Box-Cox Transforms for Realized Volatility*

Sílvia Gonçalves† and Nour Meddahi‡

Université de Montréal and Imperial College London

January 8, 2008

Abstract

The log transformation of realized volatility is often preferred to the raw version of realized volatility because of its superior finite sample properties. One of the possible explanations for this finding is the fact the skewness of the log transformed statistic is smaller than that of the raw statistic. Simulation evidence presented here shows that this is the case. It also shows that the log transform does not completely eliminate skewness in finite samples. This suggests that there may exist other nonlinear transformations that are more effective at reducing the finite sample skewness.

The main goal of this paper is to study the accuracy of a new class of transformations for realized volatility based on the Box-Cox transformation. This transformation is indexed by a parameter β and contains as special cases the log (when $\beta = 0$) and the raw (when $\beta = 1$) versions of realized volatility. Based on the theory of Edgeworth expansions, we study the accuracy of the Box-Cox transforms across different values of β. We derive an optimal value of β that approximately eliminates skewness. We then show that the corresponding Box-Cox transformed statistic outperforms other choices of β, including $\beta = 0$ (the log transformation). We provide extensive Monte Carlo simulation results to compare the finite sample properties of different Box-Cox transforms. Across the models considered in this paper, one of our conclusions is that $\beta = -1$ (i.e. relying on the inverse of realized volatility also known as realized precision) is the best choice if we want to control the coverage probability of 95% level confidence intervals for integrated volatility.

Keywords: Realized volatility, Box-Cox transformation, Edgeworth expansions.

*This work was supported by grants FQRSC, SSHRC, MITACS, NSERC and Jean-Marie Dufour's Econometrics Chair of Canada. This paper was written while the second author was visiting CREST. He thanks the CREST for its hospitality and financial support. We would like to thank a referee and the two editors for their comments that have importantly improved the paper. Thanks also to Roméo Tedongap for excellent research assistance.

†Département de sciences économiques, CIREQ and CIRANO, Université de Montréal. Address: C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: (514) 343 6556. Email: silvia.goncalves@umontreal.ca

‡Finance and Accounting Group, Tanaka Business School, Imperial College London. Address: Exhibition Road, London SW7 2AZ, UK. Phone: +44 207 594 3130. Fax: +44 207 823 7685. E-mail: n.meddahi@imperial.ac.uk.

1 Introduction

The logarithmic transformation of realized volatility (the sum of squared intraday returns) is known to have better finite sample properties than realized volatility and for this reason it is often used in empirical applications involving realized volatility. This transformation was first used by Andersen, Bollerslev, Diebold and Labys (2001) to induce normality for realized volatility measures for high frequency exchange rates returns. Andersen, Bollerslev, Diebold and Ebens (2001) applied it to high frequency stock returns. Barndorff-Nielsen and Shephard (henceforth BN-S) (2002) studied the first order asymptotic properties of the log transformation, whereas BN-S (2005) showed by simulation that the finite sample distribution of the log transformation was closer to the asymptotic standard normal distribution than the finite sample distribution of the non-transformed version of realized volatility. Gonçalves and Meddahi (2008, henceforth GM (2008)) provide a theoretical explanation of this finding based on Edgeworth expansions of the distribution of realized volatility and of its log version.

In this paper we introduce a new class of nonlinear transformations of realized volatility based on the Box-Cox transformation. This transformation is indexed by a parameter β and includes as special cases the log transformation (when $\beta = 0$) and the linear (or raw) version of realized volatility (when $\beta = 1$). BN-S (2002) briefly discuss some nonlinear transformations for realized volatility as an alternative to the log transform, but no higher order theory is presented. Here we study the higher order asymptotic accuracy of Box-Cox transforms across different values of β. Our results suggest that the log transformation can be improved upon by choosing values of β other than zero. We derive an optimal value of β for which the skewness of the higher order expansion of the distribution associated with this transformation is equal to zero and show that its finite sample properties are often superior to those associated with other Box-Cox transformations (including the log). The transformation based on the optimal value of β is infeasible because it depends on the volatility process. We therefore propose a consistent estimator of the optimal value of β and study its finite sample properties. Our results suggest that estimation of β can induce significant distortions in finite samples. In practice, a fixed value of β equal to -1 (the so-called precision) proves to be a better choice if the target is to obtain 95% level confidence intervals for integrated volatility whose actual coverage is close to the desired level of 95%. The value of $\beta = -1$ yields coverage probabilities that are roughly similar to those obtained with the i.i.d. bootstrap for $\beta = 1$ (see GM (2008)). However, the i.i.d. bootstrap for $\beta = 0$ tends to outperform the confidence intervals based on $\beta = -1$, which suggests that an i.i.d. bootstrap for $\beta = -1$ may yield even more accurate intervals than the first order asymptotic theory based intervals studied in this paper.

Power transformations have for a long time been known to improve the quality of the asymptotic normal distribution in finite samples. A leading example is the Wilson and Hilferty (1931) cube root transformation of chi-squared i.i.d. random variables. More recently, Chen and Deo (2004) propose power transformations for statistics that are positive linear combinations of positive independent random variables, thus generalizing the Wilson and Hilferty (1931) power transformation. In particular, they provide a method to choose the optimal power that approximately eliminates skewness in finite samples for statistics whose asymptotic variance is known and does not need to be estimated (for realized volatility-based statistics, these correspond to infeasible statistics because in practice we do not know the asymptotic variance of realized volatility). Chen and Deo's (2004) setup allows for linear combinations of independent and possibly heterogeneous random variables. Because conditionally on the volatility path, the returns are independent but heteroskedastic, the optimal power transformations that we derive here for the infeasible statistics can be obtained as a special case of Chen and Deo's (2004) results. Since in our context these transformations are a function of the volatility path and therefore are infeasible, we propose consistent estimators for the optimal power transformations.

In this paper, we also derive the optimal Box-Cox transform for feasible statistics based on realized volatility. These are studentized statistics which replace the asymptotic variance with a consistent estimator and are not studied in Chen and Deo (2004). Our results show that the optimal Box-Cox transformation that approximately eliminates skewness for the feasible transformed statistic is different from the optimal Box-Cox transformation that applies to the infeasible statistic. When returns are homoskedastic, the optimal transform corresponds to $\beta = -1/3$ instead of $\beta = 1/3$.

Unlike the infeasible statistic, the feasible statistic is biased (due to the estimation of the asymptotic variance). Thus we also derive the optimal Box-Cox transform that eliminates the bias. Under constant volatility, it corresponds to $\beta = -1$. Under stochastic volatility, our optimal Box-Cox transforms depend on the volatility path and need to be estimated. Our simulations show that estimation of the optimal power transforms creates some finite sample distortions. In practice, a fixed $\beta = -1$ reduces both the bias and the skewness and therefore yields good coverage probabilities.

The idea of using Edgeworth expansions to compare the accuracy of the asymptotic distribution for alternative statistics of interest has been used previously in other contexts. An example is Phillips and Park (1988), who use Edgeworth expansions to investigate alternative forms of the Wald test for nonlinear restrictions. Phillips and Park (1988) derive Edgeworth expansions for the distribution of Wald tests for different formulations of a given nonlinear restriction and utilize the associated correction terms to determine which form is more closely approximated by the asymptotic distribution.

The rest of this paper is organized as follows. In Section 2 we introduce the setup and study by simulation the finite sample properties of the log transform for realized volatility in comparison with the raw realized volatility. In Section 3 we study the higher order properties of the Box-Cox transformation. We complement our theoretical analysis with Monte Carlo simulation results. Section 3.1 considers infeasible statistics, for which the asymptotic variance is assumed known. In Section 3.2 we study the Box-Cox transformed feasible statistics, which replace the unknown asymptotic variance with a consistent estimator. Section 4 concludes. Appendix A contains details on the Monte Carlo design and on how to construct confidence intervals for integrated volatility based on Box-Cox transformed realized volatility statistics. Appendix B contains the proofs of the our theoretical results.

2 Setup and Motivation

The goal of this section is to explore by simulation some possible explanations for the improved accuracy of the normal distribution when applied to the log transform. As we will see, the skewness of the log transformed realized volatility is substantially smaller than that of the non-transformed statistic and this can explain why the log transform performs better in finite samples. Our results also show that the log transform does not completely eliminate the skewness in finite samples. This finding suggests that there may exist other nonlinear transformations that more efficiently reduce the finite sample skewness. This motivates our study of the Box-Cox transformation in the next section.

We consider the same setup as in GM (2008). More specifically, the log price process $\{\log S_t : t \geq 0\}$ follows $d \log S_t = \sigma_t dW_t$, where W_t denotes a standard Brownian motion and σ_t a volatility term. We assume the drift term is zero and we suppose the independence between W_t and σ_t. Because our higher order cumulants expansions rely on the higher order expansions for general nonlinear statistics derived in GM (2008), we exclude drift and leverage effects, consistent with Assumption H of GM (2008).

Intraday returns at a given horizon h are denoted r_i and are defined as $r_i \equiv \log S_{ih} - \log S_{(i-1)h} = \int_{(i-1)h}^{ih} \sigma_u dW_u$, for $i = 1, \ldots, 1/h$, with $1/h$ an integer. The parameter of interest is the integrated volatility over a day, $\overline{\sigma^2} = \int_0^1 \sigma_u^2 du$, where we have normalized the daily horizon to be the interval $(0, 1)$. The realized volatility estimator is defined as $R_2 = \sum_{i=1}^{1/h} r_i^2$. Following the notation of GM (2008), we let the integrated power volatility be denoted by $\overline{\sigma^q} \equiv \int_0^1 \sigma_u^q du$ for any $q > 0$. The realized q-th order power variation is defined as $R_q = h^{-q/2+1} \sum_{i=1}^{1/h} |r_i|^q$. Under certain regularity conditions (see BN-S (2004) and Barndorff-Nielsen et. al. (2006)), $R_q \xrightarrow{P} \mu_q \overline{\sigma^q}$, where $\mu_q = E|Z|^q$, $Z \sim N(0, 1)$. As $h \to 0$, it also follows that (BN-S (2002), Jacod and Protter (1998))

$$\frac{\sqrt{h^{-1}}\left(R_2 - \overline{\sigma^2}\right)}{\sqrt{V}} \to^d N(0,1), \tag{1}$$

where $V = 2\overline{\sigma^4}$ is the asymptotic variance of $\sqrt{h^{-1}} R_2$. Similarly, as first proved by BN-S (2002),

$$\frac{\sqrt{h^{-1}}\left(R_2 - \overline{\sigma^2}\right)}{\sqrt{\hat{V}}} \to^d N(0,1), \tag{2}$$

where $\hat{V} = \frac{2}{3} R_4$ is a consistent estimator of V.

The log versions of (1) and (2) are given by

$$\frac{\sqrt{h^{-1}}\left(\log(R_2) - \log\left(\overline{\sigma^2}\right)\right)}{\sqrt{\frac{V}{\left(\overline{\sigma^2}\right)^2}}} \to^d N(0,1) \tag{3}$$

and

$$\frac{\sqrt{h^{-1}}\left(\log(R_2) - \log\left(\overline{\sigma^2}\right)\right)}{\sqrt{\frac{\hat{V}}{(R_2)^2}}} \to^d N(0,1), \tag{4}$$

respectively.

We compare the log transformed statistics with the raw statistics in a Monte Carlo study. The design is the same as in GM (2008). In particular, we consider two stochastic volatility models. The first model is the GARCH(1,1) diffusion of Andersen and Bollerslev (1998) and the second model is a two-factor diffusion model analyzed by Huang and Tauchen (2006). We assume no drift and no leverage effects. 100,000 Monte Carlo replications are used throughout. Appendix A contains more details on the Monte Carlo design.

We consider two types of statistics: infeasible statistics, such as (1) and (3), for which the asymptotic variance depends on the true volatility process, and feasible statistics, such as (2) and (4), which are based on a consistent estimator of the asymptotic variance. We include infeasible statistics in our study for two reasons. First, the contrast between the infeasible and the feasible statistics is important to help us understand the finite sample performance of the feasible statistics (in particular, estimation of the asymptotic variance matters in finite samples and it deteriorates the properties of the feasible statistic). Second, as we will show next, the optimal Box-Cox transformation is different when applied to the infeasible or to the feasible statistics. By including the infeasible statistic in the Monte Carlo study we can evaluate the performance of these two choices for both types of statistics.

Tables 3 and 4 contain the results for the infeasible R_2-based statistics for the GARCH(1,1) and the two-factor models, respectively. Tables 5 and 6 contain the corresponding results for the feasible versions of these statistics. In these tables we find results for several Box-Cox transformed statistics, but for now we only discuss the results for the raw statistics ($\beta = 1$) and for their log transforms ($\beta = 0$). For each sample size ($h^{-1} = 12, 48, 288$ and 1152), we computed the finite sample summary statistics for each version of the R_2-based statistic, namely the mean, the standard error, the skewness[1] and the excess kurtosis, across the 100,000 Monte Carlo replications. We also computed the coverage probability of two-sided symmetric and lower one-sided 95% level intervals for integrated volatility. See Appendix A for more details on how to build confidence intervals for integrated volatility based on the infeasible and feasible statistics.

Tables 3 and 4 suggest the bias is smallest (and close to zero) for the raw version of the infeasible statistic. This is in line with the fact that the raw statistic has zero bias by construction. In contrast, the log transform introduces a negative bias, especially at the smallest sample sizes. The raw statistics have right skewed distributions, especially for the small sample sizes and for the two-factor diffusion. The degree of asymmetry is smaller for the log version than for the raw version of the statistics. This transformation clearly reduces the amount of finite sample skewness and is particularly effective in doing so for the two-factor diffusion. The log transform is also able to reduce excess kurtosis. The actual coverage rates of symmetric 95% level confidence intervals based on the raw statistics are close to the desired level for all sample sizes. Interestingly, the one-sided intervals are too conservative whereas the log based intervals tend to undercover (the latter is probably due to the negative bias).

[1]The tables report the finite sample third central moment of the statistic of interest and not the skewness as it is usually calculated (i.e. we do not scale this quantity by the third power of the standard error). Equivalently, we report the finite sample third cumulants. The main reason for doing so is that we want to compare this empirical measures with the Edgeworth correction predictions for the third order cumulant of the statistics of interest.

We now turn to Tables 5 and 6, which contain the results for the feasible statistics. The main difference with respect to Tables 3 and 4 is that there is a clear deterioration of the finite sample performance of the raw statistics for all sample sizes and models (although more so for the two-factor diffusion than for the GARCH(1,1) diffusion). The raw statistics are negatively biased in finite samples. This bias is introduced by the estimation of the asymptotic variance. Moreover, the amount of skewness and excess kurtosis can be quite substantial, especially for the small sample sizes and for the two-factor diffusion. The log transformation is effective in reducing the magnitude of the bias, skewness and excess kurtosis as well as in producing less variability for the two models. The coverage rates are therefore closer to the desired level of 95% than those of the raw statistics. However, significant coverage distortions remain for the log transformation, especially when $h^{-1} = 12$ and 48.

Since the normal distribution is characterized by zero mean, unit variance, zero skewness and zero excess kurtosis, a statistic that more closely matches these moments is expected to be better approximated by the normal than otherwise. Our results suggest that this is the case for the log transform. GM (2008) provide a theoretical result that confirms this simulation finding.

Our results also suggest that the log transformation is not completely effective in eliminating the bias and the skewness in finite samples, which may explain why there are still some distortions in coverage probabilities at the smaller sample sizes. This is why we consider in the next section the more general Box-Cox transformation.

3 The Box-Cox Transformation

The Box-Cox transformation for realized volatility is defined as

$$g\left(R_2;\beta\right) = \begin{cases} \frac{R_2^{\beta}-1}{\beta} & \text{when } \beta \neq 0 \\ \log\left(R_2\right) & \text{when } \beta = 0. \end{cases} \tag{5}$$

It contains the log transformation for realized volatility (when $\beta = 0$) and the raw statistic (when $\beta = 1$) as special cases. Our main goal in this section is to study the accuracy of this transformation for several values of β.

The infeasible Box-Cox transformed statistic is given by

$$S_{\beta} = \frac{\sqrt{h^{-1}}\left(g\left(R_2;\beta\right) - g\left(\overline{\sigma^2};\beta\right)\right)}{g'\left(\overline{\sigma^2};\beta\right)\sqrt{V}}.$$

Here and throughout the paper, $g'\left(x;\beta\right)$ denotes the derivative of g with respect to x. Similarly, $g''\left(x;\beta\right)$ denotes the second derivative of g with respect to x. By an application of the delta method, we have that $S_{\beta} \to^{d} N\left(0,1\right)$. When $\beta = 0$, S_{β} corresponds to (3). When $\beta \neq 0$,

$$S_{\beta} = \frac{\sqrt{h^{-1}}\left(R_2^{\beta} - \left(\overline{\sigma^2}\right)^{\beta}\right)}{\beta g'\left(\overline{\sigma^2};\beta\right)\sqrt{V}} \to^{d} N\left(0,1\right),$$

where $g'\left(\overline{\sigma^2};\beta\right)\sqrt{V}=\overline{\sigma^2}^{(\beta-1)}\sqrt{V}$. $\beta=1$ corresponds to the raw version (1).

The statistic S_β is infeasible because the scaling factor $\beta g'\left(\overline{\sigma^2};\beta\right)\sqrt{V}$ depends on the volatility path, which is unknown in practice. Therefore, we need to replace it with a consistent estimator given by $\beta g'\left(R_2;\beta\right)\sqrt{\hat{V}}$. The feasible statistic is then given by

$$T_\beta=\frac{\sqrt{h^{-1}}\left(g\left(R_2;\beta\right)-g\left(\overline{\sigma^2};\beta\right)\right)}{g'\left(R_2;\beta\right)\sqrt{\hat{V}}}=\begin{cases}\frac{\sqrt{h^{-1}}\left(R_2^\beta-\left(\overline{\sigma^2}\right)^\beta\right)}{\beta R_2^{\beta-1}\sqrt{\hat{V}}} & \text{when } \beta\neq 0\\ \frac{\sqrt{h^{-1}}\left(\log(R_2)-\log\left(\overline{\sigma^2}\right)\right)}{\sqrt{\frac{\hat{V}}{R_2^2}}} & \text{when } \beta=0.\end{cases}$$

Because $\beta g'\left(R_2;\beta\right)\sqrt{\hat{V}}$ is a consistent estimator of $\beta g'\left(\overline{\sigma^2};\beta\right)\sqrt{V}$ as $h\to 0$, $T_\beta\to^d N\left(0,1\right)$.

3.1 Infeasible statistic

We study the accuracy of the Box-Cox transformation for the infeasible statistic S_β. Our approach will be based on a second order Edgeworth expansion of the distribution of S_β. This expansion depends on the first three cumulants of S_β which we denote by $\kappa_i\left(S_\beta\right)$ for $i=1,2,3$. The following result (which follows from the results in GM (2008)) provides asymptotic expansions for $\kappa_i\left(S_\beta\right)$, $i=1,2,3$.[2]

Proposition 3.1 *Under Gonçalves and Meddahi's (2008) regularity conditions, conditional on σ, as $h\to 0$,*

a) $\kappa_1\left(S_\beta\right)=\sqrt{h}\left(\frac{1}{2}\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V}\right)+O\left(h\right)$.

b) $\kappa_2\left(S_\beta\right)=1+O\left(h\right)$.

c) $\kappa_3\left(S_\beta\right)=\sqrt{h}\left(\frac{4}{\sqrt{2}}\frac{\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^{3/2}}+3\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V}\right)+O\left(h\right)$.

Proposition 3.1 shows that β has an impact through order $O\left(\sqrt{h}\right)$ on the first and third cumulants of the statistic S_β. In particular, note that $\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}=\left(\beta-1\right)\left(\overline{\sigma^2}\right)^{-1}$. Thus, $\beta=1$ is the optimal choice if eliminating bias is the goal, i.e. the raw statistic performs best with respect to bias. This result confirms the Monte Carlo simulation results discussed previously. The log implies a negative bias for S_β, which is also in agreement with our simulation results. In general, applying a nonlinear transformation to R_2 introduces a bias term. The bias is positive if $\beta>1$ and it is negative if $\beta<1$.

The third order cumulant is also affected by the choice of β through the ratio $\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}$. Because the asymptotic normal approximation is associated with a zero third order cumulant, choosing β so as to make $\kappa_3\left(S_\beta\right)$ approximately equal to zero can induce better finite sample properties.

[2]The proof is based on Taylor expansions. Phillips (1979) and Niki and Konishi (1986) provide higher order expansions of the cumulants of a nonlinear transformation of an infeasible statistic in the univariate case whereas Marsh (2004) considers the multivariate case. The results in Proposition 3.1 can be obtained from Phillips (1979) and Niki and Konishi (1986) when applied to the realized volatility context. For the feasible statistics (to be discussed in the next section of this paper), a case by case analysis is necessary given that one has to take into account the uncertainty in the estimator of the variance of the statistic, which is case specific.

Corollary 3.1 *Conditionally on* σ*, as* $h \to 0$*, up to order* $O\left(\sqrt{h}\right)$,

a) *Choosing* β *equal to 1 implies* $\kappa_1(S_\beta) = 0$.

b) *Choosing* β *equal to* $\beta^* = 1 - \frac{2}{3}\frac{\overline{\sigma^2}\,\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^2} \leq \frac{1}{3}$ *implies* $\kappa_3(S_\beta) = 0$.

Corollary 3.1 shows that different values of β are needed to make bias and skewness equal to zero. In particular, the optimal value of β from the viewpoint of skewness is random and depends on the volatility path. If volatility is constant, then $\beta^* = \frac{1}{3}$. It corresponds to the so-called Wilson-Hilferty cube root transformation. If volatility is stochastic, the Cauchy-Schwartz inequality implies $\beta^* < \frac{1}{3}$. Recently, Chen and Deo (2004) proposed an optimal power transformation for random variables that are positive linear combinations of positive independent random variables and showed that this transformation induces normality in small samples. Their optimal transformation is equal to ours when applied to realized volatility for the infeasible statistic. Chen and Deo's (2004) results do not cover feasible statistics. As we will see in the next section, the optimal power transform β^* derived for the infeasible statistic S_β does not apply to the feasible statistic T_β.

In practice we cannot compute β^* because it depends on σ. It is nevertheless possible to consistently estimate β^* by replacing R_q/μ_q for $\overline{\sigma^q}$ (see BN-S (2004)). This yields the following estimator of β^*: $\hat{\beta}^* = 1 - \frac{2}{3}\frac{R_2\,(R_6/\mu_6)}{(R_4/\mu_4)^2} = 1 - \frac{2}{3}\frac{R_2\,(R_6/15)}{(R_4/3)^2}$, given that $\mu_4 = 3$ and $\mu_6 = 15$. Since β^* depends on the sixth order integrated power volatility, which is hard to estimate, our simulations show that estimating β^* creates some finite sample distortions. Confidence intervals based on $\hat{\beta}^*$ are nevertheless still more accurate than those based on the raw or the log statistic. Note that our Edgeworth expansions do not apply to the statistics computed with $\hat{\beta}^*$ because we have not taken into account the randomness in the estimates of β^*. However, our confidence intervals are first-order asymptotically valid.

The left-hand-side panels of Tables 7 and 8 contain some summary statistics for β^* and $\hat{\beta}^*$ for different sample sizes, for the GARCH(1,1) and for the two-factor diffusion, respectively. We complement these results by providing the kernel densities of $\frac{\overline{\sigma^2}\,\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^2}$ and $\frac{R_2\,(R_6/15)}{(R_4/3)^2}$ in Figure 1 for the GARCH(1,1) and for the two-factor diffusion models (these are sufficient statistics for β^* and $\hat{\beta}^*$ respectively).

For the GARCH(1,1) diffusion, Table 7 shows that on average β^* equals 0.331 and its distribution is very concentrated around this value, with a minimum of 0.310 and a maximum of 0.332. Thus, for the GARCH(1,1) the optimal value of the Box-Cox transformation in terms of skewness is on average very close to 1/3, which is the optimal value of β for the constant volatility case. This is not very surprising because the GARCH(1,1) diffusion implies very persistent sample paths for volatility. Estimation of β^* creates some finite sample distortions. Table 7 and Figure 1 show that $\hat{\beta}^*$ is very (positively) biased for finite samples. For instance, for $h^{-1} = 12$, the distribution of $\hat{\beta}^*$ is centered at 0.482, with a minimum value of 0.323. Its range only slightly intersects the range of β^*! The bias becomes less and less pronounced as we increase the sample size, as expected.

The results in Table 8 and Figure 1 (right-hand-side) suggest that the distribution of β^* for the two-factor diffusion is quite different from that of β^* for the GARCH(1,1) diffusion. On average, β^*

equals 0.057, but its range is now much larger, going from a minimum of -1.031 to a maximum of 0.297. As with the GARCH(1,1), for small samples $\hat{\beta}^*$ is severely biased towards positive values.

We complete this section by studying the accuracy of the normal approximation for the Box-Cox transformation for different values of β. We rely on a second order Edgeworth expansion of the distribution of S_β given by (see e.g. Hall, 1992, p. 47)

$$P\left(S_\beta \leq x\right) = \Phi\left(x\right) + \sqrt{h} p_\beta\left(x\right)\phi\left(x\right) + o\left(h\right),$$

where for any $x \in \mathbb{R}$, $\Phi\left(x\right)$ and $\phi\left(x\right)$ denote the cumulative distribution function and the density function of a standard normal random variable, respectively. The correction term $p_\beta\left(x\right)$ is defined as

$$p_\beta\left(x\right) = -\left(\bar{\kappa}_{1,\beta} + \frac{1}{6}\bar{\kappa}_{3,\beta}\left(x^2 - 1\right)\right),$$

where

$$\bar{\kappa}_{1,\beta} = \frac{1}{2}\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V} \quad \text{and} \quad \bar{\kappa}_{3,\beta} = \frac{4}{\sqrt{2}}\frac{\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^{3/2}} + 3\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V}$$

are the coefficients of the leading terms of $\kappa_1\left(S_\beta\right)$ and $\kappa_3\left(S_\beta\right)$, respectively.

Given this expansion, the error (conditional on σ) incurred by the normal approximation in estimating the distribution of S_β for a given choice of β is given by

$$\sup_{x\in\mathbb{R}}\left|P\left(S_\beta \leq x\right) - \Phi\left(x\right)\right| = \sqrt{h}\sup_{x\in\mathbb{R}}\left|p_\beta\left(x\right)\phi\left(x\right)\right| + o\left(h\right).$$

Thus, $\sup_{x\in\mathbb{R}}\left|p_\beta\left(x\right)\phi\left(x\right)\right|$ is the contribution of order $O\left(\sqrt{h}\right)$ to the normal error for the transformation indexed by β.

The following table shows the magnitude of $\bar{\kappa}_{1,\beta}$ and $\bar{\kappa}_{3,\beta}$ for several values of β when σ is constant.

Table 1. Coefficients of the correction term $p_\beta\left(x\right)$ when σ is constant

β	1	0	-1	$-1/3$	$1/3$
$\bar{\kappa}_{1,\beta}$	0	$-\frac{1}{2}\sqrt{2}$	$-\sqrt{2}$	$-\frac{2}{3}\sqrt{2}$	$-\frac{1}{3}\sqrt{2}$
$\bar{\kappa}_{3,\beta}$	$2\sqrt{2}$	$-\sqrt{2}$	$-4\sqrt{2}$	$-2\sqrt{2}$	0

Table 1 shows that for the case of constant volatility, the raw statistic is dominated by the log transformation from the viewpoint of skewness but not from the viewpoint of bias (as already discussed above). The log transformation clearly dominates the choices of $\beta = -1/3$ and $\beta = -1$. Choosing $\beta = 1/3$ (the optimal value according to skewness when σ is constant) also induces a bias smaller than the log transformation, so in this case $\beta = 0$ is dominated by the optimal choice β^*. The comparison between $\beta = 1$ and $\beta = \beta^*$ does not provide a clear ranking because $\beta = 1$ eliminates bias but not skewness whereas $\beta = \beta^*$ eliminates skewness but not bias.

For the GARCH(1,1) and the two-factor diffusion models, Tables 3 and 4 compare the finite sample value of the first and third order cumulants (in the tables, these are under the name of 'Mean' and 'Skewness', respectively) of the Box-Cox transformed statistics with the predictions of the asymptotic expansions for the first and third order cumulants given in Proposition 3.1 (reported as 'EE Mean' and

'EE Skewness', respectively. In particular, 'EE mean' is the average value of $\sqrt{h}\bar{\kappa}_{1,\beta}$ across the 100,000 Monte Carlo replications. 'EE Skewness' is the average value of $\sqrt{h}\bar{\kappa}_{3,\beta}$). For the GARCH(1,1), this comparison reveals generally close agreement between the two sets of results, which suggests that our expansions are good approximations to the true finite sample cumulants for the GARCH(1,1) model. The only exception is when $\beta = -1$, in which case the third order cumulant in finite samples is quite different from the Edgeworth expansion prediction, most especially when $h^{-1} = 12$. The same remark applies for the two-factor diffusion. For this model the finite sample quality of the asympotic expansions for the third order cumulants is inferior to the quality of these expansions for the GARCH(1,1) diffusion, at least for the smaller sample sizes. For instance, for $h^{-1} = 12$, the average sample value of $\sqrt{h}\bar{\kappa}_{3,\beta}$ when $\beta = \beta^*$ is zero (as predicted by our theory), but the finite sample value of the third order cumulant of S_{β^*} is -0.118. It reduces to -0.032 when $h^{-1} = 48$ and to 0.004 when $h^{-1} = 288$. For the larger sample sizes, the sample values of the first and third order cumulants agree with those predicted by the asymptotic expansions for the two-factor diffusion model.

The following result compares the accuracy of the normal approximation for several choices of β for the general case where volatility is not necessarily constant. As a measure of accuracy, we use the absolute value of the first higher order term in the Edgeworth expansion of the distribution of the statistic in question.

Proposition 3.2 *Conditionally on σ, as $h \to 0$, we have that for any β_1 and β_2 such that $\beta_2 < \beta_1 < \beta^* \leq 1/3$ and for any $x \neq 0$, $\left|p_{\beta^*}(x)\right| < \left|p_{\beta_1}(x)\right| < \left|p_{\beta_2}(x)\right|$ and $p_{\beta^*}(0) = p_{\beta_1}(0) = p_{\beta_2}(0)$.*

Proposition 3.2 shows that the log transform ($\beta = 0$) is dominated by the optimal Box-Cox transform when σ is constant ($\beta^* = \frac{1}{3}$). When σ is stochastic, the ranking is unclear as it depends on whether $\beta^* < 0$ and $\beta^* > 0$. If β^* is always nonnegative, Proposition 3.2 applies and we can conclude that β^* is an improvement over the log transform. Our simulations show that this is the case for the GARCH(1,1) diffusion, where $\beta^* > 0$ across the 100,000 Monte Carlo simulations. Thus, for this model we may conclude that the Box-Cox transform dominates the log transform in that the error of the normal approximation is smaller for the Box-Cox transform when β is chosen according to β^* for all values of $x \neq 0$.

For the two-factor diffusion although the average β^* is positive, it can take on negative values. In this case, we can show that the ranking between the log transform and the Box-Cox transform using β^* is not uniform in x, i.e. none of the transformations dominates uniformly the other. Similarly, for the infeasible statistic the ranking between the raw statistic and its log transform depends on the value of x, with none of them dominating the other. As Proposition 4.2 of GM (2008) shows (see Proposition 3.4 below for an extension of their result to other values of β) , this is not the case for the feasible statistic, where the log transform dominates the raw version for all values of $x \neq 0$.

To evaluate the finite sample quality of the normal approximation for the different transformations, Figure 2 presents the QQ plots for the Box-Cox transformed infeasible statistics for the GARCH(1,1) diffusion (left-hand-side panels) and for the two-factor diffusion (right-hand-side panels). The normal distribution is not a good approximation for the raw statistic ($\beta = 1$) when the sample size is small. The log transform ($\beta = 0$) dominates the raw statistic, but some distortions remain for $h^{-1} = 12$ and 48. The optimal Box-Cox transformation based on β^* induces a closer to normal QQ plot than the log transformation. The QQ plots of $\beta = 1/3$ and $\beta = \beta^*$ are almost identical, as expected since β^* is very concentrated around $\beta = 1/3$. The QQ plot for $\beta = \hat{\beta}^*$ shows some distortions when $h^{-1} = 12$. Thus, estimation of β^* induces finite sample distortions. For the two-factor diffusion, the log improves upon the raw version of realized volatility. The choice of β^* does not uniformly dominate the log transform. As for the GARCH(1,1) diffusion, estimation of β^* introduces some finite sample distortions.

To end this section we investigate by simulation the coverage probabilities of 95% level confidence intervals for $\overline{\sigma^2}$ across several values of β. Results for the GARCH(1,1) diffusion are presented in Figure 5 (one-sided intervals on the top panels and two-sided symmetric intervals on the bottom panels) whereas Figure 6 contains the corresponding results for the two-factor diffusion. Starting with the GARCH(1,1) diffusion, Figure 5 suggests that the log transform clearly improves upon the raw statistic for one-sided intervals but not for two-sided intervals, for which the raw statistic already performs very well (confirming previous results). It also suggests that other values of β different from zero may improve upon the log transformation. For instance, for one-sided intervals and for all sample sizes, $\beta = 1/3$ (or slightly above) clearly improves coverage with respect to choosing $\beta = 0$. To complement these results, note that Table 3 shows that for $h^{-1} = 12$ the coverage rates are 98.40 for $\beta = 1$, 91.00 for $\beta = 0$, 93.51 for $\beta = 1/3$, 93.50 for $\beta = \beta^*$ and 94.67 for $\beta = \hat{\beta}^*$. For the two-sided symmetric intervals, coverage rates are equal to 95.66 for $\beta = 1$, 93.60 for $\beta = 0$, 94.82 for $\beta = 1/3$, 94.81 for $\beta = \beta^*$ and 95.25 for $\beta = \hat{\beta}^*$, when $h^{-1} = 12$. Figure 5 shows that for this sample size $\beta = 0$ is dominated by β in the range 0 to 1 (or slightly above).

For the two-factor diffusion, choosing β between 0 and 1/3 induces coverage rates closer to 95% for one sided intervals (see top panel of Figure 6). For two-sided intervals, $\beta = 0$ seems to perform very well, with values of β larger than 0 but smaller than 4/3 producing too conservative intervals. $\beta = 4/3$ does as well as $\beta = 0$. Table 4 shows that in general $\beta = 0$ is dominated by β^* for one-sided intervals but not for two-sided intervals. Intervals based on $\hat{\beta}^*$ tend to overcover for $h^{-1} = 12$.

3.2 Feasible statistic

In this section we study the Box-Cox transformed feasible statistics T_β. Our goal is to compare the higher order properties of T_β for different choices of β in order to study the accuracy of the normal approximation across different values of β. We first analyze the properties of the first and third order cumulants of T_β. These are the main ingredients in the second-order Edgeworth expansions of the distribution of T_β. We let $\kappa_i(T_\beta)$ for $i = 1, 2, 3$, denote the first three cumulants of T_β. The following result provides asymptotic expansions for $\kappa_i(T_\beta)$, $i = 1, 2, 3$. It follows from an application of Theorem A.1 in GM (2008).

Proposition 3.3 *Conditionally on σ, as $h \to 0$,*

a) $\kappa_1 (T_\beta) = -\frac{1}{2}\sqrt{h}\left(\frac{4}{\sqrt{2}}\frac{\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^{3/2}} + \frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V}\right) + O(h).$

b) $\kappa_2 (T_\beta) = 1 + O(h).$

c) $\kappa_3 (T_\beta) = \sqrt{h}\left(-\frac{8}{\sqrt{2}}\frac{\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^{3/2}} - 3\frac{g''\left(\overline{\sigma^2};\beta\right)}{g'\left(\overline{\sigma^2};\beta\right)}\sqrt{V}\right) + O(h).$

Proposition 3.3 shows that the choice of β influences the higher order properties (through $O\left(\sqrt{h}\right)$) of the first and third order cumulants of T_β. Note that unlike in the infeasible case, the raw statistic ($\beta = 1$) has a negative bias equal to $-\frac{4}{2\sqrt{2}}\frac{\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^{3/2}}$. Thus, applying a nonlinear transformation to the feasible statistic has the potential to reduce bias *and* skewness. The following result states the optimal choices of β for approximately eliminating the bias and the skewness of T_β.

Corollary 3.2 *Conditionally on σ, as $h \to 0$, up to order $O\left(\sqrt{h}\right)$,*

a) *Choosing β equal to $\beta_* = 1 - 2\frac{\overline{\sigma^2}\,\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^2} \leq -1$ implies $\kappa_1 (T_\beta) = 0$.*

b) *Choosing β equal to $\beta^{**} = 1 - \frac{4}{3}\frac{\overline{\sigma^2}\,\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^2} \leq -\frac{1}{3}$ implies $\kappa_3 (T_\beta) = 0$.*

The optimal choices of β are random and depend on the volatility path. In general, different values of β are required to eliminate bias and skewness. Under constant volatility, these are equal to $\beta_* = -1$ and $\beta^{**} = -\frac{1}{3}$, respectively.

For the GARCH(1,1) diffusion, Table 7 shows that the average value of β^{**} is equal to -0.338 and thus very close to $-1/3$, the optimal value of β^{**} when volatility is constant. The distribution of β^{**} is also very concentrated around the mean value, with a range going from -0.381 up to -0.334. Similarly, the distribution of β_* is highly concentrated around its average value of -1.007, very close to the optimal value of β_* in the constant volatility case. As we remarked for the infeasible statistics, these results are not very surprising given that the GARCH(1,1) diffusion is highly persistent. For the two-factor diffusion, Table 8 shows that on average the value of β^{**} equals -0.886, with a distribution that is very dispersed (the minimum value is -3.061 and the maximum value is -0.407). Table 8 also shows that the average value of β_* is equal to -1.828, with a minimum value of -5.092 and a maximum value of -1.110. Thus, in this case the values of β^{**} and β_* are quite different from their optimal values under constant volatility (which equal $-1/3$ and -1 respectively). This is also not very surprising because the two-factor diffusion model is known to generate sample paths for prices and volatility that are very rugged, often close to those generated by a jump diffusion model.

The comparison between β^{**} and β_* shows that different values are required to eliminate bias and skewness for the feasible statistic. In addition, the skewness correction differs for the feasible and the infeasible statistics (compare β^* with β^{**}). The same remark applies to the bias correction, which corresponds to taking $\beta = 1$ for the infeasible statistic and to $\beta = \beta_*$ for the feasible statistic.

As discussed previously for β^*, β^{**} and β_* are unknown in practice because they depend on $\overline{\sigma^q}$ for $q = 2, 4, 6$. We can nevertheless estimate β^{**} and β_* consistently with the following estimators:

$$\hat{\beta}^{**} = 1 - \frac{4}{3}\frac{R_2 \ (R_6/\mu_6)}{(R_4/\mu_4)^2} = 1 - \frac{4}{3}\frac{R_2 \ (R_6/15)}{(R_4/3)^2}, \text{ and } \hat{\beta}_* = 1 - 2\frac{R_2 \ (R_6/15)}{(R_4/3)^2}.$$

Tables 7 and 8 (middle and right-hand-side panels) give the summary statistics (see also Figure 1 for the corresponding kernel densities). Overall, these results suggest that estimating β^{**} and β_* induces distortions in finite samples. The estimates are severely upward biased for the smallest sample sizes. The bias only becomes under control for $h^{-1} = 1152$, although for this sample size the variance is also larger than for the smaller sample sizes.

The log does not correspond to the optimal Box-Cox transformation from the viewpoint of bias and skewness. GM (2008) prove that the normal approximation is more accurate when applied to the log statistic as compared to the raw statistic. To end this section, we extend this result by providing a ranking of Box-Cox transforms across different values of β based on the accuracy of the normal approximation.

The second order Edgeworth expansion of the distribution of T_β is given by

$$P\left(T_\beta \leq x\right) = \Phi\left(x\right) + \sqrt{h} q_\beta\left(x\right)\phi\left(x\right) + o\left(h\right), \quad \text{with} \quad q_\beta\left(x\right) = -\left(\kappa_{1,\beta} + \frac{1}{6}\kappa_{3,\beta}\left(x^2 - 1\right)\right),$$

where $\kappa_{1,\beta}$ and $\kappa_{3,\beta}$ denote the coefficients of the leading terms (of order $O\left(\sqrt{h}\right)$) of the first and third order cumulants of T_β given in Proposition 3.3.

The following table shows the magnitude of $\kappa_{1,\beta}$ and $\kappa_{3,\beta}$ for several values of β when σ is constant.

Table 2. Coefficients of the correction term $q_\beta(x)$ when σ is constant

β	1	0	-1	$-1/3$	$1/3$
$\kappa_{1,\beta}$	$-\sqrt{2}$	$-\frac{1}{2}\sqrt{2}$	0	$-\frac{1}{3}\sqrt{2}$	$-\frac{2}{3}\sqrt{2}$
$\kappa_{3,\beta}$	$-4\sqrt{2}$	$-\sqrt{2}$	$2\sqrt{2}$	0	$-2\sqrt{2}$

When volatility is constant, Table 2 shows that the coefficients of the raw statistic ($\beta = 1$) are both larger (in absolute value) than those of the other transformations. In particular, the log transformation dominates $\beta = 1$ both in terms of skewness and bias. Choosing $\beta = \beta^{**} = -1/3$ eliminates skewness, as expected. Its bias is nonzero but it is smaller than the bias of both the log and the raw versions of realized volatility. $\beta = -1$ eliminates bias (as expected in the constant volatility case), but induces a larger amount of skewness compared to the log transform or the choice of $\beta = -1/3$. Its skewness is nevertheless smaller in absolute value than the skewness implicit in $\beta = 1$.

Similarly to the infeasible statistics, we can compare the scaled values of $\kappa_{1,\beta}$ and $\kappa_{3,\beta}$ (i.e. $\sqrt{h}\kappa_{i,\beta}$ for $i = 1, 3$) with the finite sample bias and skewness of the Box-Cox statistics given in Tables 5 and 6. The main overall feature of notice is that the cumulant expansions for the feasible statistics are in general less accurate than those for the infeasible statistics, especially when $h^{-1} = 12$. This is not surprising and it only confirms that the quality of the asymptotic theory (even to higher-order) is poorer for the feasible statistics than for the infeasible statistics. However, Table 5 shows that for the GARCH(1,1) diffusion, the first order cumulant expansion is quite accurate, even for the small sample sizes. The expansion for the third order cumulant is less accurate than for the first order cumulant, especially when $h^{-1} = 12$. It becomes reasonably accurate across the different values of β for the remaining values of h^{-1}. A comparison between β_* and β^{**} shows that β_* has smaller bias (close to zero) but larger skewness than β^{**} (whose skewness is close to zero), confirming our theoretical predictions. Estimating β^{**} induces an even larger bias compared to estimating β_* and it introduces more skewness than estimation of β_*. Thus, $\hat{\beta}^{**}$ has poorer finite sample properties than $\hat{\beta}_*$. As noted previously for $\hat{\beta}^*$, our Edgeworth expansions do not apply to the statistics computed with $\hat{\beta}^{**}$ or $\hat{\beta}_*$ because we have not taken into account the randomness in the estimates of β^{**} and β_*. The method delivers however valid confidence intervals. Comparing $\beta = -1$ with $\beta = \beta_*$ shows that these two transformations behave very similarly, which is as expected given that the GARCH(1,1) diffusion implies very persistent paths for volatility. The comparison between $\beta = -1$ and $\beta = \beta^{**}$ is similar to the comparison between $\beta = \beta_*$ and $\beta = \beta^{**}$, with $\beta = -1$ dominating in terms of bias and $\beta = \beta^{**}$ dominating in terms of skewness.

The distortions are larger for the two-factor diffusion than the GARCH(1,1) diffusion. For instance, when $h^{-1} = 12$, the sample third order cumulant of the raw statistic is equal to -78.445 whereas the cumulant expansion gives an average value of -3.365. For $h^{-1} = 48$, the values reduce to -1.609 and -1.683, respectively. Thus, the quality of the third order cumulant expansion greatly improves with this increase in the sample size when $\beta = 1$. For $\beta = 0$, not only is the amount of skewness much smaller in small samples but it is also much more in line with our cumulant expansion. Choosing $\beta = \beta^{**}$ reduces the amount of skewness relatively to the other transformations (an exception is $\beta = -1$, which has comparable, or even smaller, e.g. when $h^{-1} = 12$, skewness to β^{**}), but does not reduce it to zero, contrary to our theoretical predictions. Similarly, choosing $\beta = \beta_*$ reduces the amount of bias compared to the other transformations for all sample sizes but the smallest sample size (where $\beta = -1$ and β^{**} have smaller bias), but does not completely eliminate it. As for the GARCH(1,1) diffusion, the finite sample properties of $\hat{\beta}_*$ are superior to those of $\hat{\beta}^{**}$, with $\hat{\beta}_*$ showing a smaller finite sample mean, skewness and excess kurtosis compared to $\hat{\beta}^{**}$. $\beta = -1$ has smaller bias than β_{**} for all sample sizes and since its skewness is comparable to that of β_{**} in finite samples, $\beta = -1$ dominates β_{**}. Similarly, the sample skewness of $\beta = -1$ is much smaller than that of β_*, without a correspondingly large increase in bias ($\beta = -1$ has only slightly larger bias than β_*). Overall, we may conclude that $\beta = -1$ performs best in terms of both bias and skewness, dominating other transformations such as β_* (specifically tailored at bias reduction) and β^{**} (whose target is skewness reduction).

The following result compares the accuracy of the normal approximation for the Box-Cox transformation across several values of β. The result is general and does not assume that volatility is constant.

Proposition 3.4 *Conditionally on σ, as $h \to 0$, we have that for any β_1 and β_2 such that $\beta^{**} < \beta_1 < \beta_2$ and for any $x \neq 0$, $\left|q_{\beta^{**}}(x)\right| < \left|q_{\beta_1}(x)\right| < \left|q_{\beta_2}(x)\right|$, and $q_{\beta^{**}}(0) = q_{\beta_1}(0) = q_{\beta_2}(0)$.*

Proposition 3.4 proves that eliminating skewness helps increase the accuracy of the normal approximation. In particular, the error of the normal approximation (up to order $O\left(\sqrt{h}\right)$) is larger for any β_1 such that $\beta^{**} < \beta_1$, including the log transform ($\beta = 0$) and the raw statistic ($\beta = 1$). Proposition 3.4 also proves that the log transformation is an improvement over the raw statistic since $\left|q_{\beta_1}(x)\right| < \left|q_{\beta_2}(x)\right|$ when $\beta_1 = 0$ and $\beta_2 = 1$, and $\beta_1 > \beta^{**}$. Thus, Proposition 3.4 includes Proposition 4.2 of GM (2008) as a special case. Proposition 3.4 does not allow a comparison between β^{**} and β_* since $\beta_* \leq \beta^{**}$. In this case, we can show that there is no uniform (in x) ranking between the two choices.

The QQ plots for the feasible statistics are presented in Figures 3 (GARCH(1,1)) and 4 (two-factor). For the GARCH(1,1) model, the QQ plots confirm the ranking suggested by Proposition 3.4. These plots show that $\beta = \beta^{**}$ dominates the log transform with $\beta = 0$, which is better than $\beta = 1/3$, which in turn dominates the raw transform with $\beta = 1$. The feasible transformation based on $\hat{\beta}^{**}$ induces more finite sample distortions than the infeasible transformation, especially for the smaller sample sizes. The QQ plots for $\hat{\beta}^{**}$ are close to those for the log. Choosing $\beta = -1$ and $\beta = \beta_*$ is better than choosing $\beta = 1$, but the QQ plots for $\beta = -1$ and β_* show more distortions when compared to the other Box-Cox transforms, especially when $h^{-1} = 12$. Interestingly, the QQ plot for $\hat{\beta}_*$ shows small sample distortions.

The two-factor diffusion is associated with larger distortions than the GARCH(1,1) diffusion. This is evident when we compare the scales of the corresponding QQ plots. The log dominates $\beta = 1$. The QQ plots for $\beta = 0$ and $\beta = \beta^{**}$ are the reverse image of each other, with $\beta = 0$ dominating $\beta = \beta^{**}$ on the right-hand side of the distribution and being dominated by $\beta = \beta^{**}$ on the left-hand side. Although Proposition 3.4 does not cover the case where $\beta = -1$, the QQ plots suggest that this transformation dominates the log on the left-hand side but does worse at the right-hand side. Choosing $\beta = \beta_*$ induces larger distortions than choosing $\beta = \beta_{**}$ or $\beta = -1$, but estimation of β^{**} creates larger deviations from the normal approximation than estimation of β_* (which is very well behaved).

To end this section, we compare the Box-Cox transformations for realized volatility in terms of coverage probabilities for both one-sided and two-sided symmetric intervals for $\overline{\sigma^2}$. Our analysis is similar to that followed for the infeasible statistics, but now we concentrate on the right-hand-side panels of Figures 5 and 6, which refer to the feasible statistics. Starting with the one-sided intervals for the GARCH(1,1) diffusion, Figure 5 shows that choosing $\beta = 0$ improves upon $\beta = 1$ for all sample sizes, but $\beta = 0$ does not correspond to the best possible choice in terms of coverage accuracy. Indeed, any value of β between -1 and 0 dominates $\beta = 0$, with the optimal choice lying somewhere close to -1. For one-sided intervals, Table 5 shows that for $\beta = -1$ the coverage rate is equal to 96.07, 95.74, 95.38, 95.19 for $h^{-1} = 12$, 48, 288 and 1152, respectively. In contrast, the log transformation has coverage rates equal to 88.54, 92.36, 94.17 and 94.55. These rates are systematically worse than those for $\beta = -1$ (which are similar to those for β_*), and they are also worse than those for β^{**} (which are equal to 90.94, 93.51, 94.59 and 94.77). Estimation of β^{**} and β_* induces larger undercoverages, but $\hat{\beta}^{**}$ and $\hat{\beta}_*$ are still preferred to $\beta = 0$, with $\hat{\beta}_*$ dominating $\hat{\beta}^{**}$ (as we saw previously, the finite sample properties of $\hat{\beta}_*$ in terms of bias, skewness and kurtosis are better than those of $\hat{\beta}^{**}$). The results for the two-sided symmetric intervals are qualitatively similar to those for the one-sided intervals, with the main difference being that two-sided intervals are associated with smaller coverage distortions for almost all transformations and for all sample sizes. The exception is when $\beta = -1$ and $\beta = \beta_*$, which are associated with a slight overcoverage for one-sided intervals and with undercoverage for two-sided intervals. Nevertheless, Figure 5 shows that $\beta = -1$ is the choice of β that produces more accurate two-sided intervals for $\overline{\sigma^2}$. The difference in coverage probability with respect to other choices of β (including $\beta = 0$) is especially important for $h^{-1} = 12$. The results for the two-factor diffusion model are qualitatively similar to those for the GARCH(1,1) diffusion. The main differences are that the two-factor diffusion has larger distortions for all methods and all sample sizes (the lines in Figure 6 shift downwards in comparison to Figure 5). Choosing the optimal value of β^{**} is an improvement over choosing $\beta = 0$, which dominates $\beta = 1/3$, which dominates $\beta = 1$. Choosing $\beta = \hat{\beta}^{**}$ outperforms $\beta = 0$, but underperforms $\beta = \beta^{**}$. The comparison between $\beta = \beta_*$ and $\beta = \beta^{**}$ favors β^{**}, suggesting that for the two factor diffusion model the skewness correction is preferred over the bias correction. One possible explanation is the fact that β_* induces a smaller bias but at the cost of introducing too large a skewness (β^{**} in contrast reduces skewness without introducing too large a bias). Figure 6 suggests that the optimal choice of β in terms of coverage probability control for one-sided and two-sided intervals is slightly above -1 for the two-factor diffusion model. Overall, a choice of $\beta = -1$ appears to do best if we want to produce intervals for $\overline{\sigma^2}$ based on the Box-Cox transform with good coverage rates. In particular, this choice dominates the optimal choices β^{**}, which is based solely on eliminating the finite sample skewness, and β_*, which is tailored to eliminate bias. As we discussed before, $\beta = -1$ performs best if both bias and skewness are taken into account simultaneously. Controlling bias *and* skewness are both important if good coverage accuracy is the goal.

4 Conclusion

The log transformation is often preferred to the raw version of realized volatility because of its superior finite sample properties (including finite sample skewness). GM (2008) provide a theoretical explanation for this finding.

The fact that the log transformation improves upon the raw statistic does not imply that the log transformation is the best possible transformation available. The main contribution of this paper is to consider a broader class of analytical transformations (indexed by the parameter β) that includes the log ($\beta = 0$) and the raw ($\beta = 1$) versions of realized volatility as special cases. This is the Box-Cox transformation. Based on higher order expansions of the cumulants of the Box-Cox transformed statistics, we first derive the optimal values of β that induce zero bias and zero skewness. The values of β that eliminate bias are generally different from those that eliminate skewness. These values are also different depending on whether we consider infeasible statistics (which are constructed under the assumption that their asymptotic variances are known) or on whether we consider feasible statistics (which are the statistics used in practice). Therefore in this paper we treat the two cases separately. For each case, we rely on Edgeworth expansions of the distribution of the Box-Cox transforms to study the accuracy of these transforms across different values of β. We also provide extensive Monte Carlo simulation to investigate how the different Box-Cox transforms behave in finite samples. Our general conclusion is that there exist other Box-Cox transformations that dominate the log transformation. In particular, the optimal but infeasible transformations derived in this paper generally outperform the log. Their feasible versions are less well behaved than their infeasible versions in finite samples and are comparable to the log transform. However, if coverage probability accuracy is the goal, choosing $\beta = -1$ appears as the best choice across the two models studied in this paper.

The results in this paper show that confidence intervals for integrated volatility based on an appropriately chosen Box-Cox transform have smaller finite sample distortions than intervals based on the raw or the log versions of realized volatility. The Box-Cox transform intervals are nevertheless still based on the asymptotic normal approximation. We could use the bootstrap to further improve upon the first order asymptotic theory derived for our class of statistics. GM (2008) follow this approach for $\beta = 0$ and $\beta = 1$. Their results show that the bootstrap allows a further improvement in accuracy when applied to these statistics.

Table 3. Infeasible R_2-based statistics, GARCH(1,1) diffusion

h^{-1}		$\beta = 1$ (Raw)	$\beta = 0$ (Log)	$\beta = -1$	$\beta = 1/3$	β^*	$\hat{\beta}^*$
12	Mean	0.000	−0.210	−0.490	−0.136	−0.136	−0.105
	EE Mean	0.000	−0.204	−0.409	−0.136	−0.137	−0.106
	Skewness	0.824	−0.489	−8.170	0.009	0.006	0.186
	EE Skewness	0.821	−0.406	−1.663	0.003	0.000	0.185
	St. Error	0.999	1.042	1.471	0.998	0.998	0.989
	Ex. Kurtosis	1.064	0.387	14.419	−0.016	−0.015	0.030
	Cov. one-sided 95%	98.40	91.00	84.63	93.51	93.50	94.67
	Cov. two-sided 95%	95.66	93.60	87.91	94.82	94.81	95.25
48	Mean	0.002	−0.101	−0.211	−0.066	−0.066	−0.059
	EE Mean	0.000	−0.102	−0.204	−0.068	−0.068	−0.061
	Skewness	0.392	−0.224	−1.164	−0.010	−0.012	0.034
	EE Skewness	0.410	−0.203	−0.817	0.001	0.000	0.046
	St. Error	0.999	1.010	1.091	1.000	1.000	0.998
	Ex. Kurtosis	0.210	0.074	1.503	−0.026	−0.026	−0.023
	Cov. one-sided 95%	96.48	93.09	90.03	94.28	94.27	94.52
	Cov. two-sided 95%	95.27	94.69	92.90	95.06	95.06	95.10
288	Mean	−0.001	−0.043	−0.085	−0.029	−0.029	−0.028
	EE Mean	0.000	−0.042	−0.083	−0.028	−0.028	−0.027
	Skewness	0.160	−0.091	−0.361	−0.006	−0.007	−0.002
	EE Skewness	0.168	−0.083	−0.333	0.001	0.000	0.005
	Standard Error	0.999	1.001	1.014	0.999	0.999	0.999
	Excess Kurtosis	0.039	0.023	0.236	0.004	0.004	0.003
	Cov. one-sided 95%	95.55	94.30	93.06	94.73	94.72	94.75
	Cov. two-sided 95%	95.02	94.88	94.65	94.98	94.98	94.98
1152	Mean	0.003	−0.018	−0.039	−0.011	−0.011	−0.011
	EE Mean	0.000	−0.021	−0.042	−0.014	−0.014	−0.014
	Skewness	0.077	−0.048	−0.177	−0.006	−0.007	−0.006
	EE Skewness	0.084	−0.041	−0.167	0.000	0.000	0.001
	Standard Error	1.002	1.002	1.005	1.002	1.002	1.002
	Excess Kurtosis	−0.001	−0.003	0.051	−0.008	−0.008	−0.008
	Cov. one-sided 95%	95.22	94.62	93.99	94.81	94.81	94.81
	Cov. two-sided 95%	94.99	94.98	94.93	94.99	94.99	94.99

Notes: 100,000 Monte Carlo replications. EE Mean and EE skewness denote the Egdeworth corrections for the mean and skewness. The feasible statistic is $\sqrt{h^{-1}}\left(R_2^{\beta} - \left(\overline{\sigma^2}\right)^{\beta}\right)/\beta R_2^{(\beta-1)}\sqrt{\hat{V}}$, where $\hat{V} = \frac{2}{3}R_4$.

Table 4. Infeasible R_2-based statistics, two-factor diffusion

h^{-1}		$\beta = 1$ (Raw)	$\beta = 0$ (Log)	$\beta = -1$	$\beta = 1/3$	β^*	$\hat{\beta}^*$
12	Mean	0.000	−0.244	−0.603	−0.158	−0.230	−0.121
	EE Mean	0.000	−0.294	−0.588	−0.196	−0.280	−0.152
	Skewness	1.183	−0.206	−11.646	0.205	−0.118	0.384
	EE Skewness	1.683	−0.082	−1.846	0.507	0.000	0.771
	Standard Error	0.943	0.938	1.524	0.896	0.929	0.896
	Excess Kurtosis	3.381	0.168	26.234	0.108	0.099	0.474
	Cov. one-sided 95%	99.89	92.82	83.52	96.27	93.39	97.53
	Cov. two-sided 95%	95.88	95.60	87.03	97.16	95.88	97.25
48	Mean	0.002	−0.135	−0.288	−0.089	−0.129	−0.083
	EE Mean	0.000	−0.147	−0.294	−0.098	−0.140	−0.092
	Skewness	0.761	−0.073	−1.335	0.183	−0.032	0.219
	EE Skewness	0.841	−0.041	−0.923	0.253	0.000	0.289
	Standard Error	0.985	0.975	1.095	0.965	0.973	0.965
	Excess Kurtosis	1.134	0.026	1.980	0.081	0.024	0.123
	Cov. one-sided 95%	98.21	93.73	89.16	95.31	93.95	95.50
	Cov. two-sided 95%	95.67	95.25	92.38	95.78	95.38	95.81
288	Mean	0.003	−0.056	−0.116	−0.036	−0.053	−0.043
	EE Mean	0.000	−0.060	−0.120	−0.040	−0.057	−0.047
	Skewness	0.349	−0.013	−0.393	0.105	0.004	0.064
	EE Skewness	0.343	−0.017	−0.377	0.103	0.000	0.061
	Standard Error	0.998	0.994	1.013	0.993	0.994	0.991
	Excess Kurtosis	0.257	0.018	0.285	0.041	0.016	0.011
	Cov. one-sided 95%	96.30	94.50	92.69	95.12	94.58	94.97
	Cov. two-sided 95%	95.18	95.17	94.61	95.22	95.17	95.26
1152	Mean	0.008	−0.022	−0.052	−0.012	−0.021	−0.018
	EE Mean	0.000	−0.030	−0.060	−0.020	−0.029	−0.026
	Skewness	0.152	−0.029	−0.214	0.031	−0.020	−0.006
	EE Skewness	0.172	−0.008	−0.188	0.052	0.000	0.013
	Standard Error	0.999	0.999	1.004	0.998	0.999	0.997
	Excess Kurtosis	0.057	0.019	0.107	0.018	0.018	0.008
	Cov. one-sided 95%	95.64	94.76	93.92	95.05	94.79	94.89
	Cov. two-sided 95%	95.01	95.00	94.86	95.02	94.99	95.04

Notes: 100,000 Monte Carlo replications. EE Mean and EE skewness denote the Egdeworth corrections for the mean and skewness. The infeasible statistic is $\sqrt{h^{-1}}\left(R_2^{\beta} - \left(\overline{\sigma^2}\right)^{\beta}\right)/\beta\left(\overline{\sigma^2}\right)^{\beta-1}\sqrt{V}$.

Table 5. Feasible RV-based statistics, GARCH(1,1) diffusion

h^{-1}		$\beta=1$	$\beta=0$	$\beta=-1$	$\beta=1/3$	β^{**}	$\hat{\beta}^{**}$	β_*	$\hat{\beta}_*$
12	Mean	−0.548	−0.235	0.000	−0.325	−0.151	−0.227	0.001	−0.103
	EE Mean	−0.410	−0.206	−0.001	−0.274	−0.137	−0.198	0.000	−0.092
	Skewness	−11.944	−0.755	1.198	−2.028	−0.004	−0.730	1.212	0.320
	EE Skewness	−1.641	−0.414	0.812	−0.823	0.000	−0.370	0.821	0.266
	Standard Error	1.658	1.175	1.125	1.269	1.124	1.172	1.126	1.115
	Excess Kurtosis	14.831	0.621	1.293	2.105	0.180	0.663	1.315	0.309
	Cov. one-sided 95%	82.55	88.54	96.07	86.32	90.94	88.83	96.12	92.60
	Cov. two-sided 95%	85.87	90.33	93.04	89.07	91.50	90.46	93.05	92.03
48	Mean	−0.219	−0.105	0.001	−0.142	−0.069	−0.086	0.002	−0.023
	EE Mean	−0.205	−0.103	−0.001	−0.137	−0.068	−0.084	0.000	−0.023
	Skewness	−1.335	−0.259	0.450	−0.540	−0.011	−0.144	0.454	0.262
	EE Skewness	−0.821	−0.207	0.406	−0.412	0.000	−0.093	0.410	0.271
	Standard Error	1.134	1.051	1.039	1.070	1.039	1.045	1.039	1.038
	Excess Kurtosis	1.638	0.178	0.335	0.448	0.077	0.131	0.341	0.178
	Cov. one-sided 95%	89.38	92.36	95.74	91.32	93.51	92.97	95.76	94.94
	Cov. two-sided 95%	92.05	93.64	94.29	93.24	93.97	93.79	94.29	94.18
288	Mean	−0.086	−0.043	−0.001	−0.057	−0.029	−0.031	−0.001	−0.004
	EE Mean	−0.084	−0.042	0.000	−0.056	−0.028	−0.030	0.000	−0.002
	Skewness	−0.374	−0.096	0.162	−0.185	−0.008	−0.022	0.163	0.143
	EE Skewness	−0.335	−0.085	0.166	−0.168	0.000	−0.010	0.168	0.153
	Standard Error	1.022	1.009	1.007	1.012	1.007	1.007	1.007	1.007
	Excess Kurtosis	0.260	0.043	0.056	0.088	0.022	0.027	0.057	0.053
	Cov. one-sided 95%	92.93	94.17	95.38	93.76	94.59	94.53	95.39	95.31
	Cov. two-sided 95%	94.44	94.69	94.83	94.61	94.77	94.77	94.83	94.80
1152	Mean	−0.040	−0.019	0.002	−0.026	−0.012	−0.012	0.003	0.002
	EE Mean	−0.042	−0.021	0.000	−0.028	−0.014	−0.014	0.000	0.000
	Skewness	−0.179	−0.050	0.077	−0.092	−0.007	−0.009	0.078	0.075
	EE Skewness	−0.168	−0.042	0.083	−0.084	0.000	−0.001	0.084	0.082
	Standard Error	1.007	1.004	1.004	1.005	1.004	1.004	1.004	1.004
	Excess Kurtosis	0.059	0.004	0.006	0.016	−0.002	−0.001	0.006	0.006
	Cov. one-sided 95%	93.96	94.55	95.19	94.35	94.77	94.76	95.19	95.17
	Cov. two-sided 95%	94.86	94.91	94.96	94.90	94.92	94.93	94.96	94.94

Notes: 100,000 Monte Carlo replications. EE Mean and EE skewness denote the Egdeworth corrections for the mean and skewness. The feasible statistic is $\sqrt{h^{-1}}\left(R_2^{\beta}-\left(\overline{\sigma^2}\right)^{\beta}\right)/\beta R_2^{(\beta-1)}\sqrt{\hat{V}}$, where $\hat{V}=\frac{2}{3}R_4$.

Table 6. Feasible RV-based statistics, two-factor diffusion

h^{-1}		$\beta = 1$	$\beta = 0$	$\beta = -1$	$\beta = 1/3$	β^{**}	$\hat{\beta}^{**}$	β_*	$\hat{\beta}_*$
12	Mean	−0.972	−0.428	−0.085	−0.572	−0.115	−0.417	0.218	−0.234
	EE Mean	−0.841	−0.547	−0.253	−0.645	−0.280	−0.280	0.000	−0.385
	Skewness	−78.445	−1.709	1.864	−5.231	2.001	−1.687	88.892	0.272
	EE Skewness	−3.365	−1.601	0.163	−2.189	0.000	−1.541	1.683	−0.629
	Standard Error	2.475	1.348	1.236	1.540	1.247	1.339	1.939	1.216
	Excess Kurtosis	73.102	1.020	1.849	4.373	3.785	1.227	462.75	0.204
	Cov. one-sided 95%	75.74	82.86	93.31	80.17	92.12	83.27	99.42	88.43
	Cov. two-sided 95%	79.45	85.74	90.95	83.89	90.45	86.12	89.86	88.89
48	Mean	−0.421	−0.235	−0.074	−0.293	−0.088	−0.196	0.065	−0.097
	EE Mean	−0.421	−0.274	−0.127	−0.323	−0.140	−0.140	0.000	−0.144
	Skewness	−1.609	−0.504	0.385	−0.825	0.341	−0.329	4.778	0.284
	EE Skewness	−1.683	−0.801	0.082	−1.095	0.000	−0.578	0.841	−0.025
	Standard Error	1.373	1.158	1.111	1.207	1.112	1.138	1.225	1.114
	Excess Kurtosis	5.249	0.586	0.444	1.382	0.595	0.395	54.774	0.290
	Cov. one-sided 95%	84.56	88.62	93.32	87.14	92.93	89.77	97.23	92.65
	Cov. two-sided 95%	88.01	90.85	92.41	90.07	92.34	91.31	92.58	92.14
288	Mean	−0.166	−0.102	−0.040	−0.123	−0.046	−0.068	0.013	−0.020
	EE Mean	−0.172	−0.112	−0.052	−0.132	−0.057	−0.057	0.000	−0.030
	Skewness	−0.658	−0.307	0.109	−0.461	0.076	−0.095	0.504	0.227
	EE Skewness	−0.687	−0.327	0.033	−0.447	0.000	−0.121	0.343	0.162
	Standard Error	1.078	1.043	1.032	1.052	1.032	1.039	1.044	1.041
	Excess Kurtosis	0.910	0.226	0.125	0.381	0.111	0.145	0.583	0.244
	Cov. one-sided 95%	90.84	92.57	94.30	92.01	94.13	93.46	95.83	94.81
	Cov. two-sided 95%	93.22	93.89	94.22	93.70	94.21	93.99	94.24	94.04
1152	Mean	−0.078	−0.047	−0.017	−0.057	−0.019	−0.024	0.009	0.002
	EE Mean	−0.086	−0.056	−0.026	−0.066	−0.029	−0.029	0.000	−0.007
	Skewness	−0.391	−0.184	0.007	−0.251	−0.010	−0.049	0.171	0.117
	EE Skewness	−0.343	−0.163	0.017	−0.223	0.000	−0.027	0.172	0.132
	Standard Error	1.024	1.014	1.011	1.017	1.010	1.014	1.012	1.015
	Excess Kurtosis	0.311	0.112	0.050	0.163	0.046	0.072	0.103	0.102
	Cov. one-sided 95%	93.02	93.87	94.77	93.58	94.70	94.51	95.45	95.18
	Cov. two-sided 95%	94.49	94.65	94.70	94.57	94.70	94.61	94.72	94.59

Notes: 100,000 Monte Carlo replications. EE Mean and EE skewness denote the Egdeworth corrections for the mean and skewness. The feasible statistic is $\sqrt{h^{-1}}\left(R_2^{\beta} - \overline{\sigma^2}^{\beta}\right)/\beta R_2^{(\beta-1)}\sqrt{\hat{V}}$, where $\hat{V} = \frac{2}{3}R_4$.

Table 7. Summary statistics for the infeasible and feasible versions of the optimal power

GARCH(1,1) diffusion

		Infeasible statistic		Feasible statistic			
h^{-1}	Statistics	β^*	$\hat{\beta}^*$	β^{**}	$\hat{\beta}^{**}$	β_*	$\hat{\beta}_*$
12	Mean	0.331	0.482	−0.338	−0.036	−1.007	−0.554
	St. Error	0.002	0.045	0.004	0.089	0.006	0.134
	Max	0.333	0.591	−0.334	0.182	−1.004	−0.226
	Min	0.310	0.323	−0.381	−0.354	−1.071	−1.032
48	Mean	0.331	0.407	−0.338	−0.187	−1.007	−0.780
	St. Error	0.002	0.069	0.004	0.138	0.006	0.206
	Max	0.333	0.542	−0.334	0.083	−1.004	−0.376
	Min	0.310	0.061	−0.381	−0.877	−1.071	−1.816
288	Mean	0.331	0.351	−0.338	−0.298	−1.007	−0.947
	St. Error	0.002	0.067	0.004	0.133	0.006	0.200
	Max	0.333	0.477	−0.334	−0.046	−1.004	−0.569
	Min	0.310	−0.409	−0.381	−1.817	−1.071	−3.226
1152	Mean	0.331	0.336	−0.338	−0.327	−1.007	−0.991
	St. Error	0.002	0.045	0.004	0.089	0.006	0.134
	Max	0.333	0.439	−0.334	−0.122	−1.004	−0.684
	Min	0.310	−0.344	−0.381	−1.688	−1.071	−3.032

Notes: 100,000 Monte Carlo replications.

Table 8. Summary statistics for the infeasible and feasible versions of the optimal power

Two-factor diffusion

		Infeasible statistic		Feasible statistic			
h^{-1}	Statistics	β^*	$\hat{\beta}^*$	β^{**}	$\hat{\beta}^{**}$	β_*	$\hat{\beta}_*$
12	Mean	0.057	0.482	−0.886	−0.036	−1.828	−0.554
	St. Error	0.120	0.043	0.240	0.086	0.359	0.129
	Max	0.297	0.596	−0.407	0.192	−1.110	−0.211
	Min	−1.031	0.325	−3.061	−0.350	−5.092	−1.025
48	Mean	0.057	0.373	−0.886	−0.254	−1.828	−0.882
	St. Error	0.120	0.076	0.240	0.152	0.359	0.229
	Max	0.297	0.561	−0.407	0.122	−1.110	−0.318
	Min	−1.031	0.038	−3.061	−0.925	−5.092	−1.887
288	Mean	0.057	0.216	−0.886	−0.567	−1.828	−1.351
	St. Error	0.120	0.136	0.240	0.271	0.359	0.407
	Max	0.297	0.464	−0.407	−0.072	−1.110	−0.608
	Min	−1.031	−0.615	−3.061	−2.230	−5.092	−3.845
1152	Mean	0.057	0.126	−0.886	−0.748	−1.828	−1.622
	St. Error	0.120	0.164	0.240	0.327	0.359	0.491
	Max	0.297	0.412	−0.407	−0.177	−1.110	−0.765
	Min	−1.031	−1.458	−3.061	−3.916	−5.092	−6.373

Notes: 100,000 Monte Carlo replications.

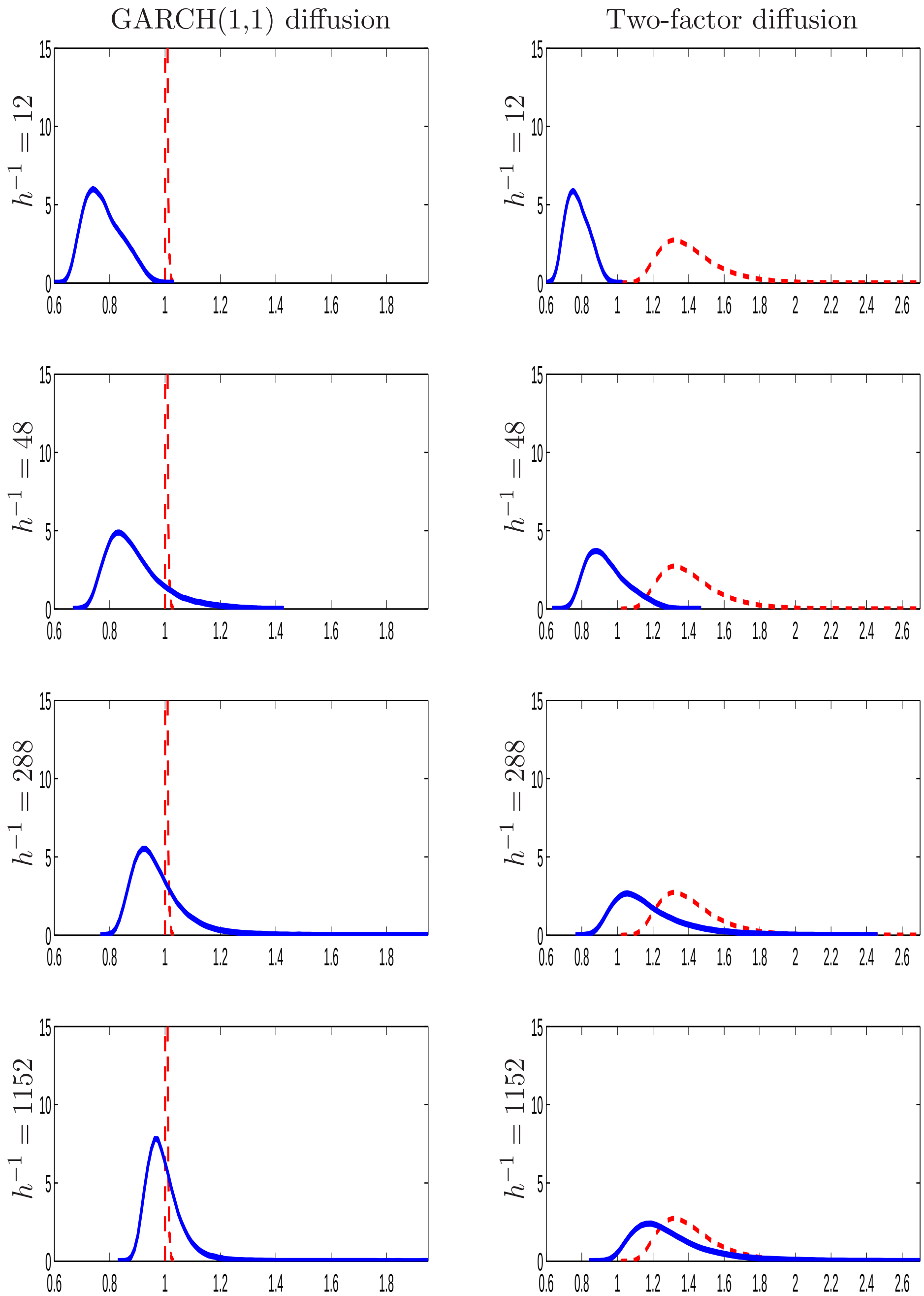

Figure 1: The left-hand-side panel shows the kernel densities of the infeasible and feasible power statistics for the GARCH(1,1) diffusion. The right-hand-side panel shows these statistics for the two-factor diffusion. The dashed line refers to $\frac{\overline{\sigma^2}\,\overline{\sigma^6}}{\left(\overline{\sigma^4}\right)^2}$ and the solid line refers to its realized version $\frac{RV\ (R_6/15)}{(R_4/3)^2}$.

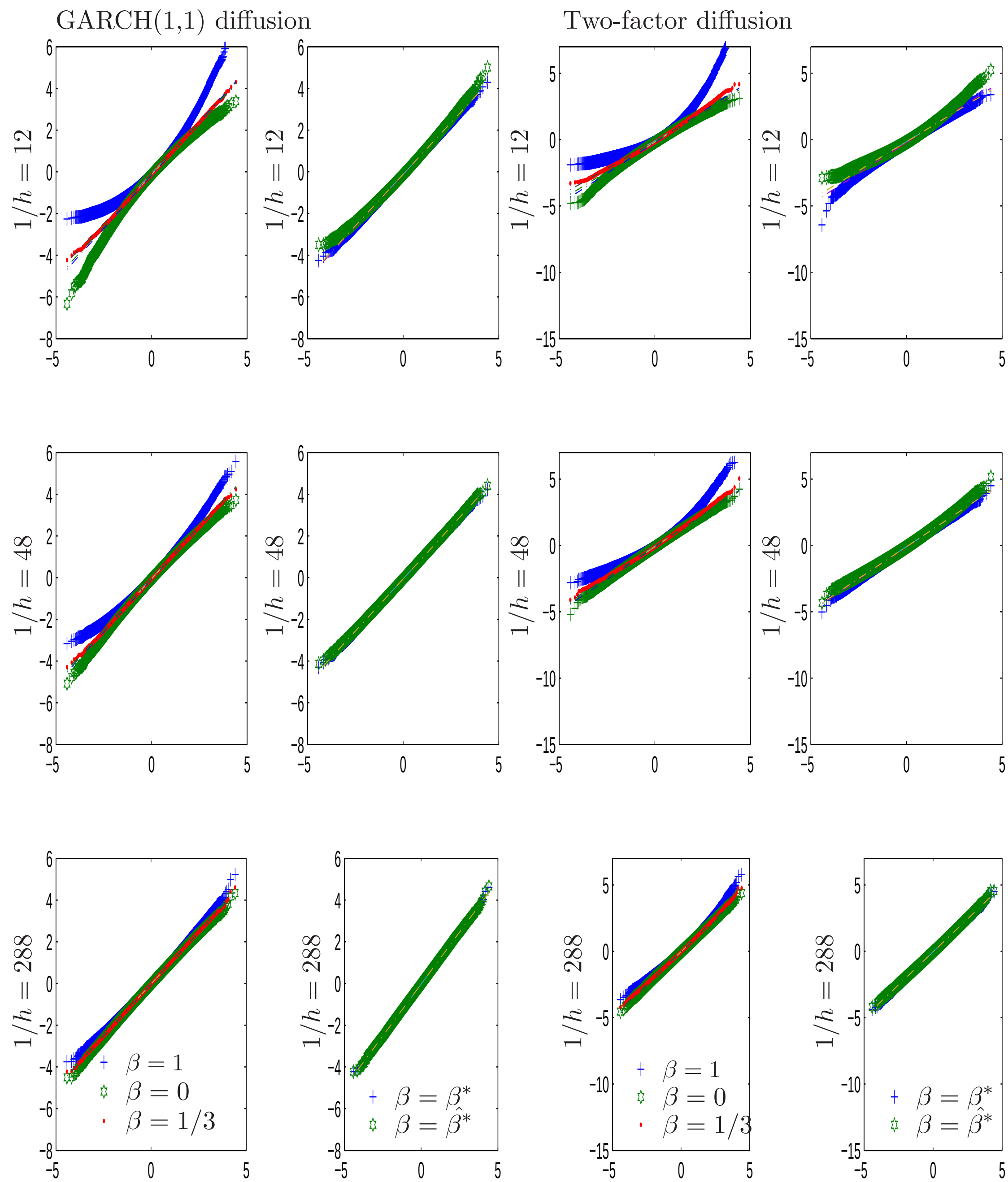

Figure 2: The left-hand-side panels give the QQ-plots for the infeasible statistics for the GARCH(1,1) diffusion when $\beta = 1, \beta = 0$ and $\beta = 1/3$, and when $\beta = \beta^*$, $\beta = \hat{\beta}^*$, respectively. The right-hand-side panels give the QQ-plots for the two-factor diffusion.

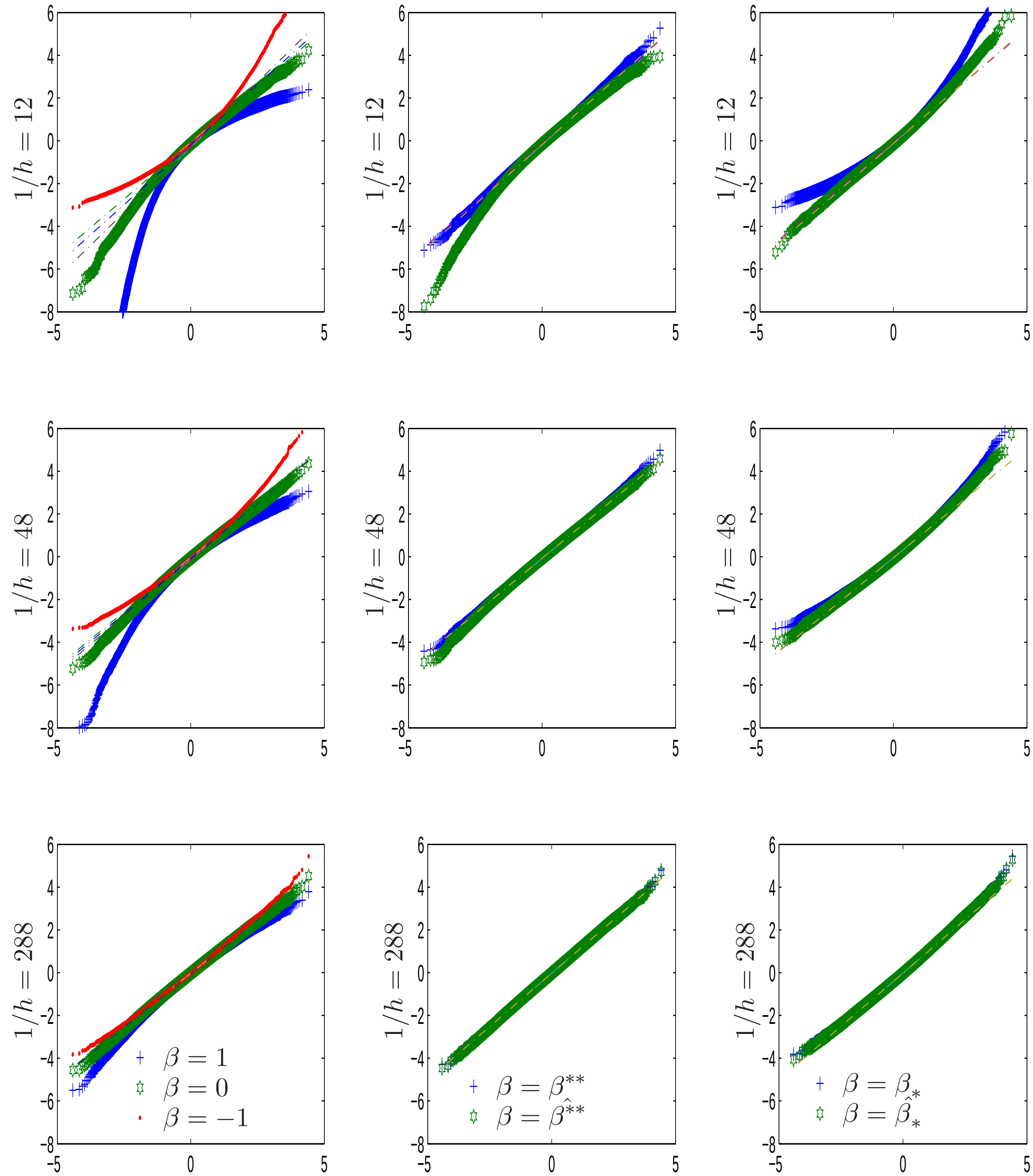

Figure 3: QQ-plots for feasible statistics for the GARCH(1,1) diffusion.

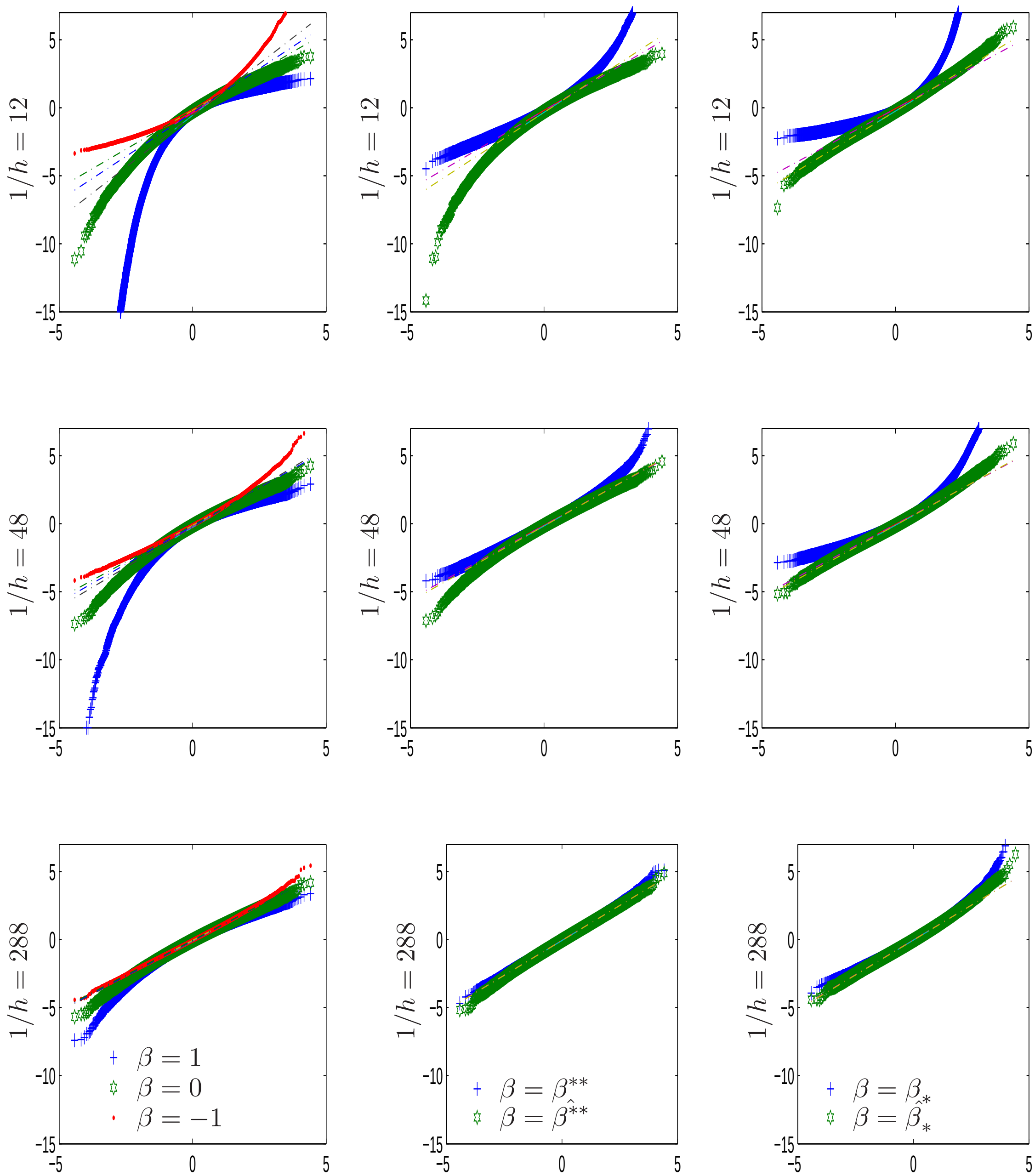

Figure 4: QQ-plots for feasible statistics for the two-factor diffusion.

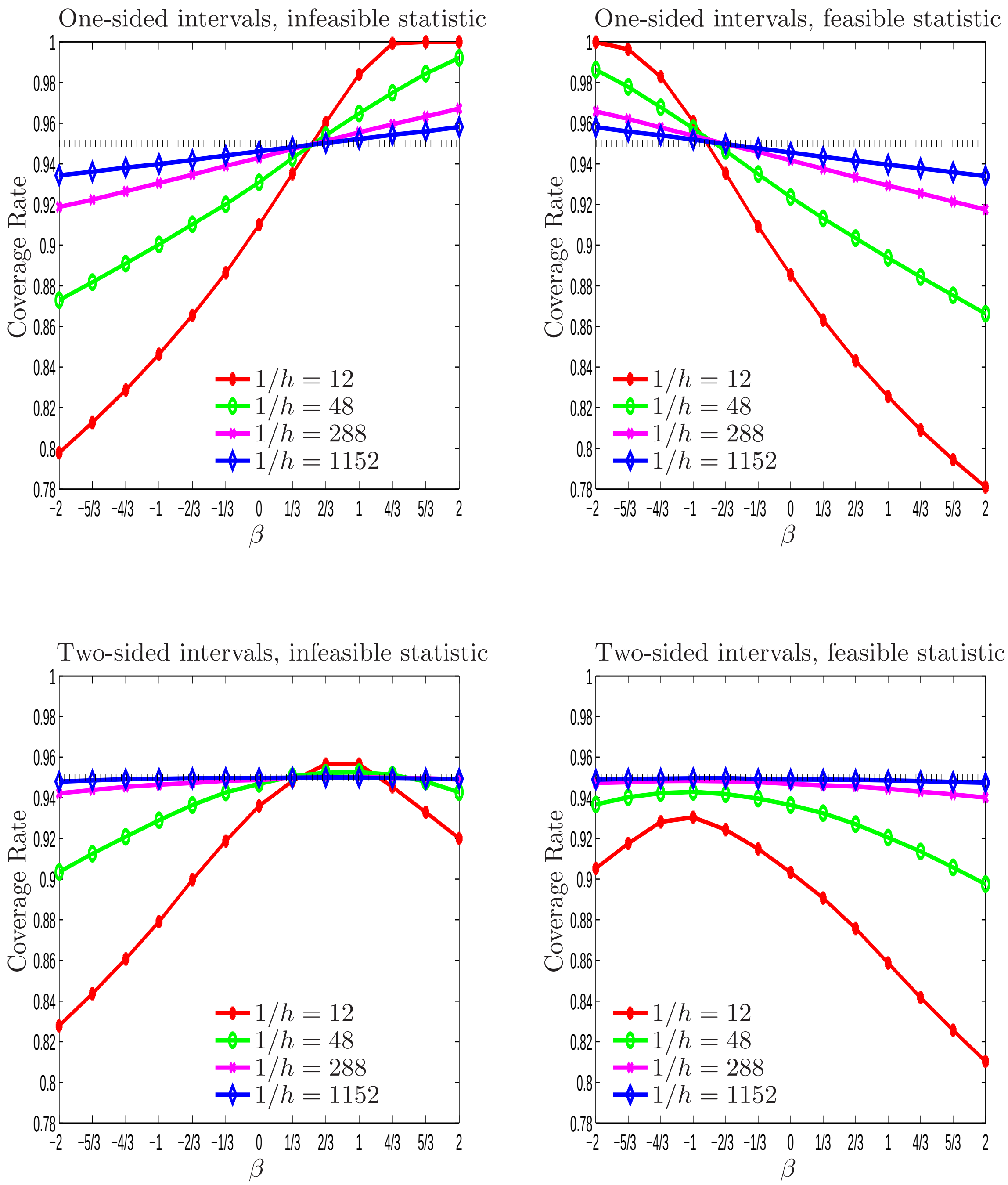

Figure 5: Coverage probabilities of confidence intervals across several values of β, GARCH(1,1) diffusion.

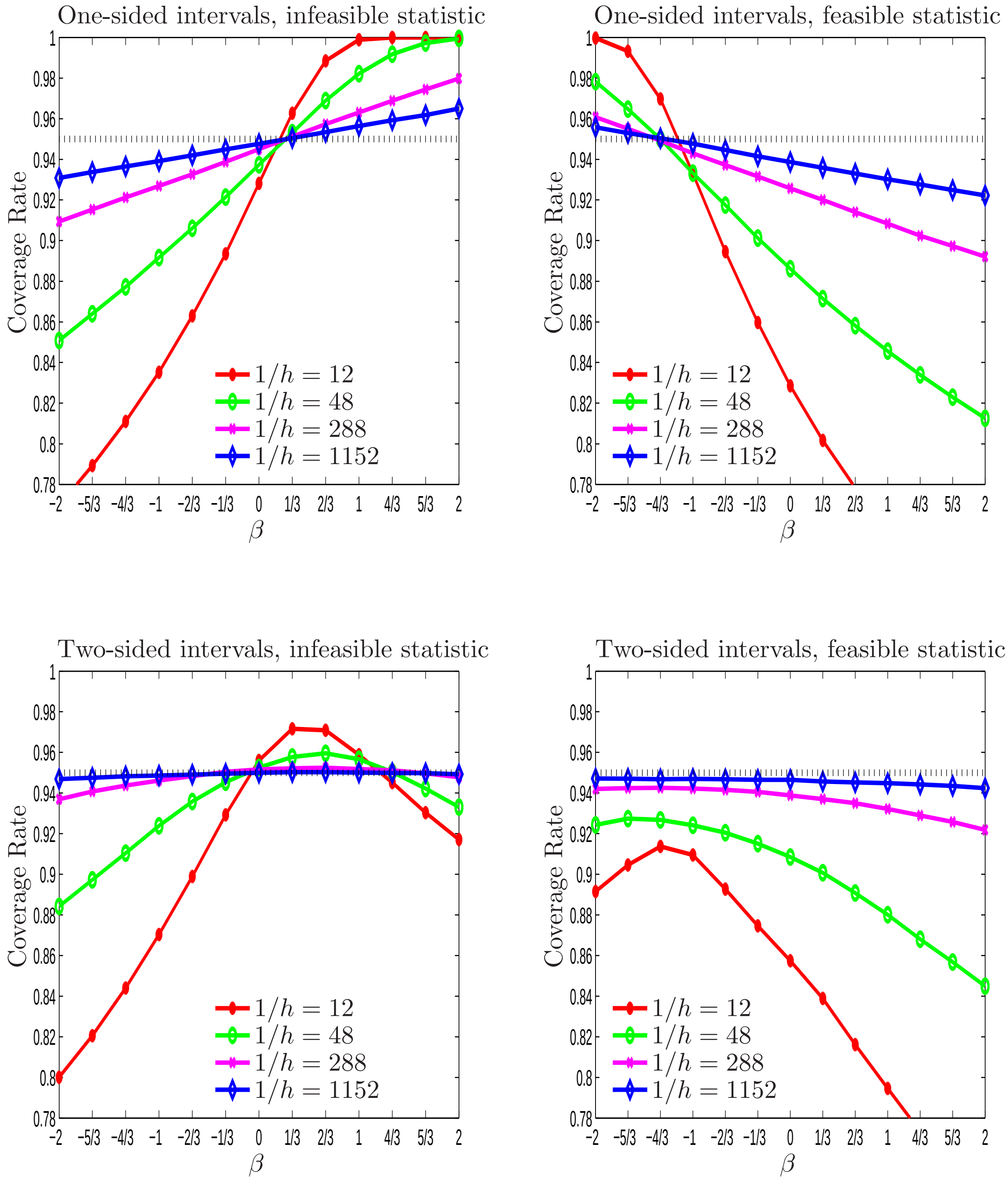

Figure 6: Coverage probabilities of confidence intervals across several values of β, two-factor diffusion.

Appendix A – Details on the Monte Carlo experiment

We consider the following stochastic volatility model

$$d\log S_t = \mu dt + \sigma_t\left[\rho_1 dW_{1t} + \rho_2 dW_{2t} + \sqrt{1-\rho_1^2-\rho_2^2}dW_{3t}\right],$$

where W_{1t}, W_{2t} and W_{3t} are three independent standard Brownian motions. Since we assume no drift and no leverage, $\mu = \rho_1 = \rho_2 = 0$, implying that $d\log S_t = \sigma_t dW_{3t}$.

We consider two different models for σ_t. Our first model is the GARCH(1,1) diffusion studied by Andersen and Bollerslev (1998): $d\sigma_t^2 = 0.035\,(0.636-\sigma_t)\,dt + 0.144\sigma_t^2 dW_{1t}$. We also consider the two-factor diffusion model analyzed by Huang and Tauchen (2006): $\sigma_t =$s-exp$\left(-1.2+0.04\sigma_{1t}^2+1.5\sigma_{2t}^2\right)$, where $d\sigma_{1t}^2 = -0.00137\sigma_{1t}^2 dt + dW_{1t}$, $d\sigma_{2t}^2 = -1.386\sigma_{2t}^2 dt + \left(1+0.25\sigma_{2t}^2\right)dW_{2t}$, and where the function s-exp is the usual exponential function with a linear growth function splined in at high values of its argument: s-exp$(x) = \exp(x)$ if $x \le x_0$ and s-exp$(x) = \frac{\exp(x_0)}{\sqrt{x_0}}\sqrt{x_0 - x_0^2 + x^2}$ if $x > x_0$, with $x_0 = \log(1.5)$.

The explicit form of a confidence interval for $\overline{\sigma^2}$ based on the Box-Cox transform statistic depends on whether $\beta > 0$ or $\beta < 0$. For $\beta > 0$, a lower one sided $(1-\alpha)\%$ confidence interval for $\overline{\sigma^2}$ based on the infeasible statistic S_β is given by $\left(0, \left[R_2^\beta - z_\alpha v_\beta\right]^{1/\beta}\right)$, where $v_\beta = \beta\overline{\sigma^2}^{\beta-1}\sqrt{hV}$, with $V = 2\overline{\sigma^4}$, is the scaling factor for S_β, and where z_α is such that $\Phi(z_\alpha) = \alpha$ for any α. For $\beta < 0$, it is given by $\left(0, \left[R_2^\beta - z_{1-\alpha} v_\beta\right]^{1/\beta}\right)$. When $\beta > 0$, a two-sided symmetric $(1-\alpha)\,\%$ confidence interval based on S_β is given by $\left(\left[R_2^\beta - z_{1-\alpha/2} v_\beta\right]^{1/\beta}, \left[R_2^\beta + z_{1-\alpha/2} v_\beta\right]^{1/\beta}\right)$, whereas for $\beta < 0$, it is given by $\left(\left[R_2^\beta + z_{1-\alpha/2} v_\beta\right]^{1/\beta}, \left[R_2^\beta - z_{1-\alpha/2} v_\beta\right]^{1/\beta}\right)$. The confidence intervals for $\overline{\sigma^2}$ based on the feasible statistic T_β are defined similarly with v_β replaced with $\hat{v}_\beta = \beta R_2^{\beta-1}\sqrt{h\hat{V}}$, where $\hat{V} = \frac{2}{3}R_4$.

Appendix B – Proofs

Proof of Proposition 3.1. We apply Lemmas S.1 and S.2 in GM (2008).

Proof of Corollary 3.1. The goal is to characterize the value of β such that the leading term of $\kappa_1(S_\beta)$ (resp $\kappa_3(S_\beta)$) equals zero; given that $g''(\overline{\sigma^2},\beta)/g'(\overline{\sigma^2},\beta) = (\beta-1)\left(\overline{\sigma^2}\right)^{-1}$, and given Proposition 3.1, the solution is $\beta = 1$ (resp $\beta = \beta^*$). The Cauchy-Schwartz inequality implies $(\overline{\sigma^4})^2 \le \overline{\sigma^2}\;\overline{\sigma^6}$, which leads to $\beta^* \le 1/3$.

Proof of Proposition 3.2. We have

$$\begin{aligned} p_{g_\beta}(x) &= -\frac{1}{\sqrt{2}}\frac{(\overline{\sigma^4})^{1/2}}{\overline{\sigma^2}}((\beta-\beta^*)(x^2-1)+(\beta-1)) = -\frac{1}{\sqrt{2}}\frac{(\overline{\sigma^4})^{1/2}}{\overline{\sigma^2}}((\beta-\beta^*)x^2+(\beta^*-1)) \\ &= \frac{1}{\sqrt{2}}\frac{(\overline{\sigma^4})^{1/2}}{\overline{\sigma^2}}((\beta^*-\beta)x^2+(1-\beta^*)). \end{aligned}$$

When x is fixed and non-zero, the function that appears in the last equation is positive and decreasing when β varies with $\beta \le \beta^* < 1$; therefore, $\left|p_{\beta^*}(x)\right| < \left|p_{\beta_1}(x)\right| < \left|p_{\beta_2}(x)\right|$. When $x = 0$, $p_{g_\beta}(x)$ does not depend on β; hence, $p_{\beta^*}(0) = p_{\beta_1}(0) = p_{\beta_2}(0)$.

Proof of Proposition 3.3. See GM (2008).

Proof of Corollary 3.2. The goal is to characterize the value of β such that the leading term of $\kappa_1(T_\beta)$ (resp $\kappa_3(T_\beta)$) equals zero; given that $g''(\overline{\sigma^2},\beta)/g'(\overline{\sigma^2},\beta) = (\beta-1)\overline{\sigma^2}^{-1}$, and given Proposition 3.3, the solution is $\beta = \beta_*$ (resp $\beta = \beta^{**}$). The Cauchy-Schwartz inequality implies $(\overline{\sigma^4})^2 \le \overline{\sigma^2}\;\overline{\sigma^6}$, which leads to $\beta_* \le -1$ and $\beta^{**} \le -1/3$.

Proof of Proposition 3.4. We have

$$q_{g_\beta}(x) = \frac{\sqrt{2}}{3}\,\frac{\overline{\sigma^6}}{(\overline{\sigma^4})^{3/2}}(2x^2+1) + \frac{(\overline{\sigma^4})^{1/2}}{\sqrt{2}}\frac{g''_\beta(\overline{\sigma^2})}{g'_\beta(\overline{\sigma^2})} = \frac{1}{\sqrt{2}}\frac{(\overline{\sigma^4})^{1/2}}{\overline{\sigma^2}}((\beta-\beta^{**})x^2 + \frac{1}{2}(1-\beta^{**})).$$

When x is fixed and non-zero, the function that appears in the last equation is positive and increasing when β varies with $\beta^{**} < \beta$; therefore, $\left|q_{\beta^{**}}(x)\right| < \left|q_{\beta_1}(x)\right| < \left|q_{\beta_2}(x)\right|$. When $x = 0$, $q_{g_\beta}(x)$ does not depend on β; hence, $q_{\beta^{**}}(0) = q_{\beta_1}(0) = q_{\beta_2}(0)$.

References

[1] Andersen, T.G. and T. Bollerslev, 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, International Economic Review, 39, 885-905.

[2] Andersen, T.G., T. Bollerslev, F. X. Diebold and P. Labys, 2001. The distribution of realized exchange rate volatility, Journal of the American Statistical Association, 96, 42-55.

[3] Andersen, T.G. , T. Bollerslev, F.X. Diebold, and H. Ebens, 2001. The Distribution of Realized Stock Return Volatility, Journal of Financial Economics, 61, 43-76.

[4] Barndorff-Nielsen, O. and N. Shephard, 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society, Series B, 64, 253-280.

[5] Barndorff-Nielsen, O. and N. Shephard, 2004. Power and bipower variation with stochastic volatility and jumps, Journal of Financial Econometrics, 2, 1-48.

[6] Barndorff-Nielsen, O. and N. Shephard, 2005. How accurate is the asymptotic approximation to the distribution of realised volatility? in Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg, (edited by Donald W.K. Andrews and James H.Stock), Cambridge University Press, 306–331.

[7] Barndorff-Nielsen, O., S.E. Graversen, J. Jacod, and N. Shephard, 2006. Limit theorems for bipower variation in financial econometrics, Econometric Theory, 22, 677-719.

[8] Chen, W. C. and R. S. Deo, 2004. Power transformation to induce normality and their applications. Journal of the Royal Statistical Society, Series B, 66, 117-130.

[9] Gonçalves, S. and N. Meddahi, 2008. Bootstrapping realized volatility, Université de Montréal, mimeo.

[10] Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer-Verlag, New York.

[11] Huang, X. and G. Tauchen, 2006. The relative contribution of jumps to total price variance, *Journal of financial econometrics* 3, 456-499.

[12] Jacod, J. and P. Protter, 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. Annals of Probability 26, 267-307.

[13] Marsh, P., 2004. Transformations for multivariate statistics. Econometric Theory, 20, 963-987.

[14] Niki, N. and S. Konishi, 1986. Effects of transformations in higher order asymptotic expansions. Annals of the Institute of Statistical Mathematics 38, Part A, 371-383.

[15] Phillips, P.C.B., 1979. Expansions for Transformations of Statistics, Research note, University of Birmingham, UK.

[16] Phillips, P.C.B, and J. Y. Park, 1988. On the formulation of Wald tests of nonlinear restrictions. Econometrica, 56. 1065-1083.

[17] Wilson, E. B. and M. M. Hilferty, 1931. The distribution of chi square. Proc. Natn. Acad. Sci. USA, 17, 684–688.