Bias Corrections for Exponentially Transformed Forecasts: is it worth the effort?

Matei Demetrescu* — CAU Kiel

Vasyl Golosnoy† — Ruhr-University Bochum

Anna Titova‡ — CAU Kiel

November 3, 2017

Abstract

In many economic applications the log transformation of the process of interest allows to model and to forecast log values as linear time series. However, a reverse transformation of the log forecasts introduces a bias which should accounted for. In this paper we compare different bias correction methods for the reverse transformation of log series following a linear autoregressive process. We find that the correction method to choose in finite samples depends much on the empirical error distribution whereby for some cases no bias correction is advantageous. Our results are illustrated both in Monte Carlo simulations and in an empirical study.

Key words: bias correction, linear autoregression, linex forecast, log transformation

JEL classification: C22, C53, C58

*Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40, D-24118 Kiel, Germany, Email: mdeme@stat-econ.uni-kiel.de.

†Faculty of Management and Economics, Ruhr-Universität Bochum, Universitätsstr. 150, D-44801, Bochum, Germany, Email: vasyl.golosnoy@rub.de

‡Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40, D-24118 Kiel, Germany, Email: atitova@stat-econ.uni-kiel.de.

1 Introduction

The log transformation is widely used in applied time series econometrics to linearize relations or to stabilize variances. It is nowadays a standard transformation for modeling autoregressive (AR) times series in various economic and financial applications, see among others Andersen et al. (2011), Bauer and Vorkink (2011), Lütkepohl and Xu (2012), Brechmann et al. (2016), Gribisch (2017). Along these lines, models in logs often turn out to be better suited both for estimation purposes and for making forecasts. Then the log forecast should be transformed back in order to predict the original variables of interest. Such reverse exponential transformation, however, introduces bias into the forecasting procedure, as already emphasized by Granger and Newbold (1976). Since ignoring this bias could provide substantial losses in forecasting precision (cf. Lütkepohl and Xu, 2012; Proietti and Lütkepohl, 2013), there are several approaches for the bias correction purpose.

The practically relevant question is how to deal with bias corrections in finite sample for various types of model error distributions. Of course, ignoring the bias and simply transforming the forecast in logs through the exponential function is one possible course of action, even if a naive one. At the other end of the spectrum of possibilities, one finds bootstrap-based corrections (cf. Thombs and Schucany, 1990) which could be, however, computationally demanding.

In this paper we concentrate on more straightforward bias correction procedures, such as the method which exploits a residual variance for the correction (particularly suitable by a normality assumption) as well as the method which relies on computing (estimating) the mean (expectation) of the exponentially transformed model residuals. Whereas the variance-based possibility is optimal for normally distributed innovation, the mean-based correction requires only existence of such an expectation. We additionally consider a semiparametric approach based on estimation of the model in logs under the Linex loss (Varian, 1975), which we show to provide unbiased forecasts of the untransformed (original) series such that there is no need for further adjustments. However, this Linex approach leads to a non-linear estimation procedure which could cause losses in estimation efficiency especially undesired in finite samples.

We consider AR settings in logs with model errors following different types of distributions with the interest to make a one step ahead forecast of the original (exponentially transformed) variable. We compare the effectiveness of these bias correction methods and of the naive approach without correction. The forecasting performance of different correction methods has been already studied for several setting, e.g. for a family of data generating processes with Markov switching (cf. Patton and Timmermann, 2007). We extend this strand of literature by specifically focusing on error distributions that exhibit deviations from normality which are of high empirical relevance. In particular, we study the effect of skew-normal (Azzalini, 1985), mixture normal (Everitt and Hand, 1981; McLachlan and Peel, 2004), contaminated normal (Seidel, 2011) as well as t-distributed innovations (cf. Tarami and Pourahmadi, 2003). Since we investigate several AR models with different degrees of persistence, our setup covers a broad class of practically important situations.

We find that in finite samples the variance-based correction appears to be the preferable approach even for various deviations from normality, whereas the expectation-based correction of the residual exponent is a close competitor. Despite of its theoretical attractiveness, the Linex-based approach shows losses in estimation efficiency and appears to be dominated by two above-mentioned alternatives in terms of the considered forecasting loss functions. However, by increasing sample size as the estimation error diminishes, the Linex-based approach gets a preferable one. Surprisingly, a naive prediction without bias correction is found to be suitable for highly persistent AR processes in logs in finite samples. These findings are supported by the empirical results of applying the log heterogenous autoregressive (HAR) model (Corsi et al., 2012) for the purpose of daily realized volatility prediction for highly liquid U.S. stocks.

The rest of the paper is structured as follows. In section 2 we discuss the necessity to make a bias correction, give a summary of established methods suitable for this purpose, and discuss the approach based on estimation under Linex loss. The Monte Carlo study covering various types of AR processes and error distributions is presented in section 3, whereas in section 4 we contrast the behavior of the different bias corrections in an empirical application. The final section 5 concludes the paper, while the appendix collects some technical arguments.

2 Problem Setting and Bias Correction Techniques

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

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

where $\rho(L) = 1 - \sum_{j=1}^{p} \rho_j L^j$ is an invertible lag polynomial of order p.

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:

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

We note that the distribution of x_t (and thus of ϵ_t) must have light tails[1] if $\mathrm{E}[y_{T+1}]$ is to be finite. Denote by $x_T(1) = \mathrm{E}[x_{T+1}|\mathcal{F}_T]$ the one step head MSE-optimal forecast of the log series x_t so that $x_T(1) = \rho(L)x_T$.

Although the representation (1) in logs is quite useful for modeling and estimation purposes, we are often more interested in the out-of-sample one step ahead MSE-optimal forecast $y_T(1) = \mathrm{E}[\exp(x_{T+1})|\mathcal{F}_T]$ of the original variable y_{T+1} which is then given as

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

because of the representation

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

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

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

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

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

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 would estimate these forecast functions by plugging in consistent estimators $\hat{\mu}$ and $\hat{\rho}_j$, leading to $\hat{x}_T(1)$, and address the issue of forecasting bias subsequently. The sample $x_1, ..., x_T$ is used in order to estimate the model parameters in

[1] We take x_t to have light tails if $\mathrm{E}[|x_t|^k] < \infty$ for all $k \in \mathbb{R}^+$.

Equation (1) and any of the bias correction terms.

Hence, the MSE has essentially three sources: the bias, the volatility of the shocks, and the estimation error. Note that estimation error does play a surprisingly important role which gets larger, expectedly, for smaller sample sizes. Although it vanishes as $T \to \infty$, efficient estimation of the parameters may become an issue when the sample is not too large. E.g. for Gaussian errors ϵ_t, a least squares (LS) estimation is maximum likelihood (ML) and, thus, asymptotically efficient, while estimation under the Linex loss may be advantageous under skewed error distributions as it resembles the logarithm of the errors' density in this case.

2.1 Variance-based bias corrections

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

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

Then the feasible variance-corrected forecast is given by

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

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

2.2 Mean-based bias correction

One could estimate the expectation $\mathrm{E}(\exp(\epsilon_{T+1}))$ in (2) directly from the sample, e.g. as the sample average:

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

Then this mean-corrected forecast would be

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

where $\hat{\epsilon}_t$ are the in-sample model residuals.

Since our forecast is based on the conditional mean of y_t, the expectation is implicitly assumed to exist. However, if higher-order moments of y_t do not exist, convergence of the sample average to $\mathrm{E}[\exp(\epsilon_{T+1})]$ may be slow. Of course, more robust estimates of the central tendency (e.g., median or truncated mean) could be applied here as well. Note also that in practice we have to resort to residuals to compute an estimate of $\mathrm{E}[\exp(\epsilon_{T+1})]$, and, therefore, estimation error enters this step.

2.3 Forecasts based on the Linex loss

The considered above variance-based and mean-based bias corrections are two step procedures, as in the first step one should estimate the AR model in (1) and in the second step compute the bias correction factor. We now consider a distribution-free approach which we show to make unbiased forecasts in a single step for exponentially transformed values.

To obtain such a single step forecast let

$$m_{t+1} = \log \mathrm{E}\left(y_{t+1} | y_t, \ldots\right),$$

such that $\exp(m_{t+1}) = \mathrm{E}\left(\exp(y_{t+1}) | y_t, \ldots\right) = \mathrm{E}\left(\exp(x_{t+1}) | y_t, \ldots\right)$, or, equivalently,

$$\mathrm{E}\left(e^{x_{t+1} - m_{t+1}} - 1 \middle| x_t, x_{t-1}, \ldots\right) = 0. \tag{7}$$

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

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

Clearly, m_{T+1} is not the conditional expectation of x_{T+1} given $\mathcal{F}_T$. In a nutshell, this approach predicts the logs in a biased manner, but the bias in the forecasts of the log series is such that the transformation to the original variable y_t leads to its unbiased forecast.

To impose the condition (7), the generalized method of moments (GMM) estimation suggests itself; for our case a particular choice of instruments leads to an estimator with a nice interpretation. To show this, we have from (1) that

$$m_{t+1} = \mu + \sum_{j=1}^{p} \rho_j \, x_{t+1-j},$$

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

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

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

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

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

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

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

and by computing

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

We establish 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 the appendix. This guarantees the MSE-optimality of the forecasts for y_t for large T values.

3 Monte Carlo analysis

Now we examine in Monte Carlo simulations how the shape of the error distribution influences the MSE of the estimated forecasts under consideration in finite samples. In order to contrast various bias corrections, we concentrate on a simple AR(1) process for the log-transformed values because of its immense practical importance. The stationary AR(1) model is given as

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

and we experiment with $\rho \in \{0, 0.5, 0.9\}$. Further we set $\mu = 0$ without loss of generality but estimate it from the data. The estimation of the AR(1) models in logs is conducted based on samples of size $T \in \{200, 500, 1000\}$ with 10^4 iterations.

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

3.1 Distribution of innovations

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

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

We set $\xi = 0$, $\omega = 1$ and compute the standardized innovations for various values of the skewness parameter β:

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

In Case II we assume that ϵ_t follows a standardized symmetric normal mixture distribution (NMD). This is as another very important deviation from normality (cf. McLachlan and 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 the 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

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

In Case III we assume the innovations to follow a contaminated normal distribution which allows for higher kurtosis values (cf. Seidel, 2011). This case is rather similar to the previous Case II specification. The difference lies in the mixture probability that is now set to $\pi = 0.95$. We use again standardized innovations ϵ_t for model estimation and evaluation.

Finally, in Case IV we assume u_t follow a central t-distribution with $v \in [5,30]$ degrees of freedom (dof). The standardized errors ϵ_t are obtained by $\epsilon_t = u_t/(v/(v-2))^{1/2}$. This choce is motivated as a robustness check, since the t-distribution has fat tails, and therefore y_t would not have finite expectation.

3.2 Methods for bias correction

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

1. Naive 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 $\hat{\epsilon}_T, ..., \hat{\epsilon}_1$ as $\hat{\sigma}^2 = 1/(T-1) \cdot \sum_{t=1}^{T} \hat{\epsilon}_t^2$.
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 $\hat{y}_T(1) = \exp(\hat{m}_{T+1})$.

For Cases I–IV we plot in Figures 1-4 the log MSE differences of naive, mean-based and Linex-based forecasts to the baseline variance-based forecast correction which is optimal is case of normal innovations.

[Figures 1–4 about here.]

3.3 Monte Carlo results

The results for Case I with skew-normal distribution of innovations is shown in Figure 1, where we plot the log MSE differences depending on β-value. The obtained evidence is quite similar for $T = 500, 1000$ but to some extent different for $T = 200$. However, it differs much with respect to the autocorrelation parameter value ρ.

For no autocorrelation $\rho = 0$, the naive forecast is dominated by all correction methods. The Linex-based correction is better (but is close to the mean correction) for a pronounced negative skewness with parameter values $\beta < -2$, while variance correction is worse than the Linex-based approach for these β-values. However, the variance-based correction is the best method for positive skewness. This may be explained by the nature of the effect of estimated parameters; for negative skewness, the Linex loss function corresponds to the log-density of innovations and, hence, the estimation under Linex is more efficient than OLS.

For $\rho = 0.5$ the findings remain similar, whereas the Linex-based approach is characterized by some undesired outliers for positive values of skewness β. On the contrary, for strong autocorrelation $\rho = 0.9$ the naive forecast without bias correction dominates all other three correction methods. However, the advantages of naive forecasts get smaller with the increase of the sample size T. This is most likely again due to the nature of the estimation error.

The log differences of MSEs in Case II with mixture of distributions are shown in Figure 2 for different values of σ_2^2. For AR(1) parameters $\rho = 0$ and $\rho = 0.5$ the variance correction method appears to be the best one for $T = 200$, whereas for $T = 500, 1000$ the mean correction is a very close to it. For $\rho > 0$, the Linex-based MSE exhibits undesired peaks which get smaller (but do not disappear) with the increasing T. Hence, it could not be recommended for correction in case of mixtures due to its numerical instability. As this problem gets even more acute in Cases III and IV, we do not report Linex-based results for them. Again, the naive uncorrected forecast appears to be preferable for strong autocorrelation $\rho = 0.9$ where all forecasting methods exhibit more fluctuations in their MSEs.

The results for Case III with contaminated distribution are presented in Figure 3. The mean correction is worse than variance correction in all situations. However, making no bias correction with naive forecast is not worse than the variance-based correction with the exception of small contaminations by $\rho = 0$. In all other cases naive forecast provides more stability in case of high autocorrelation coefficients.

Finally, in Case IV with t-distributed innovations shown in Figure 4 we observe that variance-based forecast is the best one for $\rho = 0$ and $\rho = 0.5$, whereas for $\rho = 0.9$ the naive uncorrected forecast should be used. Again, advantages of the naive forecast decrease with the increase of the sample size T.

Hence, our major findings for no and medium autocorrelation coefficients $\rho \in \{0, 0.5\}$ are as follows. Case I: negative skewness β is in favor of the Linex-based method, whereas for positive β values this method gets instable and variance-based correction is preferable. Case II: variance-based correction is slightly better than mean-based correction in case of mixture distribution. Case III: variance-based correction is not better than naive uncorrected forecasts but is substantially better than mean-based prediction in case of contaminated normal distribution. Case IV: variance correction is suitable for t-distributed innovations.

In Cases I–IV, increase in T leads to smaller fluctuations of the MSEs due to reduction of estimation errors. Linex-based corrections improves much with the increase in T. However, higher values of the autoregressive parameter $|\rho|$ lead to more instability for all studied bias correction methods, so that for e.g. $\rho = 0.9$ and $T = 200$ no bias correction appears to be preferable.

4 Empirical Illustration

The availability of intra-day data allows to estimate the true daily volatility σ_t^2 consistently by its realized measure y_t (cf. Andersen et al., 2007) which is the given time series of interest $\{y_t\}$ in our study. 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 naive uncorrected forecasts with those from the variance-based, mean-based and Linex-based methods for bias correction with the purpose of one step ahead prediction of daily realized volatility.

4.1 HAR model for daily realized volatility

The heterogenous autoregressive (HAR) model of Corsi (2009) appears to be very successful for modeling and forecasting daily realized volatility. In order to asses the complex autoregressive structure of the process $\{y_t\}$, we exploit the HAR model which includes daily, weekly, monthly, and quarterly components (cf. Andersen et al., 2011):

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

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 5, 22, and 65 are the average number of weekly, monthly, and quarterly trading days, respectively.

A disadvantages of the specification in (8) is that the symmetry assumption for the distribution of ε_t is obviously violated due a pronounced asymmetry between positive and negative volatility shocks (cf. Tsay, 2010). For this reason a log transformation $x_t = \log y_t$ is commonly applied for modeling purpose (Andersen et al., 2007) which provides more symmetrically distributed innovations. Then the corresponding HAR model in logs (Corsi et al., 2012) is:

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

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

Due to the documented evidence for the non-normality of the innovations ϵ_{t+1} in log volatility processes (cf. Lanne, 2006), it holds $\exp(\mathrm{E}(x_{t+1}|\mathcal{F}_t)) \neq \mathrm{E}(\exp(x_{t+1})|\mathcal{F}_t)$, so that a forecast bias correction is required. Hence, we estimate the model in (9) and make a forecast of y_{t+1} given the information set $\mathcal{F}_t$ by applying different types of bias corrections.

4.2 Data and descriptive statistics

The data set is from Noureldin et al. (2012) and consists of 10 highly liquid stocks ranging from February 1, 2001 until December 31, 2009 which results in 2242 daily realized variances for each asset. It is available under the link 'http://qed.econ.queensu.ca/jae/datasets/noureldin001/' at the internet site of *Journal of Applied Econometrics*. The considered data set covers both calm and turmoil periods on U.S. financial markets.

In order to investigate the properties of residuals from the log-HAR model in (9), we first estimate it based on the full sample information. The parameter estimates are given in Table 1 whereas the descriptive statistics for model residuals are provided in Table 2.

company	intercept β_0	daily β_1	weekly β_2	monthly β_3	quarterly β_4	variance $\hat{\sigma}^2$
Bank of America	0.002	0.413	0.433	0.057	0.082	0.253
	(0.011)	(0.025)	(0.038)	(0.041)	(0.029)	$R^2 = 0.87$
JP Morgan	0.008	0.399	0.414	0.064	0.106	0.252
	(0.013)	(0.025)	(0.039)	(0.043)	(0.032)	$R^2 = 0.84$
IBM	-0.004	0.3	0.474	0.13	0.062	0.223
	(0.010)	(0.026)	(0.041)	(0.048)	(0.034)	$R^2 = 0.74$
Microsoft	0.003	0.29	0.477	0.109	0.096	0.222
	(0.011)	(0.026)	(0.042)	(0.048)	(0.035)	$R^2 = 0.75$
Exxon Mobil	0.008	0.309	0.526	0.06	0.061	0.217
	(0.011)	(0.026)	(0.040)	(0.044)	(0.033)	$R^2 = 0.70$
Alcoa	0.032	0.258	0.496	0.143	0.071	0.227
	(0.018)	(0.026)	(0.042)	(0.050)	(0.037)	$R^2 = 0.73$
American Express	0.004	0.334	0.467	0.073	0.11	0.250
	(0.012)	(0.026)	(0.040)	(0.045)	(0.033)	$R^2 = 0.85$
Du Pont	0.011	0.302	0.481	0.094	0.09	0.216
	(0.012)	(0.026)	(0.041)	(0.048)	(0.036)	$R^2 = 0.73$
General Electric	0.002	0.296	0.478	0.102	0.105	0.260
	(0.012)	(0.026)	(0.042)	(0.049)	(0.036)	$R^2 = 0.80$
Coca Cola	-0.011	0.3	0.451	0.123	0.089	0.203
	(0.010)	(0.026)	(0.042)	(0.049)	(0.037)	$R^2 = 0.71$

Table 1: Parameter estimates (st. deviations) for the full sample log-HAR model in (9).

All regressor coefficients are significantly larger than zero supporting the extended HAR specification. The R^2 measures for all assets are quite high, between 0.7 and 0.87. The estimated models for the considered 10 companies show no unit root behavior as $\sum_{i=1}^{4} \hat{\beta}_i < 1$ for all of them. However, the degree of autoregressive persistence appears to be rather high.

company	sample skewness	sample kurtosis	AR(1) coefficient	Ljung-Box(20) p-value	dof t-distr.
Bank of America	0.553	4.957	-0.021	0.233	8.608
JP Morgan	0.214	4.080	-0.015	0.044	8.779
IBM	0.125	4.628	-0.015	0.115	6.576
Microsoft	0.139	4.042	-0.015	0.058	8.733
Exxon Mobil	0.250	4.114	-0.016	0.035	8.745
Alcoa	0.253	3.932	-0.011	0.202	10.433
American Express	0.199	5.287	-0.024	0.003	8.159
Du Pont	0.186	4.125	-0.013	0.392	8.710
General Electric	0.247	3.855	-0.016	0.001	9.970
Coca Cola	0.232	4.129	-0.017	0.266	10.035

Table 2: Descriptive statistics of residuals from the full sample log-HAR model in (9).

Residuals appear to be right-skewed, exhibit kurtosis around 4 and 5, and show no series autocorrelation as we report in Table 2. The estimated degrees of freedom for the t-distribution refer to the interval $(6, 10)$. In general, these HAR models seem to provide a reasonable time series specifications for the log realized volatilities. Based on our Monte Carlo results, we could expect that these deviations from normality would lead to a better performance of variance-based forecast bias correction.

4.3 Comparison of bias correction methods

For making one step ahead volatility predictions, the log-HAR model in (9) is re-estimated based on the moving window of size $T \in \{200, 500, 750, 1000\}$ days. We set variance-based correction as a benchmark and compare it with the naive, mean-based and Linex-based methods. The corresponding log MSE differences for all 10 stocks are presented in Table 3.

Bank of America			
T	naive/var	mean/var	Linex/var
200	−0.0336	0.0111	0.1520
500	−0.0286	0.0150	0.0357
750	−0.0142	0.0182	0.0258
1000	−0.0038	0.0109	0.0381
IBM			
T	naive/var	mean/var	Linex/var
200	−0.0201	0.0095	0.2557
500	0.0107	0.0042	0.0320
750	0.0208	0.0013	0.0284
1000	0.0286	−0.0003	0.0181
Exxon Mobil			
T	naive/var	mean/var	Linex/var
200	−0.0237	0.0035	0.1811
500	0.0207	0.0003	0.0256
750	0.0352	−0.0009	0.0073
1000	0.0417	−0.0016	−0.0056
American Express			
T	naive/var	mean/var	Linex/var
200	−0.0160	0.0029	0.1585
500	−0.0010	0.0020	0.0199
750	−0.0007	0.0033	0.0156
1000	0.0069	0.0030	0.0115
General Electric			
N	naive/var	mean/var	Linex/var
200	−0.0464	0.0141	0.2347
500	0.0088	0.0043	0.0820
750	0.0190	0.0015	0.0751
1000	0.0246	0.0004	0.0659

JP Morgan			
T	naive/var	mean/var	Linex/var
200	−0.0069	0.0016	0.0961
500	−0.0019	0.0032	0.0309
750	0.0120	0.0018	0.0229
1000	0.0203	0.0007	0.0192
Microsoft			
T	naive/var	mean/var	Linex/var
200	−0.0076	0.0069	0.0921
500	0.0299	0.0011	0.0017
750	0.0368	0.0001	0.0072
1000	0.0433	−0.0007	0.0155
Alcoa			
T	naive/var	mean/var	Linex/var
200	−0.0390	0.0013	0.0763
500	0.0176	0.0028	−0.0073
750	0.0316	0.0004	−0.0038
1000	0.0360	−0.0006	−0.0056
Du Pont			
T	naive/var	mean/var	Linex/var
200	0.0057	0.0030	0.0894
500	0.0195	0.0016	0.0092
750	0.0276	0.0008	0.0033
1000	0.0360	−0.0005	−0.0022
Coca Cola			
T	naive/var	mean/var	Linex/var
200	−0.0241	0.0102	0.2112
500	0.0033	0.0029	0.0619
750	0.0118	0.0012	0.0446
1000	0.0195	0.0001	0.0234

Table 3: Log of MSE ratios for one day ahead volatility forecasts from the log-HAR model in (9) estimated based on moving windows of size T.

The major findings for MSE are summarized as follows. Naive uncorrected prediction is absolutely the best (by 9 out of 10 stocks, see Table 4) for $T = 200$; variance-based correction is the best for $T = 500$, for larger values of T the mean-based correction is a close competitor. Naive uncorrected approach is slightly worse than variance-based one for $T = 500$ and gets even worse with increase of T. The Linex-based correction gets better with the increase of T and is a decent alternative for $T = 1000$. These results were to expect based on the time series and residual properties summarized in Table 2. In particular, the HAR model is persistent but we observe there almost no residual autocorrelation. The residuals exhibit right-side skewness and slight excess kurtosis. These features correspond to Cases I and II in our Monte Carlo where we report that variance-based correction is the best one for small T whereas the Linex-based correction improves by increase in T.

T	naive	mean	Linex	var
200	9/10	0/10	0/10	1/10
500	3/10	0/10	0/10	7/10
700	2/10	1/10	0/10	7/10
1000	1/10	2/10	3/10	4/10

Table 4: How often the bias correction method provides the smallest MSE out of 10 stocks

Note that although the numerical differences in the MSE in Table 3 are not so large, looking for the best point volatility forecast is still of much economic relevance. E.g., since volatility prediction is of importance for pricing derivative financial instruments, such as European and American options etc. (cf. Tsay, 2010), even a small improvement of daily volatility forecasts could lead to substantial economic gains or losses.

5 Summary

Making forecasts with an autoregressive model for log-transformed variables is a convenient possibility in numerous applications. A reverse transformation in order to get the forecast of the original variable, however, would introduce a bias which should be accounted for. For normally distributed innovations in the log-autoregressive models, the variance-based correction appears to be optimal. The alternative mean-based and Linex-based correction approaches require no distributional assumption.

In this paper we investigate a finite sample MSE forecasting performance of several bias correction methods. Namely, we contrast a naive no-correction approach, the variance-based correction with the mean-based and Linex-based corrections under important deviations from normality of the error distribution.

We find that the sample size and the degree of autoregressive persistence are of most importance for the optimal correction strategy. For large samples where the estimation risk gets negligible, the Linex-based correction shows decent performance, however, in finite samples it is subject to numerical instabilities. The variance-based correction seems to be the best approach in finite samples, closely followed by the mean-based correction. However, in case of small samples and highly persistent autoregression, no correction at all appears to be a reasonable alternative.

Acknowledgements

This research has been in part financially supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

References

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

Andersen, T. G., T. Bollerslev, and F. X. Diebold (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., T. Bollerslev, and X. Huang (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.

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

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

Brechmann, E. C., M. Heiden, and Y. Okhrin (2016). A multivariate volatility vine copula model. *Econometric Reviews*, 1–28.

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

Corsi, F., F. Audrino, and R. Renò (2012). HAR Modeling for Realized Volatility Forecasting. In *Handbook of Volatility Models and Their Applications*, pp. 363–382. Hoboken, NJ: John Wiley & Sons.

Everitt, B. S. and D. J. Hand (1981). *Finite Mixture Distributions.* London: Chapman and Hall.

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

Gribisch, B. (2017). A latent dynamic factor approach to forecasting multivariate stock market volatility. *Empirical Economics*, 1–31. in press.

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

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

McLachlan, G. and D. Peel (2004). *Finite Mixture Models*. New York, NY: John Wiley & Sons.

Noureldin, D., N. Shephard, and K. Sheppard (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 27*(6), 907–933.

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

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

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

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

Thombs, L. A. and W. R. Schucany (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* (3 ed.). Hoboken, NJ: John Wiley & Sons.

Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment. In S. E. Fienberg and A. Zellner (Eds.), *Studies in Bayesian Econometric and Statistics in Honor of Leonard J. Savage*, pp. 195–208. Amsterdam: North Holland.

Appendix

Appendix A: consistency of 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 exists, and $(x_t, \epsilon_t)'$ is a jointly stationary and ergodic process. Define now

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

i.e. the M-measure of location of ϵ_t under $\mathcal{L}$. The empirical loss to be minimized is

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

Since b is such that the expected loss of $\mathcal{L}(\epsilon_t - b)$ is smallest, minimizing the empirical loss will result in estimators consistent for $\mu + b$ and ρ_j as we show below. Note that $\hat{\mu}$ is biased asymptotically as we show it to converge a.s. to $\mu + b$, and $b \neq 0$ in general, but we also prove that $\exp(m_{T+1}(\hat{\theta}))$ is then a consistent estimator of $\mathrm{E}\left[\exp\left(x_{T+1}\right)\right]$.

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

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

is a strictly convex function of θ^*. To show convergence, note that the ergodic theorem indicates

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

pointwise in θ^*. Therefore, using Thm. 10.8 in Rockafellar (1970), it follows that

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

and e.g. Thm. 4.1.1 in Amemiya (1985) 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, it follows that $\mathrm{E}\left[\mathcal{L}(\cdot)\right]$ is minimized at $\mu + b$ and ρ_j. To show that the limit is the desired one, we need to establish that $\mathrm{E}\left[\exp\left(\epsilon_t\right)\right] = \exp(b)$, 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 $\mathrm{E}\left[\exp\left(\epsilon_t\right)\right] / \exp(b) = 1$ as required.

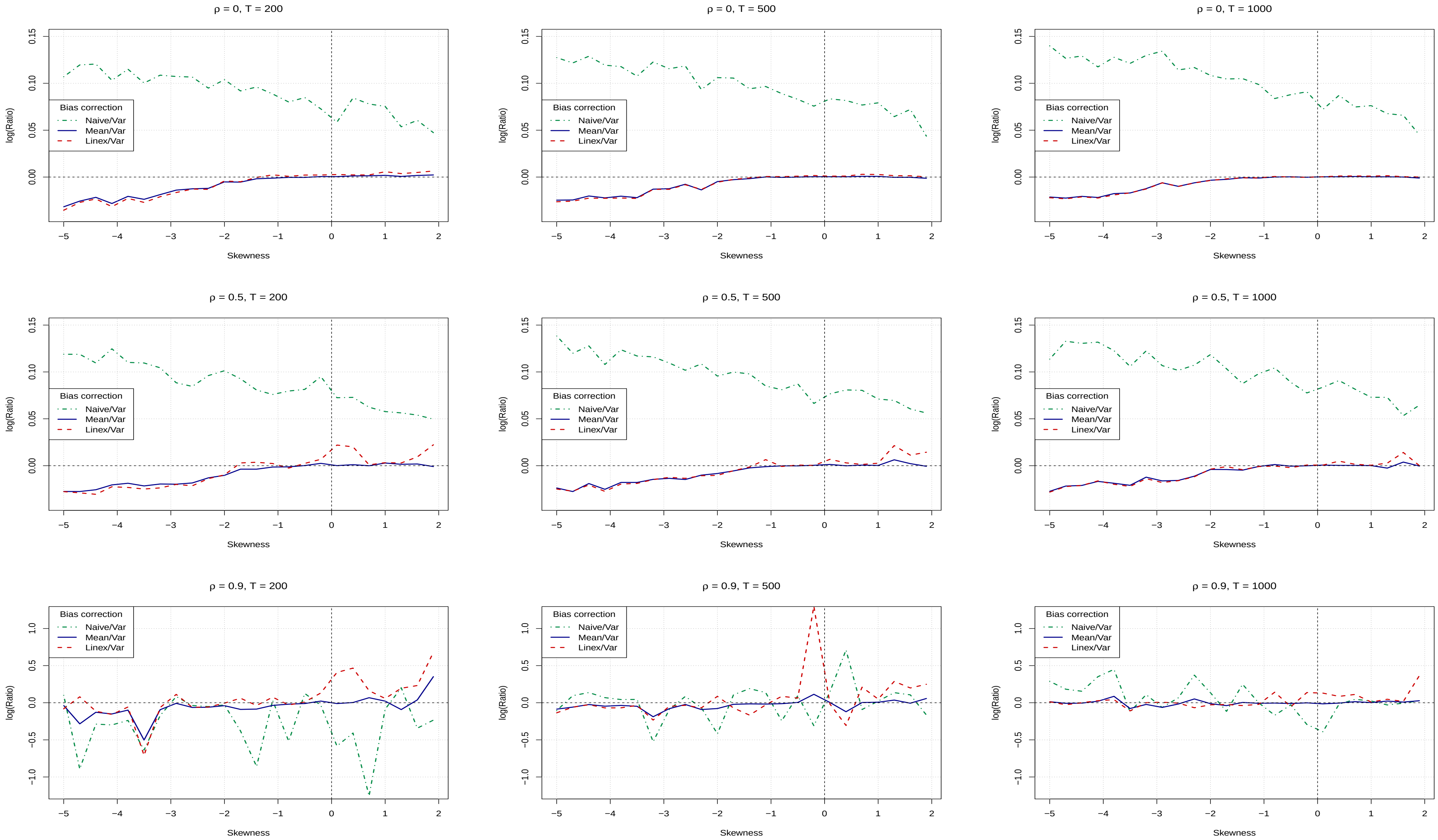

Figure 1: Log MSE differences for $\rho \in \{0, 0.5, 0.9\}$ from above to below, $T \in \{200, 500, 1000\}$ from left to right, skewness parameter $\beta \in (-5,\ 2)$ with $\beta = 0$ for a symmetric distribution.

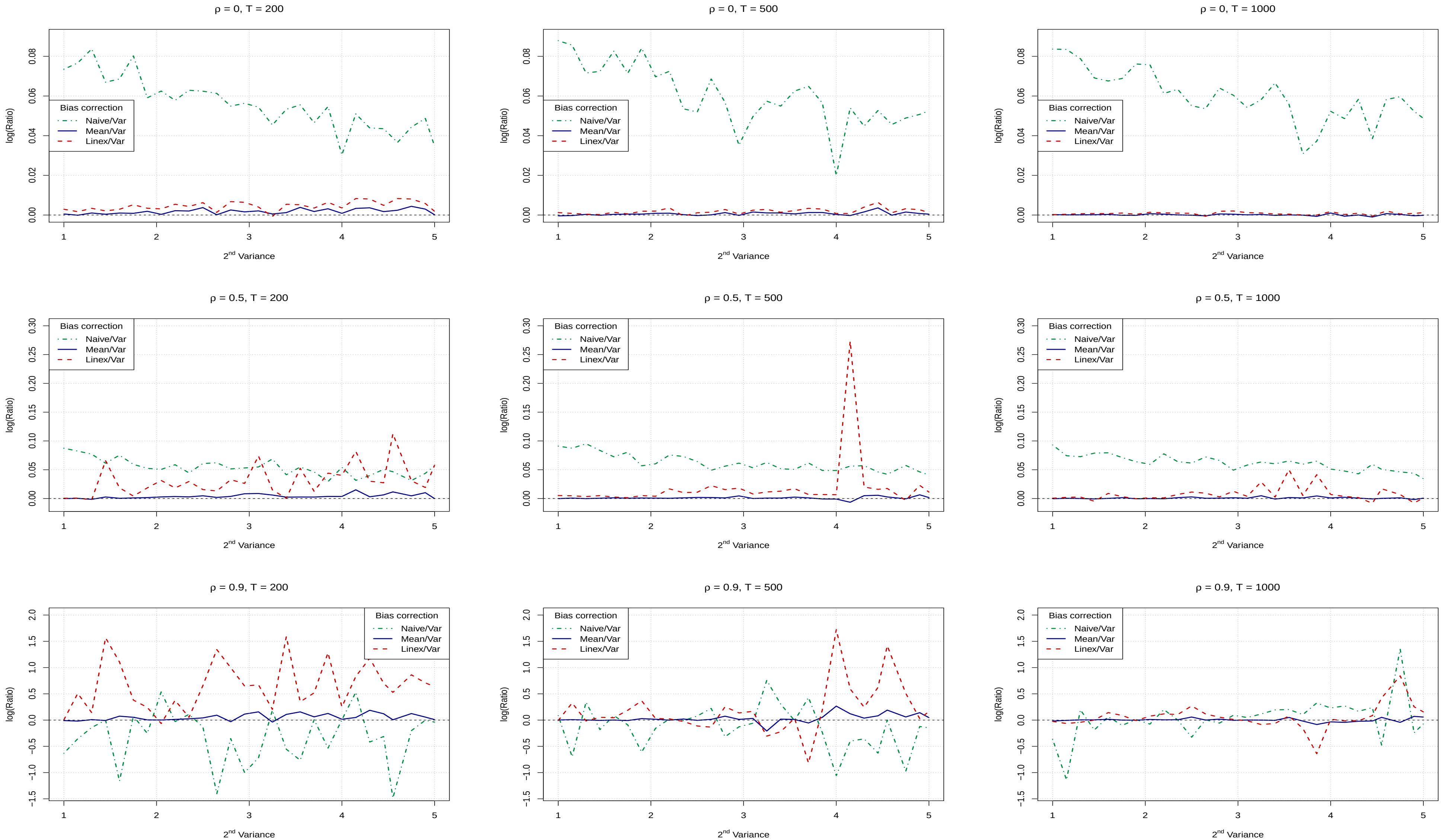

Figure 2: Log MSE differences for $\rho \in \{0, 0.5, 0.9\}$ from above to below, $T \in \{200, 500, 1000\}$ from left to right, mixture of $\mathcal{N}(0,1)$ and $\mathcal{N}(0,\sigma_2^2)$ with probability $\pi = 0.5$ and $\sigma_2^2 \in (1,\ 5)$.

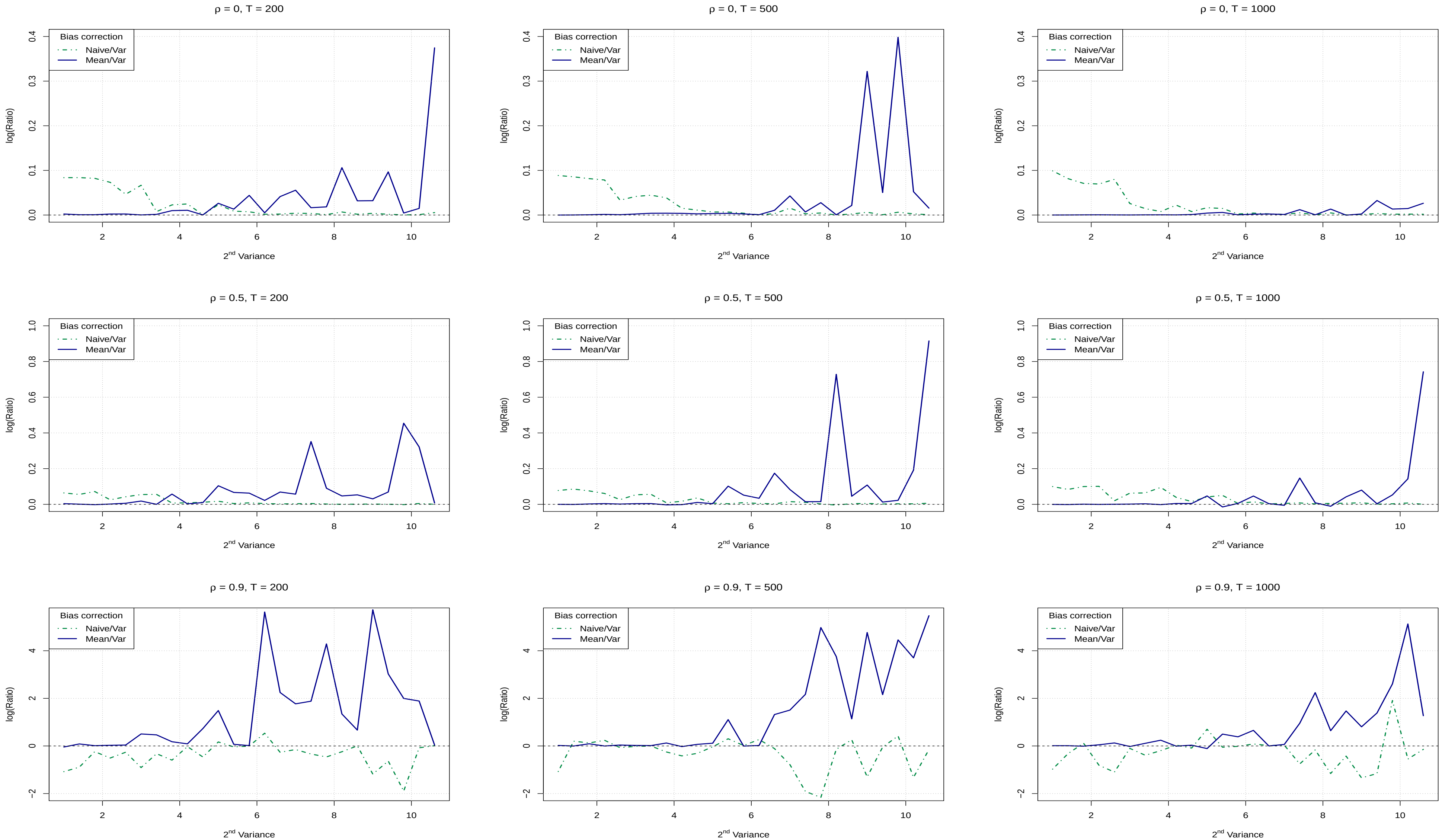

Figure 3: Log MSE differences for $\rho \in \{0, 0.5, 0.9\}$ from above to below, $T \in \{200, 500, 1000\}$ from left to right, $\mathcal{N}(0,1)$ contaminated with probability $1-\pi = 0.05$ by $\mathcal{N}(0, \sigma_2^2)$, $\sigma_2^2 \in (1,\ 30)$.

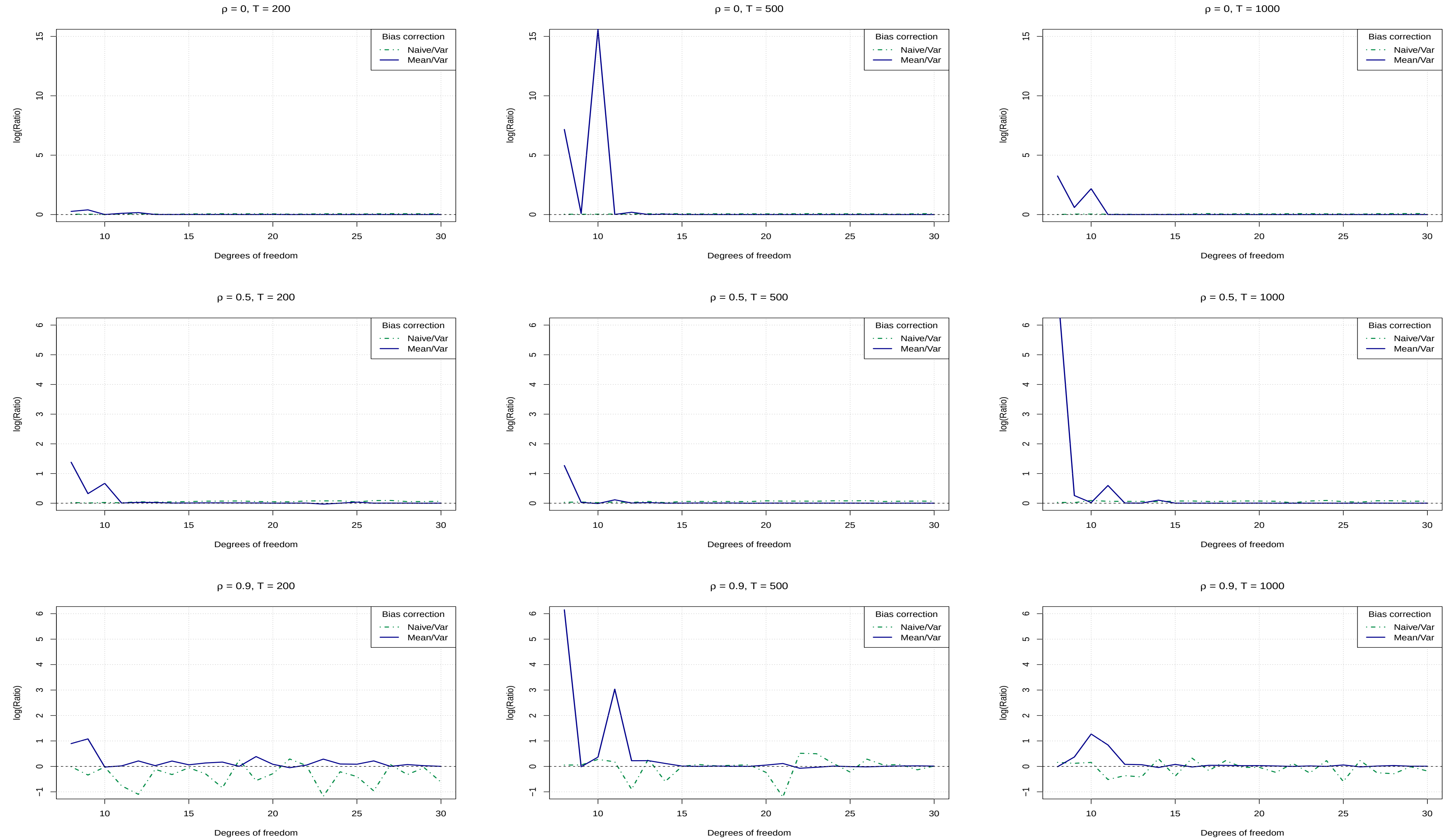

Figure 4: Log MSE differences for $\rho \in \{0, 0.5, 0.9\}$ from above to below, $T \in \{200, 500, 1000\}$ from left to right, t-distribution with $v \in (5, 30)$ degrees of freedom.