis around one. The autocorrelation functions (ACFs) for the original data and the residuals from the untransformed HAR models are shown in Fig. 8, whereas the ACFs for the log-transformed data and the residuals from the log HAR model are shown in Fig. 9.

The hyperbolic decay of the ACF is observed for both original (untransformed) data and the log-transformed series. The remaining residual autocorrelation is more pronounced for the untransformed HAR in Fig. 8 than for the log HAR in Fig. 9. This evidence could be seen as supporting the modeling of log realized measures by the HAR specifications. To summarize, these HAR models appear to provide reasonable time series specifications for the log realized volatilities.

4.3. Comparison of bias correction methods

We make one-step-ahead volatility predictions by re-estimating the log-HAR model in Eq. (9) based on moving windows of sizes $T \in \{200, 500, 750, 1000\}$ days. We set variance-based correction as a benchmark and compare it to the naïve, mean-based and Linex-based methods. In addition, we also consider the average of the mean- and variance-based forecasts, the hybrid approach, and the untransformed forecasts from the model without logs in Eq. (8). The corresponding logs of MSE ratios for all assets (i.e., the MSE increase compared to the variance-based corrections, in %) are presented in Table 2.

The major findings for the MSE can be summarized as follows. The untransformed forecast is the worst one for all constellations. For $T = 200$, the naïve approach leads to the smallest MSE for individual stocks, whereas the Linex correction is the best for the S&P500 index. However, both the naïve and Linex approaches are worse than the variance-based correction for larger values of the estimation window T. The mean-based approach is slightly worse than either the variance-based correction or the average of the mean-based and variance-based approaches. The hybrid approach appears to be mostly relevant for $T = 200$, which is in line with our Monte Carlo simulation results.

Note that although the numerical differences in the MSE in Table 2 are not very large, looking for the best point volatility forecast is still of considerable economic relevance. For example, since volatility prediction is important for pricing derivative financial instruments, such as European and American options (cf. Tsay, 2010), even a small improvement in daily volatility forecasts could lead to substantial economic gains or losses. However, trying

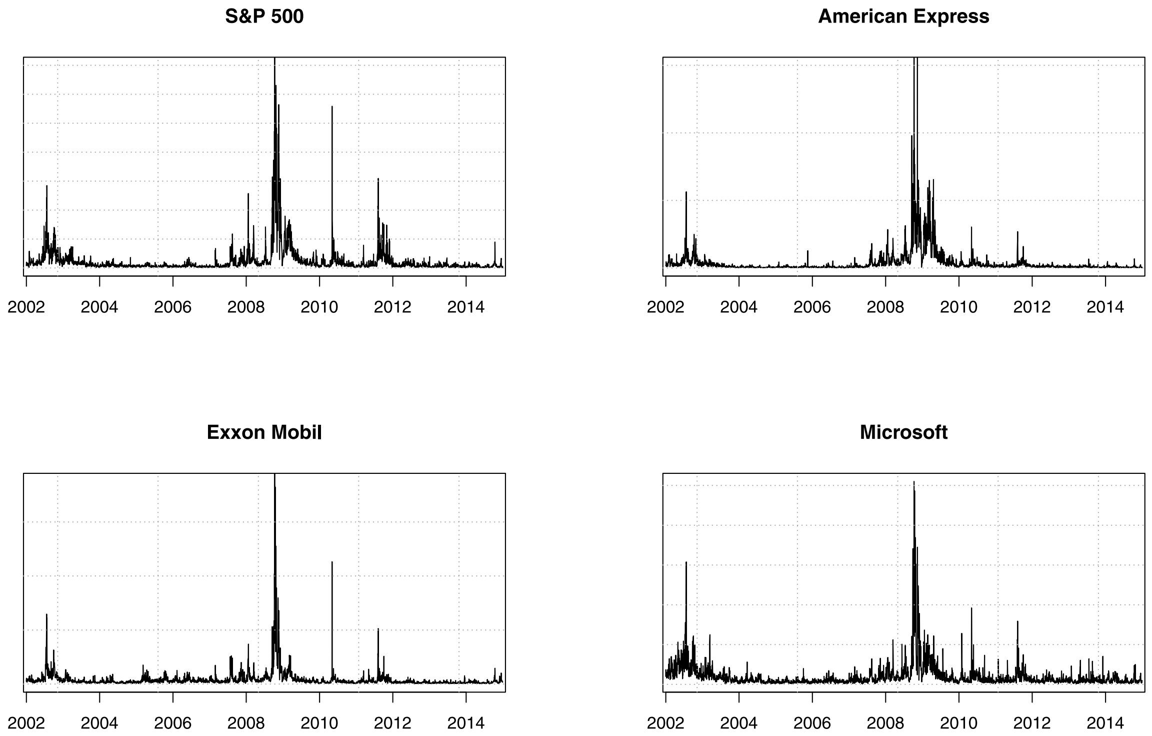

Fig. 7. Empirical time series of daily volatility measures based on realized kernel estimators.

Table 2
Log of MSE ratios (MSE increase in % compared to variance-based corrections) for one-day-ahead volatility forecasts from the log-HAR model in Eq. (9), estimated based on moving windows of size T.

T	naïve/var	mean/var	Linex/var	Avg/var	UT/var	ADF/var
S&P 500						
200	−0.025	0.004	−0.033	0.002	0.441	−0.031
500	0.008	0.003	0.009	0.001	0.457	0.000
750	0.017	0.002	0.021	0.001	0.467	0.000
1000	0.025	0.000	0.024	0.000	0.463	0.000
American Express						
200	−0.034	0.002	−0.032	0.001	0.303	−0.066
500	0.001	0.002	0.092	0.001	0.318	0.000
750	0.001	0.003	0.000	0.001	0.318	0.000
1000	0.011	0.003	0.010	0.002	0.323	0.000
Exxon Mobil						
200	−0.038	0.004	−0.038	0.002	0.599	−0.071
500	0.017	0.001	0.017	0.000	0.601	0.000
750	0.029	0.000	0.029	0.000	0.579	0.000
1000	0.037	−0.001	0.036	−0.001	0.553	0.000
Microsoft						
200	−0.127	0.031	−0.100	0.015	−0.035	−0.045
500	−0.041	0.028	−0.043	0.013	0.012	0.000
750	−0.063	0.041	−0.026	0.019	−0.035	0.000
1000	−0.057	0.035	0.025	0.017	−0.037	0.000

Note: The lag order for the ADF pretest was chosen by AIC.

to use our approach to construct a profitable trading strategy clearly remains beyond the scope of our paper.

We assess the statistical significance of our results by considering the popular Diebold–Mariano test for equal predictive accuracy in order to compare the competing correction approaches. We compare the approaches under consideration pairwise, and report in Table 3 the rejections of the benchmark at the 5% significance level in columns by '+' and the non-rejections by '−'.

The forecasts from the untransformed model (UT) are rejected statistically in all settings. For the small estimation window $T = 200$, the Linex is significantly the best approach for the index S&P500, whereas the naïve forecast is the best for the three individual stocks. Statistically, the variance-based correction appears to be the best approach for almost all settings for $T \in \{500, 750, 1000\}$, with the exception of three cases where the Linex and mean-based forecasts are better. Summarizing our evidence, one should rely on the proposed hybrid approach for small estimation windows, while the variance-based correction is mostly appropriate for larger windows.

4.4. Evaluation of the economic effect

Let us now evaluate the economic effect of bias corrections on realized volatility forecasts via a utility-based approach. For this purpose, we consider a risk-averse investor who, at time $t-1$, allocates part of his wealth $\omega_t \in [0, 1]$ in a broadly diversified portfolio, represented by the S&P 500 index, and holds the remaining part of his wealth, $1-\omega_t$, in cash. For this purpose, we exploit the S&P500 daily returns and realized volatility series from January 3rd, 2000, to December 4th, 2017.

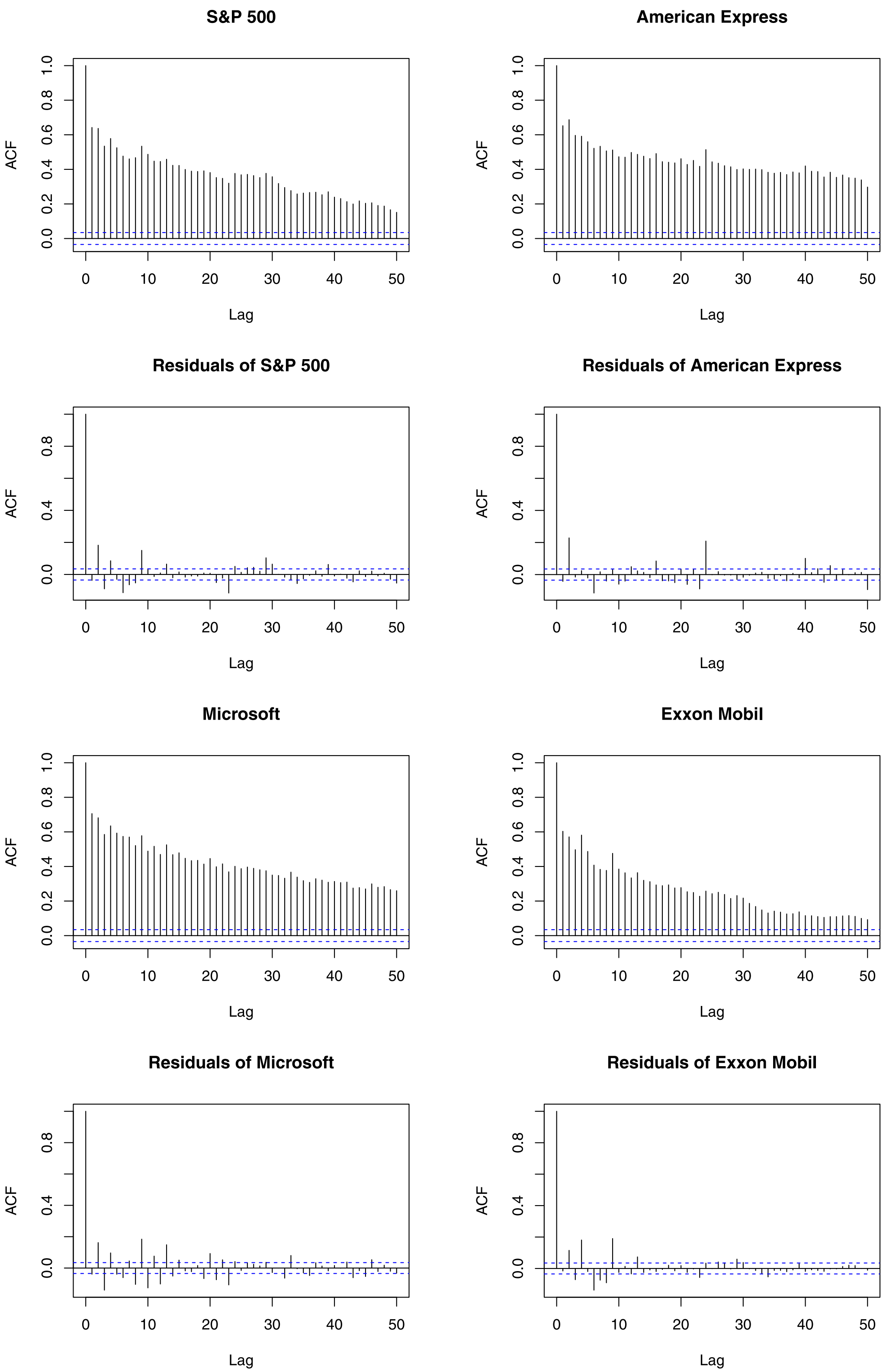

Fig. 8. ACF of the data and corresponding HAR residuals.

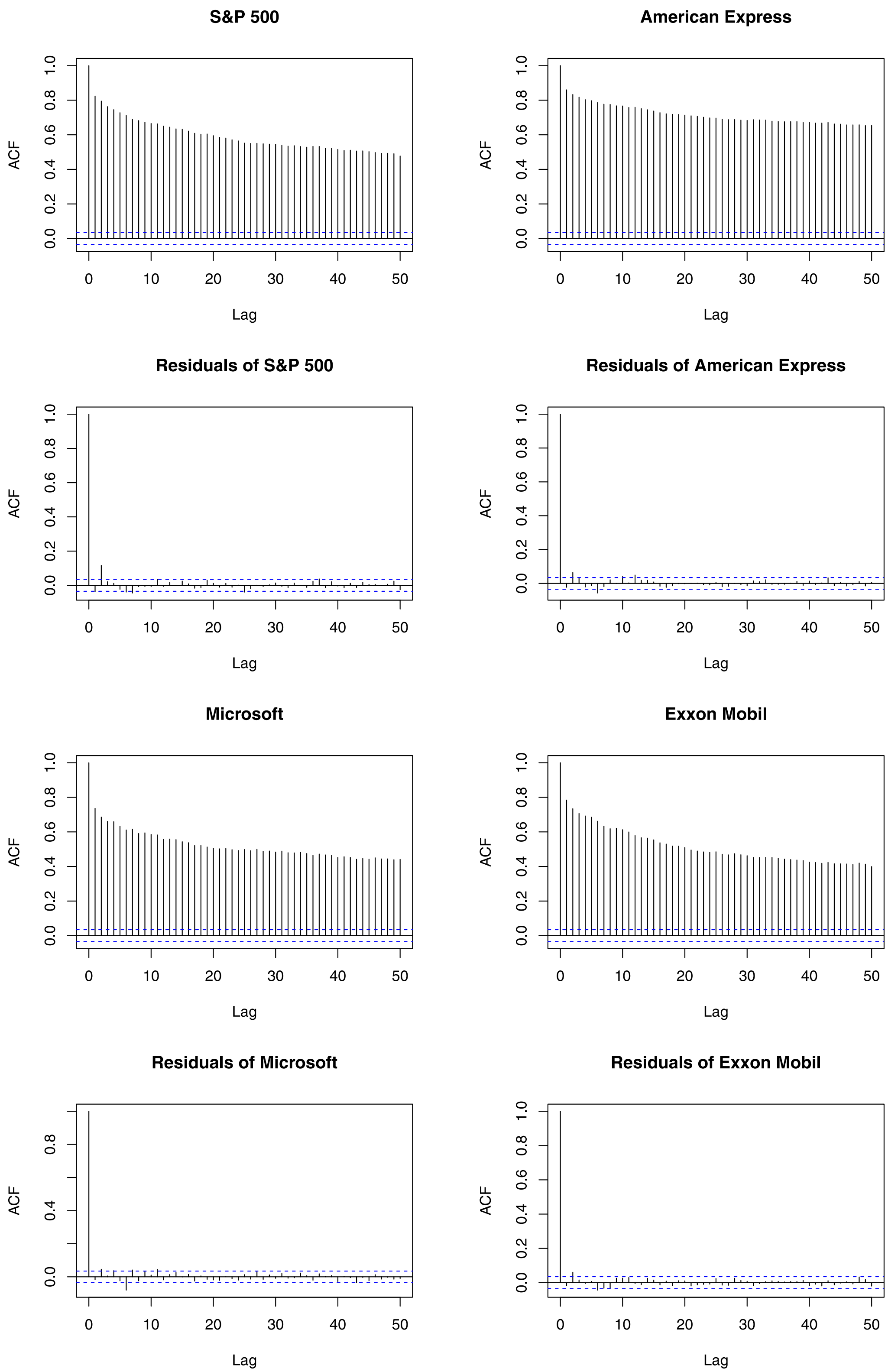

Fig. 9. ACF of the data in logs and corresponding HAR residuals.

Table 3
Diebold–Mariano test results for daily realized volatilities of S&P 500, American Express, Exxon, and Microsoft.

benchmark ⟶	Naïve				Variance corrected				Mean corrected				Linex				Average				Untransformed			
window size	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000
competitor ↓	S&P 500																							
Naïve					−	+	+	+	−	+	+	+	+	−	−	+	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	−	−	+	+	+	+					+	−	−	−	+	+	+	+	−	−	−	−
Linex	−	+	+	−	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	American Express																							
Naïve					−	+	+	+	−	−	−	+	−	−	+	+	−	−	−	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	+	−	+	+	+	+					+	−	+	−	+	+	+	+	−	−	−	−
Linex	+	+	−	−	−	+	−	+	−	+	−	+					−	+	−	−	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	+	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	Exxon Mobil																							
Naïve					−	+	+	+	−	+	+	+	−	+	−	+	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	+	+	+	−	−	−	−	−	+	+	−	−	−	−
Mean	+	−	−	−	+	+	−	−					+	−	−	−	+	+	−	−	−	−	−	−
Linex	+	−	+	−	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	−	−	−	−	+	+	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	Microsoft																							
Naïve					−	+	+	+	−	+	+	+	−	−	+	−	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	−	−	+	+	+	+					+	−	−	−	+	+	+	+	−	−	−	−
Linex	+	+	−	+	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				

Notes: The columns represent benchmark models with rolling window sizes, $t \in \{200, 500, 750, 1000\}$; the rows represent competitors. Under H_0, the benchmark and competitor have equal predictive accuracies; under the alternative, the benchmark is more accurate. Rejection at the 5% level in favor of the benchmark is shown by '+', non-rejection by '−'.

For a risk aversion coefficient $\gamma > 0$, the popular mean–variance utility function is given as

$$EU(R_t) = \mathrm{E}[\omega_t R_t|\mathcal{F}_{t-1}] - \frac{\gamma}{2}\,\mathrm{Var}[\omega_t R_t|\mathcal{F}_{t-1}],$$

where ω_t is the proportion of wealth invested in the S&P 500 at $t-1$, and R_t is the daily log-return of the S&P 500. The optimal allocation is then given as

$$\omega_t = \frac{1}{\gamma}\frac{\mathrm{E}[R_t|\mathcal{F}_{t-1}]}{\mathrm{Var}[R_t|\mathcal{F}_{t-1}]},$$

where $\mathrm{E}[R_t|\mathcal{F}_{t-1}]$ and $\mathrm{Var}[R_t|\mathcal{F}_{t-1}]$ are replaced by their forecasts conditional on the available information.

As the expectation $\mathrm{E}[R_t|\mathcal{F}_{t-1}]$ is hardly predictable in the short run, we replace it with the simple average $\hat{\mu}_t = (1/T)\sum_{i=1}^{T} R_{t-i}$ of the previous T daily asset returns. The variance $\mathrm{Var}[R_t|\mathcal{F}_{t-1}]$ is replaced by the realized volatility forecast $\widehat{y}_t$ based on the information set $\mathcal{F}_{t-1}$, so that the predicted weight of the S&P 500 index is $\hat{\omega}_t^{(\bullet)} = \hat{\mu}_t/(\gamma\,\widehat{y}_t^{(\bullet)})$, with the forecasts exhibiting several bias corrections. Moreover, we make the investment approach feasible by setting

$$\hat{\mathfrak{w}}_t^{(\bullet)} = \begin{cases} 0 & \text{if } \hat{\omega}_t^{(\bullet)} < 0 \\ \hat{\omega}_t & \text{if } \hat{\omega}_t^{(\bullet)} \in [0, 1] \\ 1 & \text{if } \hat{\omega}_t^{(\bullet)} > 1, \end{cases}$$

in order to prohibit the short-selling operations. The resulting returns $R_t^{(\bullet)} = \hat{\mathfrak{w}}_t^{(\bullet)} R_t$ are then used for computing economic performance measures.

We report several popular performance measures that are used frequently for economic evaluation purposes (cf. Brandt, Santa-Clara, & Valkanov, 2009). First, we provide the mean $\overline{R}^{(\bullet)}$ and standard deviation $\mathcal{S}^{(\bullet)}$ of the returns $R_t^{(\bullet)}$ for all volatility forecasting approaches. Moreover, we compute the Sharpe ratio as $SR^{(\bullet)} = \overline{R}^{(\bullet)}/\mathcal{S}^{(\bullet)}$, which is rather important for practitioners. Finally, we consider the mean–variance objective function, given as

$$EU^{(\bullet)} = \overline{R}^{(\bullet)} - \frac{\gamma}{2}\left(\mathcal{S}^{(\bullet)}\right)^2,$$

for two commonly-used values of the risk aversion, $\gamma = 5$ and $\gamma = 10$. The results obtained are presented in Table 4.

For risk aversion $\gamma = 5$, we find that the naïve (uncorrected) approach provides the best $SR^{(\bullet)}$ and $EU^{(\bullet)}$ values for $T = 500$, followed closely by the Linex approach. However, the picture is different for $T = 750$ and $T = 1000$, where the Linex correction provides the best results. It should be mentioned that the performance of the variance-based correction improves as T increases. For $\gamma = 10$, where a correct volatility prediction is more important, the variance- and mean-based corrections outperform the naïve approach for all values of T. However, the untransformed forecast (UT) appears to be the best approach for $T = 1000$ for both the $SR^{(\bullet)}$ and $EU^{(\bullet)}$ measures. This is due to a surprisingly high average return $\overline{R}^{(ut)}$, because the standard deviation $\mathcal{S}^{(ut)}$ is the highest one. The hybrid strategy coincides with the var-based correction because the null 'unit root' is rejected in all cases here. In summary, the Linex-based and variance-based corrections are most suitable for risk-averse investors in our setting for $\gamma = 5$ and $\gamma = 10$, respectively.

5. Summary

Forecasting with an autoregressive model for log-transformed variables is a convenient option in numerous applications. However, a reverse transformation in order

Table 4
Economic measures for $\gamma = 5$ and $\gamma = 10$ and different estimation windows T.

Risk aversion	Window T	Method (•)	Mean $\overline{R}^{(\bullet)}$	St. deviation $\mathcal{S}^{(\bullet)}$	Sharp ratio $SR^{(\bullet)}$	Expected utility $EU^{(\bullet)}$
		Naive	0.0095	0.0386	0.2462	0.0058
		Var	0.0010	0.0378	0.0265	−0.0026
	500	Mean	0.0020	0.0375	0.0534	−0.0015
		Linex	0.0090	0.0389	0.2314	0.0052
		UT	0.0055	0.0362	0.1519	0.0022
		Naive	0.0105	0.0391	0.2689	0.0067
		Var	0.0060	0.0375	0.1601	0.0025
$\gamma = 5$	750	Mean	0.0060	0.0376	0.1594	0.0026
		Linex	0.0125	0.0389	0.3214	0.0088
		UT	0.0040	0.0368	0.1086	0.0006
		Naive	0.0098	0.0391	0.2497	0.0059
		Var	0.0082	0.0384	0.2147	0.0045
	1000	Mean	0.0045	0.0387	0.1162	0.0008
		Linex	0.0098	0.0389	0.2507	0.0060
		UT	0.0028	0.0373	0.0737	−0.0006
		Naive	0.0208	0.0386	0.5378	0.0104
		Var	0.0217	0.0391	0.5569	0.0150
	500	Mean	0.0215	0.0389	0.5528	0.0135
		Linex	0.0172	0.0391	0.4417	0.0091
		UT	0.0230	0.0405	0.5682	0.0148
		Naive	0.0175	0.0381	0.4593	0.0104
		Var	0.0228	0.0391	0.5825	0.0150
$\gamma = 10$	750	Mean	0.0210	0.0391	0.5377	0.0135
		Linex	0.0165	0.0387	0.4259	0.0091
		UT	0.0168	0.0395	0.4237	0.0089
		Naive	0.0192	0.0386	0.4990	0.0119
		Var	0.0198	0.0384	0.5140	0.0124
	1000	Mean	0.0198	0.0386	0.5119	0.0124
		Linex	0.0175	0.0391	0.4481	0.0099
		UT	0.0235	0.0402	0.5851	0.0155

to get the forecast of the original variable would introduce a bias that should be accounted for.

This paper investigates the finite sample MSE forecasting performances of several bias correction methods under empirically-relevant deviations from the normality of the error distribution. Specifically, we contrast a naïve no-correction approach and a variance-based correction (which is known to be optimal for normally distributed innovations in log-autoregressive models) with the mean-based and Linex-based corrections (which require no distributional assumptions).

We find that the sample size and the degree of autoregressive persistence are the most important for the choice of the optimal correction strategy. The Linex-based correction shows a decent performance for large samples where the estimation risk is negligible, but is subject to numerical instabilities in finite samples. The variance-based correction seems to be the best approach in finite samples, followed closely by the mean-based correction. The untransformed forecasts appear to be less suitable when the model in logs is the correct one. Finally, no correction at all appears to be a reasonable alternative in the case of small samples and highly persistent autoregression. As a practical recommendation, we suggest a hybrid approach where the naïve approach should be applied only when a unit root test does not reject the unit root hypothesis, otherwise some bias correction should be applied, preferably a variance-based correction.

Acknowledgments

This research has received financial support from the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

Appendix A. Consistency of the Linex-based approach for log-transformation

We take the optimization to be conducted over a compact subset Θ of the parameter space, guaranteeing stable autoregressions. Then, given the fact that the innovations ϵ_t are *iid*, the process x_t (which has a causal moving average representation in terms of ϵ_t with absolutely summable coefficients) exists a.s., and $(x_t, \epsilon_t)'$ is a jointly strictly stationary and ergodic process. Now, define

$$b = \arg\min_{b^*} \mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b^*\right)\right],$$

i.e., the M-measure of location of ϵ_t under $\mathcal{L}$. Recall that ϵ_t (and thus x_t) have thin tails, and therefore the above expectation is finite given the linear-exponential behavior of $\mathcal{L}$.

Note that $b = \log\left(\mathrm{E}\left[\exp\left(\epsilon_t\right)\right]\right)$, which is seen to be true since b must satisfy the f.o.c.

$$\mathrm{E}\left[\mathcal{L}'\left(\epsilon_t - b\right)\right] = 0,$$

i.e., $\mathrm{E}\left[\exp\left(\epsilon_t - b\right) - 1\right] = 0$ or $\mathrm{E}\left[\exp\left(\epsilon_t\right)\right] = \exp\left(b\right)$ as required.

Now, the empirical loss to be minimized is

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p}\rho_j^* x_{t-j}\right)$$
$$= \frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(\epsilon_t - b - \left(\mu^* - (\mu + b)\right)\right.$$
$$\left. - \sum_{j=1}^{p}\left(\rho_j^* - \rho_j\right) x_{t-j}\right).$$

Since b is such that the expected loss of $\mathcal{L}\left(\epsilon_t - b\right)$ is smallest, minimizing the empirical loss will result in estimators that are consistent for $\mu + b = \tilde{\mu}$ and ρ_j, as we show below.

Since $\mathcal{L}$ is a strictly convex function of its argument, it follows that

$$\mathrm{E}\left[\mathcal{L}(\cdot)\right] = \mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b - \left(\mu^* - (\mu + b)\right)\right.\right.$$
$$\left.\left. - \sum_{j=1}^{p}\left(\rho_j^* - \rho_j\right) x_{t-j}\right)\right]$$

is a strictly convex function of θ^*. Note also that any linear combination of ϵ_t and (lags of) x_t must have thin tails, as an application of Minkowski's inequality shows; thus, the above expectation is finite and the ergodic theorem indicates that

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p}\rho_j^* x_{t-j}\right) \xrightarrow{a.s.} \mathrm{E}\left[\mathcal{L}(\cdot)\right]$$

pointwise in θ^*. The compactness of Θ and convexity of allow us to use Thm. 10.8 of Rockafellar (1970) to conclude that pointwise a.s. convergence implies uniform convergence,

$$\sup_{\theta^*\in\Theta}\left|\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p}\rho_j^* x_{t-j}\right) - \mathrm{E}\left[\mathcal{L}(\cdot)\right]\right| \xrightarrow{a.s.} 0;$$

furthermore, Thm. 4.1.1 of Amemiya (1985), for example, indicates that

$$\arg\min_{\theta^*}\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p}\rho_j^* x_{t-j}\right)$$
$$\xrightarrow{a.s.} \arg\min_{\theta^*}\ \mathrm{E}\left[\mathcal{L}(\cdot)\right].$$

Given that $\mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b^*\right)\right]$ is minimized at $b^* = b$, it follows that $\mathrm{E}\left[\mathcal{L}(\cdot)\right]$ is minimized at $\tilde{\mu}$ and ρ_j as required. □

Appendix B. Box–Cox transformation

Here, we discuss the possibility of implementing the forecast bias correction methodologies for a Box–Cox (BC) transformation, given as

$$BC\left(y\right) = \frac{y^\lambda - 1}{\lambda} \quad \text{for} \quad y \geq 0, \quad \lambda \neq 0,$$

where the log transformation is obtained as the limit for $\lambda \to 0$. First, note that a simple multiplicative decomposition of the optimal forecast, as in Eq. (2), is not available for any $\lambda \in (0, 1)$.

Under the simplifying assumption that the distribution of $x_t = BC_\lambda(y_t)$ is approximately normal (which could be made reasonably for $0 < \lambda \ll 1$), it can be shown that (cf. Freeman & Modarres, 2006, Lemma 1)

$$\mathrm{E}\left[y_{T+1}|\mathcal{F}_T\right] \approx \left(\lambda\,\mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda} + \sum_{k\geq 1}\frac{\sigma^{2k}}{2^k k!}$$
$$\times\ \left(\lambda\,\mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda - 2k}\left(\prod_{j=0}^{2k-1}(1 - j\lambda)\right).$$

For the special case of $1/\lambda \in \mathbb{N}$, this simplifies to

$$\mathrm{E}\left[y_{T+1}|\mathcal{F}_T\right]$$
$$\approx \sum_{i=0}^{1/\lambda}\binom{1/\lambda}{i}\lambda^i\left(\lambda\,\mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda - i}\mathrm{E}\left[\epsilon_{T+1}^i\right],$$

which is hardly tractable in practice, even if one were to truncate the sum on the right hand side for computational reasons. For this reason, we would recommend relying on bootstrap-based bias correction methods in the case of BC-transformed series.

Still, one may obtain an analog to the Linex-based correction when $\lambda \ll 1$: write

$$y_{T+1}(1) = \mathrm{E}\left[\left(\lambda x_{T+1} + 1\right)^{1/\lambda}|\mathcal{F}_T\right];$$

if we require the point forecast of x_{T+1}, m_{T+1}, to be transformed back for forecasting y_{T+1} using the inverse of the BC transformation, we arrive, as for the case $\lambda = 0$, at the moment condition

$$\mathrm{E}\left[\left(\frac{\lambda x_{T+1} + 1}{\lambda m_{T+1} + 1}\right)^{1/\lambda} - 1\middle|\mathcal{F}_T\right] = 0.$$

This is a legitimate GMM condition that we may employ for estimating any parameters of the model for m_{t+1} in the same way as in the case of the log transformation; however, optimization is more demanding numerically than for the Linex loss. For $\lambda \ll 1$, we may write approximately

$$\mathrm{E}\left[\left(\frac{\lambda x_{t+1} + 1}{\lambda m_{t+1} + 1}\right)^{1/\lambda} - 1\middle|\mathcal{F}_T\right]$$
$$\approx \mathrm{E}\left[\left(1 + \lambda\left(x_{t+1} - m_{t+1}\right)\right)^{(\lambda+1)/\lambda} - 1|\mathcal{F}_T\right] = 0,$$

which again may be written as an extremum estimator that minimizes the observed loss under the loss function

$$\mathcal{L}_\lambda(u) = \frac{1}{\lambda + 1}\left(1 + \lambda u\right)^{(\lambda+1)/\lambda} - u - \frac{1}{\lambda + 1}.$$

This loss has the advantage of being in difference form, and is therefore less difficult numerically. Furthermore, $\mathcal{L}_\lambda$ converges to the Linex function for $\lambda \to 0$ and is actually the squared-error loss function for $\lambda = 1$.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, *89*(4), 701–720.

Andersen, T. G., Bollerslev, T., & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics*, *160*(1), 176–189.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*(2), 171–178.

Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., & Shephard, N. (2011). Subsampling realized kernels. *Journal of Econometrics*, *160*(1), 204–219.

Bauer, G. H., & Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, *160*(1), 93–101.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, *2*(3), 489–502.

Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals of Statistics*, *6*(1), 224–231.

Bondon, P. (2009). Estimation of autoregressive models with epsilon-skew-normal innovations. *Journal of Multivariate Analysis*, *100*(8), 1761–1776.

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, *22*, 3411–3447.

Brechmann, E. C., Heiden, M., & Okhrin, Y. (2018). A multivariate volatility vine copula model. *Econometric Reviews*, *37*(4), 281–308.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Economic*, *7*(2), 174–196.

Corsi, F., Audrino, F., & Renò, R. (2012). HAR modeling for realized volatility forecasting. In L. Bauwens, C. M. Hafner, & S. Laurent (Eds.), *Handbook of volatility models and their applications* (pp. 363–382). Hoboken, NJ: John Wiley & Sons.

Demetrescu, M., & Hassler, U. (2016). (When) do long autoregressions account for neglected changes in parameters? *Econometric Theory*, *32*(6), 1317–1348.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman and Hall.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, *56*(1), 329–352.

Fleming, J., Kirby, C., & Ostdiek, B. (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economic*, *67*, 473–509.

Freeman, J., & Modarres, R. (2006). Inverse Box–Cox: The power-normal distribution. *Statistics & Probability Letters*, *76*(8), 764–772.

Golosnoy, V., Hamid, A., & Okhrin, Y. (2014). The empirical similarity approach for volatility prediction. *Journal of Banking & Finance*, *40*, 321–329.

Golosnoy, V., Okhrin, I., & Schmid, W. (2012). Statistical surveillance of volatility forecasting models. *Journal of Financial Economic*, *10*(3), 513–543.

Granger, C. W. J., & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society. Series B.*, *38*(2), 189–203.

Gribisch, B. (2018). A latent dynamic factor approach to forecasting multivariate stock market volatility. *Empirical Economics*, *55*(2), 621–651.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Heidelberg: Springer.

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Economic*, *4*(4), 594–616.

Lütkepohl, H., & Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics*, *42*(3), 619–638.

Mayr, J., & Ulbricht, D. (2015). Log versus level in VAR forecasting: 42 million empirical answers – Expect the unexpected. *Economics Letters*, *126*, 40–42.

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. New York, NY: John Wiley & Sons.

Patton, A. J., & Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, *140*(2), 884–918.

Proietti, T., & Lütkepohl, H. (2013). Does the Box–Cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, *29*(1), 88–99.

R-Core-Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation of Statistical Computing.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press.

Seidel, W. (2011). Mixture models. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 827–829). Springer Berlin Heidelberg.

Sharafi, M., & Nematollahi, A. (2016). AR(1) model with skew-normal innovations. *Metrika*, *79*(8), 1011–1029.

Stock, J. H., & Watson, M. W. (2012). *Introduction to econometrics* (3rd ed.). Boston, MA: Pearson.

Tarami, B., & Pourahmadi, M. (2003). Multivariate t autoregressions: Innovations, prediction variances and exact likelihood equations. *Journal of Time Series Analysis*, *24*(6), 739–754.

Taylor, N. (2017). Realised variance forecasting under Box–Cox transformation. *International Journal of Forecasting*, *33*, 770–785.

Thombs, L. A., & Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, *85*(410), 486–492.

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In S. Fienberg, & A. Zellner (Eds.), *Studies in Bayesian Econometric and Statistics in Honor of Leonard J. Savage* (pp. 195–208). Amsterdam: North Holland.