# Efficient option pricing for Rough Bergomi model

## 1 The goal and outline of the project

The main goal of the project is to design a fast option pricer, based on multi-index stochastic collocation (MISC), for options whose dynamics follow rBergomi model. We may later investigate QMC.

## 2 Review of literature

Extending the Black-Scholes model, in which volatility is assumed to be constant, to the case where the volatility is stochastic has proved to be successful in explaining certain phenomena observed in option price data, in particular the implied volatility smile. The main drqwback of such stochastic volatility models, however, is that they are unable to capture the true steepness of the implied volatility smile close to maturity. While choosing to add jumps to stock price models, for example modelling the stock price process as an exponential Lévy process, does indeed produce steeper implied volatility smiles, the issue of the presence of jumps in stock price processes remains controversial[1, 12].

As an alternative to diffusive stochastic volatility models, rough stochastic volatility has emerged as a new paradigm in quantitative finance, motivated by the statistical analysis of realised volatility by Gatheral, Jaisson and Rosenbaum [20] and the theoretical results on implied volatility by Fukasawa [17]. In these models, the trajectories of volatility are less regular than those of the standard Brownian motion. As shown in [20, 2], these models are a family of (continuous-path) stochastic volatility models where the driving noise of the volatility process has Hölder regularity lower than Brownian motion, typically achieved by modeling the fundamental noise innovations of the volatility process as a fractional Brownian motion with Hurst exponent (and hence Hölder regularity)$0 < H < 1/2$. A major advantage of such rough volatility models is the fact that they allow to explain crucial phenomena observed in financial markets both from a statistical [20, 6] and an option-pricing point of view [2]. In particular, with regards to the latter point of view, a widely observed empirical phenomenon in equity markets is the steepness of the smile on the short end" describing the fact that as time to maturity becomes small the empirical implied volatility skew follows a power law with negative exponent, and thus becomes arbitrarily large near zero. While standard stochastic volatility models with continuous paths struggle to capture this phenomenon, predicting instead a constant at-the-money implied volatility behaviour on he short end [18], models in the fractional stochastic volatility family (and more specifically so-called rough volatility models) constitute a class, well-tailored to fit empirical implied volatilities for short dated options. Consequently, they have become the go-to models capable of reproducing stylised facts of financial markets.

A fractional Brownian motion is a centred Gaussian process, whose covariance structure depends on the Hurst parameter $H \in (0,1)$. If $H \in (0,1/2)$, then the fractional Brownian motion has negatively correlated increments and "rough" sample paths, and if $H \in (1/2,1)$ then it has positively correlated increments and "smooth" sample paths, when compared with a standard Brownian motion, which is recovered by taking $H = 1/2$. Gatheral, Jaisson, and Rosenbaum [20] provide strong empirical justification for such models; in particular, they argue that log-volatility in practice behaves essentially as fBM with the Hurst exponent $H \approx 0.1$ at any reasonable time scale (see also [19]). This finding is corroborated by Bennedsen, Lunde and Pakkanen [6], who study over a thousand individual US equities and find that the Hurst parameter $H$ lies in $(0,1/2)$ for each equity.

The rough Bergomi model (hereafter rBergomi) is one of the recent rough volatility models, developed by Bayer, Friz and Gatheral [2], that is consistent with the stylised fact of implied volatility surfaces being essentially time-invariant, and are able to capture the term structure of skew observed in equity markets. In [2], the authors discuss issues related to changing from physical to pricing measure and using simulated prices to fit well the implied volatility surface in the case of the S&P 500 index with few parameters-just three!. They argue that the fractional model generates strong skews or "smiles" in the implied volatility even for very short time to maturity so that this modeling provides an alternative to using jumps to model such an effect. In [2] the model is so named because of its relationship with the Bergomi variance curve model [8], and may be seen as a non-Markovian generalisation of the latter.

From a practical perspective the natural question arises: What does the mantra of rough volatility mean for a trader with a view to hedging his positions? Due to the non-Markovian nature of the fractional driver, hedging under rough volatility poses a delicate challenge making even the very definition of hedging strategies difficult.

Typically, the popularity of asset pricing models hinges on the availability of efficient numerical pricing methods. In the case of diffusions, these include Monte Carlo estimators, PDE discretization schemes, asymptotic expansions and transform methods. With fractional Brownian motion being the prime example of a process beyond the semimartingale framework, most currently prevalent option pricing methods -particularly the ones assuming semimartingality or Markovianity - may not easily carry over to the rough setting. In fact, due to the lack of Markovianity or affine structure, conventional analytical pricing methods, such as PDEs or Fourier transform, do not apply. At the moment, the only known method for pricing mere vanilla options, and computing the corresponding implied volatility smiles, in this setting is Monte Carlo simulation. Recent advances in simulation methods for the rough Bergomi model have been achieved in [2, 3, 26, 7, 22]. For isntance, in [26], the authors employ a novel composition of variance reduction methods, immediately applicable to any conditionally log-normal stochastic volatility model. They got a substatantial computation gain in the pricing method over the exisiting Monte carlo methods for this model. On the other hand, more analytical results of option pricing and implied volatility under this model has been done in [23, 4, 16]. For isntance, in [23], they characterise the small-time behaviour of implied volatility using large deviations theory and related results, concerning the small-time near-the-money skew, have been obtained by Bayer, Friz, Gulisashvili, Horvath and Stemper [4]. Pricing and model calibration in rough volatility models remain time consuming.

In this paper, we suggest to design a fast option pricer, based on multi-index stochastic collocation (MISC), for options whose dynamics follow rBergomi model. We may later investigate QMC.

## 3 Background on Gaussian and fBM processes

A zero-mean real-valued Gaussian process $(Z_t)_{t\geq 0}$ is a stochastic process such that on any finite subset $\{t_1, \dots t_n\} \subset \mathbb{R}, (Z_{t_1}, \dots, Z_{t_n})$ has a multivariate normal distribution with mean zero. The law of a Gaussian process is entirely determined by the covariance function $K(s,t) = \mathrm{E}\left[Z_t Z_s\right]$ and $Z$ induces a Gaussian probability measure on $(E, \mathcal{B}(E))$, where $E$ denotes the Banach space $C_0[0,1]$ with the usual sup norm topology (see, e.g., section 3.1.1 of [10] for details).

Fractional Brownian motion (fBM) is a natural generalization of standard Brownian motion which preserves the properties of stationary increments, self-similarity, and Gaussian finite-dimensional distributions, but it has a more complex dependence structure which exhibits long-range dependence when $H > 1/2$. In this section, we recall the definition and summarize the basic properties of fBM.

A zero-mean Gaussian process $B_t^H$ is called standard fractional Brownian motion (fBM) with Hurst parameter $H \in (0,1)$ if it has covariance function

$$R_H = \mathrm{E}\left[B_t^H B_s^H\right] - \mathrm{E}\left[B_t^H\right]\mathrm{E}\left[B_s^H\right] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{1}$$

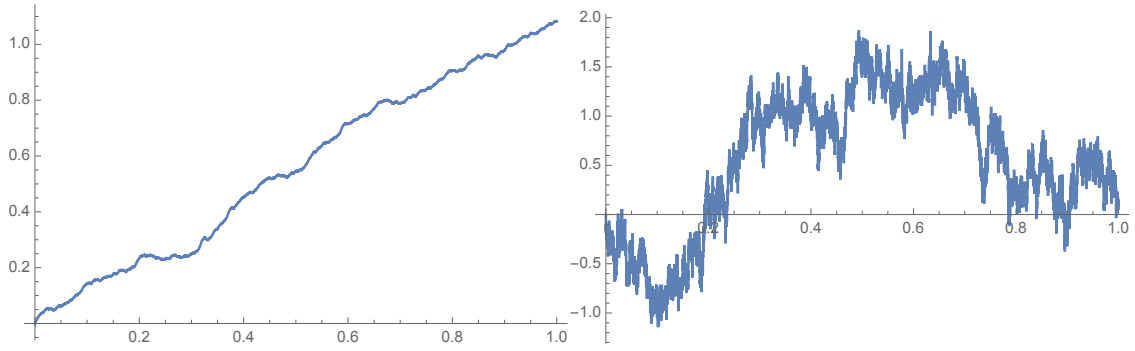
In order to specify the distribution of a Gaussian process, it is enough to specify its mean and its covariance function; therefore, for each $H$, the law of $B^H$ is uniquely determined by $R^H(s,t)$. However, this definition by itself does not guarantee the existence of fBM; to show that fBM exists, one needs to verify that the covariance function is nonnegative definite.

We now recall some fundamental properties of fBM (see also Figure 1):

- fBM is continuous a.s. and H-self-similar (H-ss), i.e., for $a > 0, (B_{at})_{t\geq 0} \overset{(d)}{=} a^H (B_t)_{t\geq 0}$ where $\overset{(d)}{=}$ means both processes have the same finite-dimensional distributions. For $H \neq 1/2$, $B^H$ does not have independent increments; for $H = 1/2, B_t^H$ is the standard Brownian motion.
- From (1), we see that
$$\mathrm{E}\left[B_t^H - B_s^H)^2\right] = |t-s|^{2H},$$
so $B_t^H - B_s^H \sim \mathcal{N}(0, |t-s|^{2H})$; thus $B^H$ has stationary increments.
- If we set $X_n = B_n^H - B_{n-1}^H$, then $X_n$ is a discrete-time Gaussian process with covariance function
$$\begin{aligned}\rho_n = \mathrm{E}\left[X_{k+n} X_n\right] &= \mathrm{E}\left[\left(B_{k+n}^H - B_{k+n-1}^H\right)\left(B_k^H - B_{k-1}^H\right)\right] \\ &\sim H(2H-1)n^{2H-2} \quad (n \to \infty),\end{aligned}$$
and thus (by convexity of the function $g(n) = n^{2H}$), we see that two increments the form $B_k - B_{k-1}$ and $B_{k+n} - B_{k+n-1}$ are positively correlated if $H \in (1/2, 1)$ and negatively correlated if $H \in (0, 1/2)$. Thus $B^H$ is persistent (i.e., it is more likely to keep a trend than to break it) when $H > 1/2$, the relatively stronger positive correlation for the consecutive increments of the associated fBm process with increasing $H$ values gives a relatively smoother process whose correlations decay relatively slowly.And antipersistent when $H < 1/2$ (i.e., if $B^H$ was increasing in the past, it is more likely to decrease in the future, and vice versa). The enhanced negative correlation with smaller Hurst exponent gives a relatively rougher process.

- If $H \in (1/2, 1)$, we can show that $\sum_{n=1}^{\infty} \rho_n = \infty$, which means that the process exhibits long-range dependence, but if $H \in (0, 1/2)$, then $\sum_{n=1}^{\infty} \rho_n < \infty$.
- Using that $\mathrm{E}\left[(B_t^H - B_s^H)^2\right] = (t-s)^{2H}$, we can show that sample paths of $B^H$ are $\alpha$-H older continuous for all $\alpha \in (0, H)$.
- fBM is the only self-similar Gaussian process with stationary increments (see, e.g., [25]), and for $H \neq 1/2$, $B_t^H$ is neither a Markov process nor a semimartingale (see, e.g., [27]).

More details regarding the fBm processes we refer, respectively, to [9, 14, 24].




(a) (b)

Figure 1: Monte Carlo simulation of fBM for $H = 0.9$ (left) and $H = 0.3$ (right).

## 4 The rBergomi model

We use the rBergomi model(rough stochastic volatility model) for the price process $S_t$ as defined in [2], normalized to $r = 0$ and, which is defined by

$$dS_t = \sqrt{v_t(\tilde{W}^H)} S_t dZ_t, \tag{2}$$

$$v_t = \xi_0(t) \exp\left(\eta \tilde{W}_t^H - \frac{1}{2}\eta^2 t^{2H}\right), \tag{3}$$

where for $0 < H < 1$ and $\nu > 0$. We have $\tilde{W}^H$ is a certain Volterra process (Riemann-Liouville process), defined by

$$\tilde{W}_t^H = \int_0^t K^H(t,s) dW_s^1, \quad t \geq 0 \tag{4}$$

where the kernel $K^H : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ reads

$$K^H(t,s) = \sqrt{2H}(t-s)^{H-1/2}, \quad \forall\, 0 \leq s \leq t. \tag{5}$$

We note that the map $s \to K^H(s,t)$ belongs to $L^2$, so that the stochastic integral (4) is well defined.

$W^1, Z$ denote two *correlated* standard Brownian motions with correlation $\rho \in [-1, 1]$, so that

$$Z := \rho W^1 + \overline{\rho} W^{\perp} \equiv \rho W^1 + \sqrt{1-\rho^2} W^{\perp}, \tag{6}$$

where $(W^1, W^{\perp})$ are two independent standard Brownian motions, Therefore, Eq 2 can be written as

$$\begin{aligned} S_t &= S_0 \exp\left(\int_0^t \sqrt{v(s)} dZ(s) - \frac{1}{2}\int_0^t v(s) ds\right), \quad S_0 > 0 \\ v(u) &= \xi_0(u) \exp\left(\eta \tilde{W}_u^H - \frac{\eta^2}{2} u^H\right), \quad \xi_0 > 0 \end{aligned} \tag{7}$$

The filtration $(\mathcal{F}_t)_{t\geq 0}$ can here be taken as the one generated by the two-dimensional Brownian motion $(W^1, W^{\perp})$ under the risk neutral measure $\mathbb{Q}$, resulting in a filtered probability space $(\Omega, \mathcal{F}; \mathcal{F}_t, \mathbb{Q})$. The stock price process $S$ is clearly then a local $(\mathcal{F}_t)_{t\geq 0}$-martingale and a supermartingale, therefore integrable. We shall henceforth use the notation $\mathrm{E}[.] = E^{\mathbb{Q}}[. \mid \mathcal{F}_0]$ unless we state otherwise.

We refer to $v_u$ as the variance process, where $\xi_0(u) = \mathrm{E}[v_u] \in \mathcal{F}_0$ a.s. the forward variance curve. $\tilde{W}^H$ is a centered, locally $(H-\epsilon)$- Hölder continuous, Gaussian process with $\mathrm{var}[\tilde{W}_t^H] = t^{2H}$.

In [2], the approach consists in sampling the Gaussian process $Z$ and $\tilde{W}^H$ on a discrete time grid using exact simulation and then approximating $S$ and $v$ using Euler discretization.

We refrain from formally naming the model parameters $(\nu, \rho, H)$, but those seeking an intuitive understanding of their influence over implied volatilities might like smile for $\nu$,skew for $\rho$, and explosion (of smile and skew) for $H - 1/2$.

## 5 Option pricing under rBergomi model

Assuming $S_0 = 1$, and using the conditioning argument on the $\sigma$-algebra generated by $W^1$ (argument first used by [28] in the context of Markovian SV models), we can show that the call price is given by

$$\begin{aligned} C_{RB}(T, K) &= E\left[(S_T - K)^+\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[(S_T - K)^+ \mid \sigma(W^1(t), t \leq T)\right]\right] \\ &= E\left[C_{BS}\left(S_0 = \exp\left(\rho \int_0^T \sqrt{v_t} dW_t^1 - \frac{1}{2}\rho^2 \int_0^T v_t dt\right),\ K = K,\ T = 1,\ \sigma^2 = (1-\rho^2)\int_0^T v_t dt\right)\right], \end{aligned} \tag{8}$$

where $C_{BS}$ denotes the Black-Scholes price.

In fact, if we use the orthogonal decomposition of $S_t$ into $S_t^1$ and $S_t^2$, where

$$S_t^1 := \mathcal{E}\{\rho \int_0^t \sqrt{v_s} dW_s^1\},\ S_t^2 := \mathcal{E}\{\sqrt{1-\rho^2} \int_0^t \sqrt{v_s} dW_s^{\perp}\}, \tag{9}$$

where $\mathcal{E}()$ denotes the stochastic exponential, then, we obtain by conditional log-normality

$$\log S_t \mid \mathcal{F}_t^1 \sim \mathcal{N}\left(\log S_t^1 - \frac{1}{2}(1-\rho^2)\int_0^t v_s ds, (1-\rho^2)\int_0^t v_s ds\right), \tag{10}$$

where $\mathcal{F}_t^1 = \sigma\{W_s^1 : s \le t\}$. Therefore, we obtain (8).

# 6 Simulation of the rBergomi model

The main challenge is the computation of $S = \int_0^T \sqrt{v_t} dW_t^1$ and $V = \int_0^T v_t dt$. As was mentioned in [3], we may try to avoid any sampling related to $W^2$ by a brute-force approach that consists in simulating a scalar Brownian motion $W^1$, followed by computing $\tilde{W}^H = \int K dW^1$ by Itô/Riemann Stieltjes approximations of $(S, V)$. However, this is not advisable given the singularity of the Volterra kernel $K(s,t)$ at the diagonal $s = t$. Therefore, one needs to jointly simulate the two-dimensional Gaussian process $(W_t^1, \tilde{W}_t^H : 0 \le t \le T)$, resulting in $W_{t_1}^1, \dots, W_{t_N}$ and $\tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}^H$ along a given grid $t_1 < \cdots < t_N$. There are essentially three possible ways to achieve this:

1. Euler discretization of the integral defining $\tilde{W}^H$ together with classical simulation of increments of $W^1$. This is horribly inefficient because the integral is singular and adaptivity probably does not help, as the singularity moves with time. For this method, we need an $N$-dimensional random Gaussian input vector to produce one (approximate, inaccurate) sample of $W_{t_1}^1, \dots, W_{t_N}^1, \tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}$.

2. Given that $W_{t_1}^1, \dots, W_{t_N}^1, \tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}$ together forms a $(2N)$-dimensional Gaussian random vector with computable covariance matrix. We can use Cholesky decomposition of the covariance matrix to produce exact samples of $W_{t_1}^1, \dots, W_{t_N}^1, \tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}$, but unlike the first way, we need $2N$-dimensional Gaussian random vectors as input. This method is exact but slow (See Section 4 in [4]). The simulation requires $\mathcal{O}\left(N^3\right)$. flops.

3. The hybrid scheme of [7] uses a different approach, which is essentially based on Euler discretization as the first way but crucially improved by moment matching for the singular term in the left point rule. It is also inexact in the sense that samples produced here do not exactly have the distribution of $W_{t_1}^1, \dots, W_{t_N}^1, \tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}$,however they are much more accurate then samples produced from method 1), but much faster than method 2). As in method 2), in this case we need a $2N$-dimensional Gaussian random input vector to produce one sample of $W_{t_1}^1, \dots, W_{t_N}^1, \tilde{W}_{t_1}^H, \dots, \tilde{W}_{t_N}$.

In this project, we adopt the last approach for the simulation of the rBergomi model. We utilise the first order variant ($\kappa = 1$) of the hybrid scheme [7], which is based on the approximation

$$\tilde{W}_{\frac{i}{N}}^H \approx \overline{W}_{\frac{i}{N}} := \sqrt{2H}\left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} \left(\frac{i}{N} - s\right)^{H-\frac{1}{2}} dW_u^1 + \sum_{k=2}^{i} \left(\frac{b_k}{N}\right)^{H-\frac{1}{2}} \left(W_{\frac{i-(k-1)}{N}}^1 - W_{\frac{i-k}{N}}^1\right)\right) \tag{11}$$

where $N$ is the number of time steps and

$$b_k := \left(\frac{k^{H+\frac{1}{2}} - (k-1)^{H+\frac{1}{2}}}{H+\frac{1}{2}}\right)^{\frac{1}{H-\frac{1}{2}}}$$

Employing the fast Fourier transform to evaluate the sum in (11), which is a discrete convolution, a skeleton $\overline{W}_0^H, \overline{W}_1^H, \ldots, \overline{W}_{\frac{[Nt]}{N}}^H$ can be generated in $\mathcal{O}(N \log N)$ foating point operations.

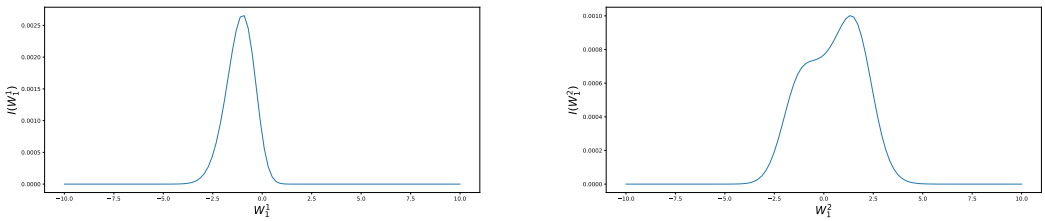
The variates $\overline{W}_0^H, \overline{W}_1^H, \ldots, \overline{W}_{\frac{[Nt]}{N}}^H$ can be generated by sampling $[nt]$ iid draws from a $\kappa + 1$-dimensional Gaussian distribution and computing a discrete convolution. We call these pairs of Gaussian random variables from now on as $(W^1, W^2)$.

# 7 Numerical tests

## 7.1 Integrand plotting wrt different random inputs

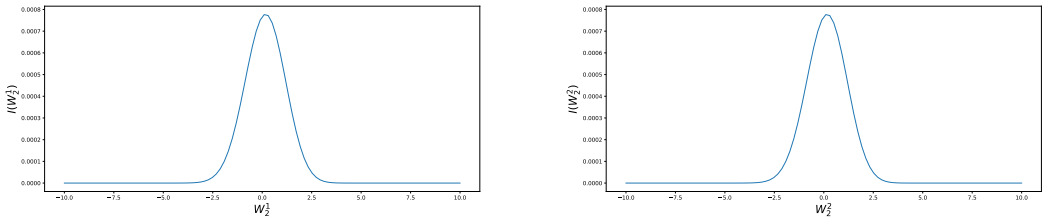
In this section, we plot the integrand, given by the term inside the expectation in (8)(including the Gaussian density), wrt different random inputs $(W_1, W_2)$ ($W_2$ stands for $W^{\text{perp}}$ in the code). This is important to check if we need a measure change and if needed for which variables. We show the results for $H = 0.43$ and $H = 0.07$ and for two scenarios of number of time steps $N \in \{2, 4\}$. We also show the two dimensional plots (See figures (25,4,7,9,8). As it seems from the plots, we may just need change of measure wrt to $W_1$ coordinates and we do not need a measure chnage for $W_2$ coordinates.

### 7.1.1 N=2, H=0.07



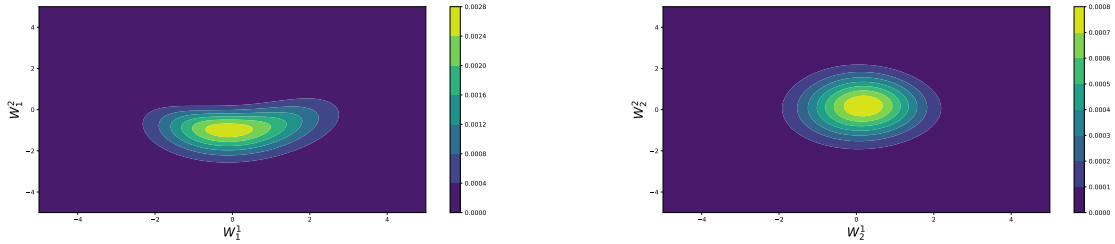
(a) (b)

Figure 2: Plotting the integrand $I$ (in (8)) as a function of $W_1$ coordinates for $H = 0.07$ and $N = 2$.



(a) (b)

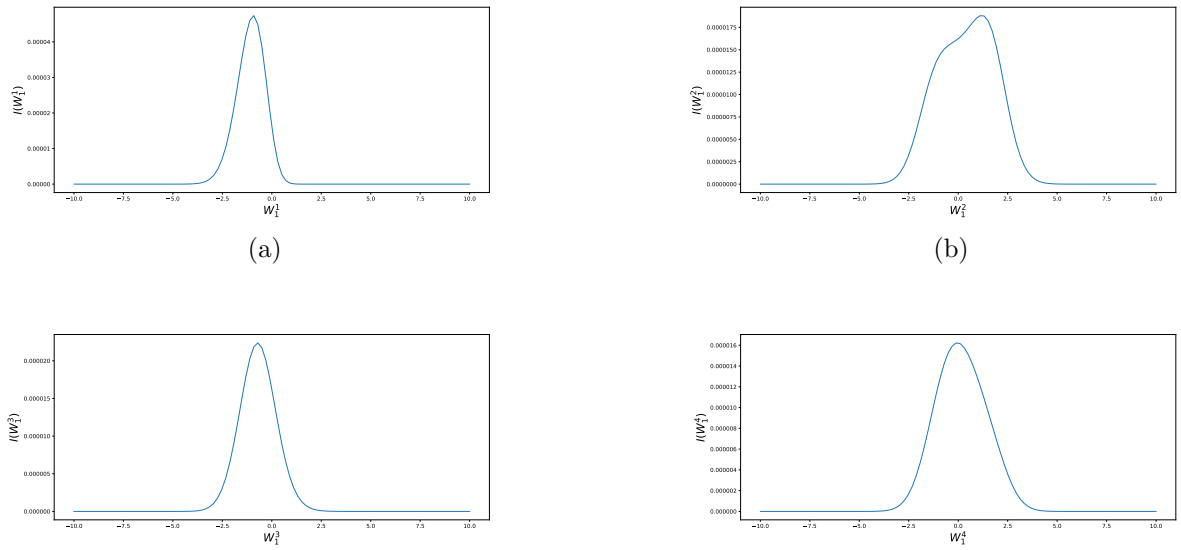
Figure 3: Plotting the integrand $I$ (in (8)) as a function of $W_2$ coordinates for $H = 0.07$ and $N = 2$.



(a) (b)

Figure 4: Two dimensional Plotting of the integrand $I$ (in (8)) for $H = 0.07$ and $N = 2$, a) function of $W_1$ coordinates, b) function of $W_2$ coordinates

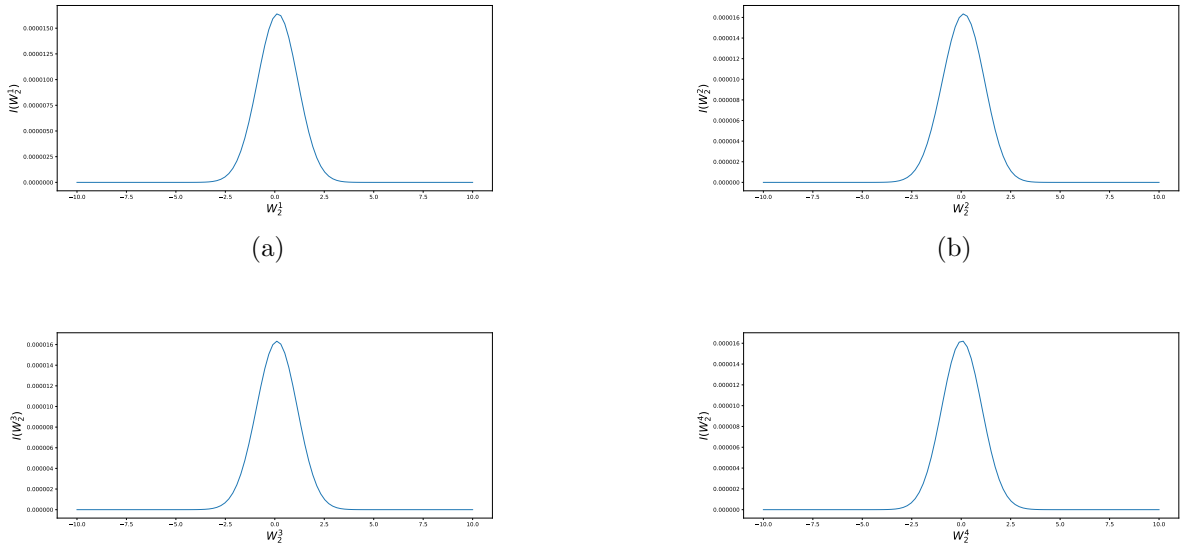
### 7.1.2 N=4, H=0.07



(a) (b)

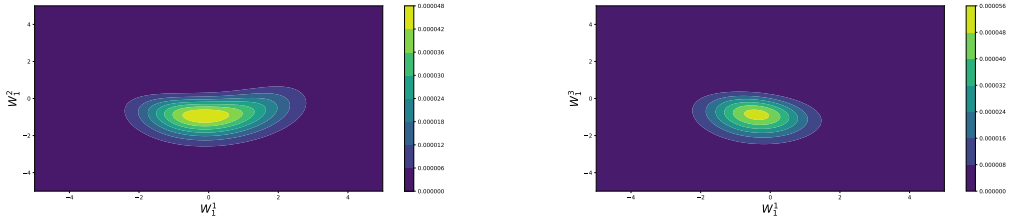
(c) (d)

Figure 5: Plotting the integrand $I$ (in (8)) as a function of $W_1$ coordinates for $H = 0.07$ and $N = 4$.



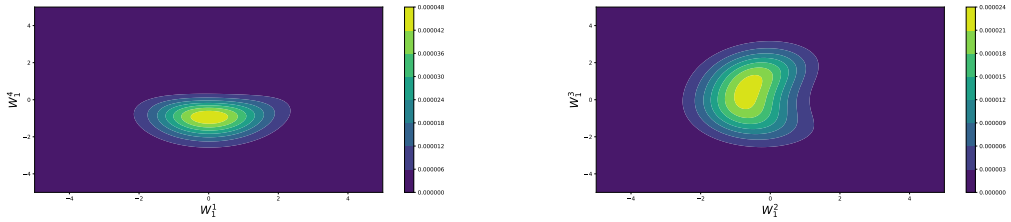
(a) (b) (c) (d)

Figure 6: Plotting the integrand $I$ (in (8)) as a function of $W_2$ coordinates for $H = 0.07$ and $N = 4$.



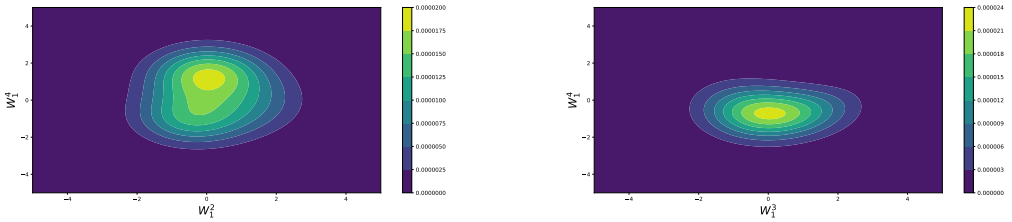
(a) (b)

Figure 7: Two dimensional Plotting of the integrand $I$ (in (8)) for $H = 0.07$ and $N = 4$, a) function of $W_1^1$ and $W_1^2$ , b) function of $W_1^1$ and $W_1^3$



(a) (b)

Figure 8: Two dimensional Plotting of the integrand $I$ (in (8)) for $H = 0.07$ and $N = 4$, a) function of $W_1^1$ and $W_1^4$ , b) function of $W_1^2$ and $W_1^3$

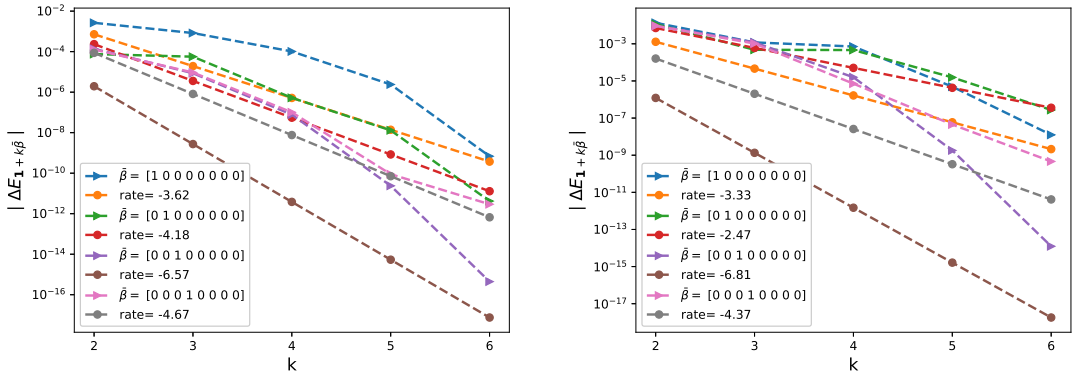


(a) (b)

Figure 9: Two dimensional Plotting of the integrand $I$ (in (8)) for $H = 0.07$ and $N = 4$, a) function of $W_1^2$ and $W_1^4$ , b) function of $W_1^3$ and $W_1^4$

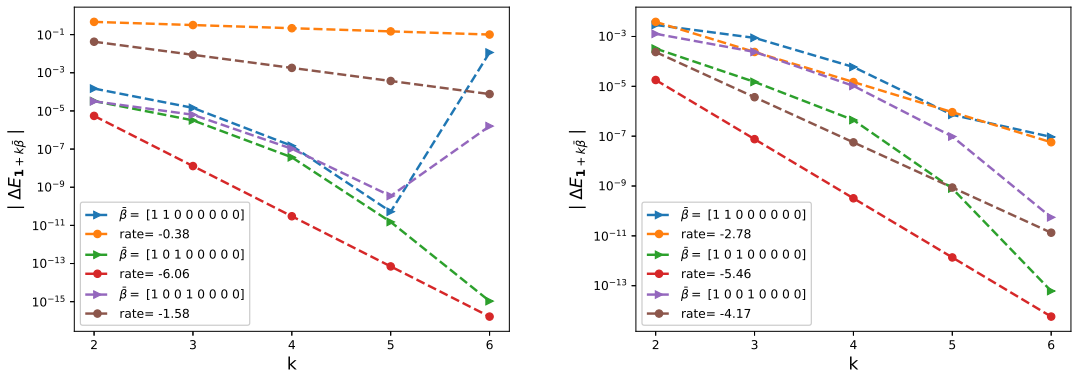
## 7.2 Comparing the mixed differences rates

In this section, we compare the mixed differences rates for the standard case against the case where we do a partial change of measure wrt $W_1$ coordinates (see Appendix A), for the case of $N = 4$ time steps. From the plots, we may notice that we face a bad behavior for the second differences, for the case without change of measure, which may explain the poor performance observed for MISC.This bad behavior is resolved when doing the partial change of measure. We obtained better results when using a measure change based on spectral decomposition rather than Cholesky decomposition. therefore by doing the change of measure, we obtained a more robust MISC solver.



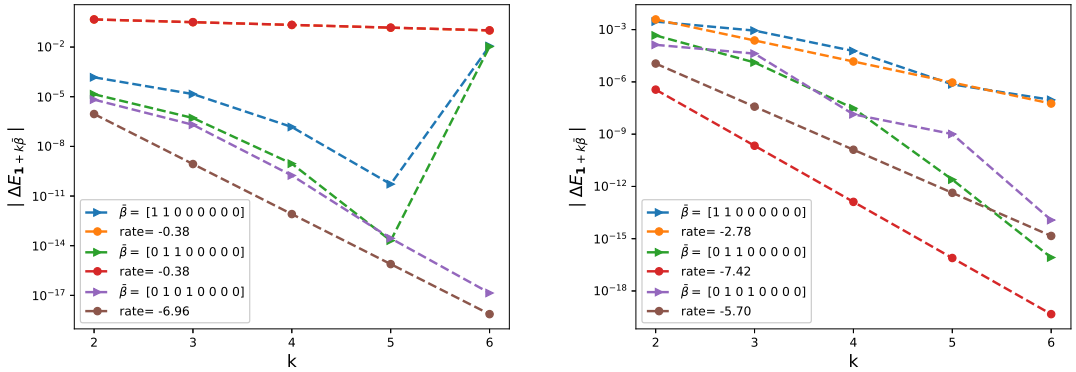

(a) (b)

Figure 10: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1}+k\bar{\beta}$)), for $W_1$, for $K = 1$, $H = 0.07$: a) Without measure change b) With measure change




(a) (b)

Figure 11: The rate of convergence of second order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$, for $W_1$, for $K = 1$, $H = 0.07$: a) Without measure change b) With measure change

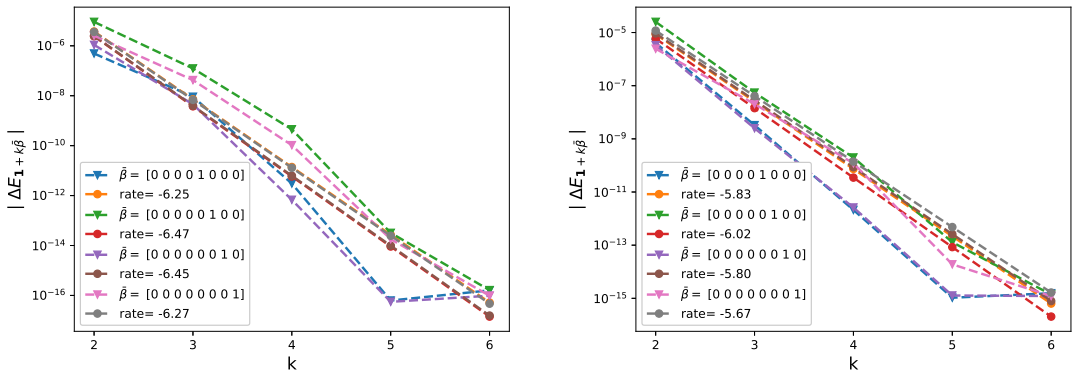



(a) (b)

Figure 12: The rate of convergence of second order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$, for $W_1$, for $K = 1$, $H = 0.07$: a) Without measure change b) With measure change

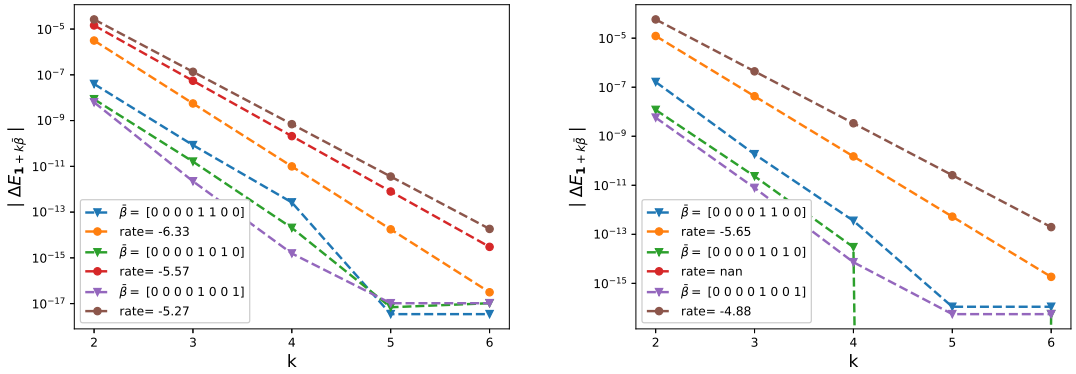
### 7.3 Weak error plots

In this section, I include the results of weak error rates for both cases: without and with change of meausure and without and with Richardson extrapolation, for $H \in \{0.43, 0.07\}$. The reference solution was computed with $N = 500$ time steps. We note that for the case where we do a partial chnage of measure, we limit the maximum number of changed coordinates up to 4, due to practical purposes related to the optimization procedure.




(a) (b)

Figure 13: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1}+k\bar{\beta}))$, for $W_2$, for $K = 1$, $H = 0.07$: a) Without measure change b) With measure change



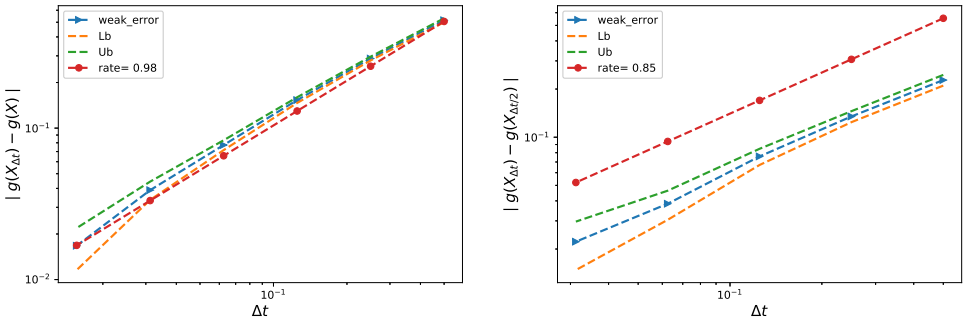

(a) (b)

Figure 14: The rate of convergence of second order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta}))$, for $W_2$, for $K = 1$, $H = 0.07$: a) Without measure change b) With measure change

### 7.3.1 Without change of measure

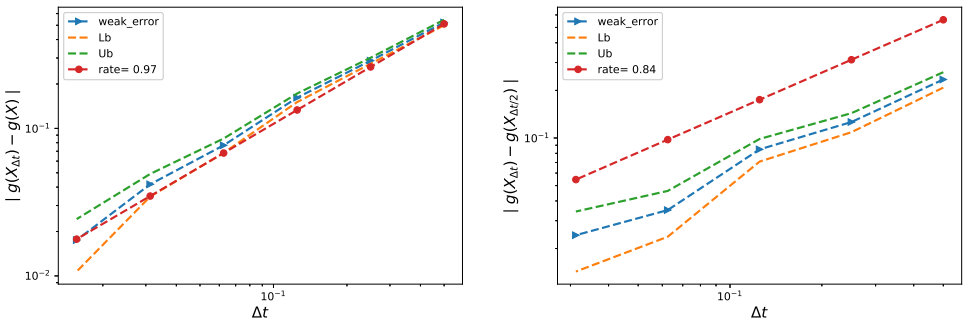
**Without Richardson extrapolation**

From figures (15 and 16), we see that for both cases $H \in \{0.43, 0.07\}$, we get a weak error of order $\Delta t$. The upper and lower bounds are 95% confidence interval.




(a) (b)

Figure 15: The rate of convergence of the weak error for $H = 0.43$ $K = 1$, without Richardson extraploation, using MC with $M = 10^5$: a) $|\mathrm{E}\left[g(X_{\Delta t})\right] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$

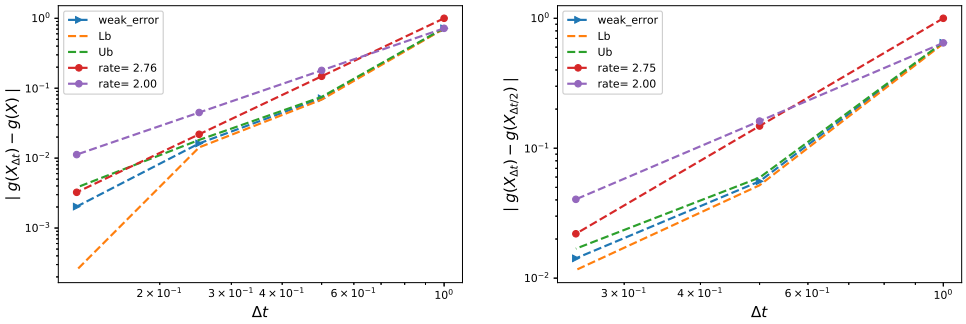



(a) (b)

Figure 16: The rate of convergence of the weak error for $H = 0.07$ $K = 1$, without Richardson extraploation, using MC with $M = 10^5$: a) $|\mathrm{E}\left[g(X_{\Delta t})\right] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$

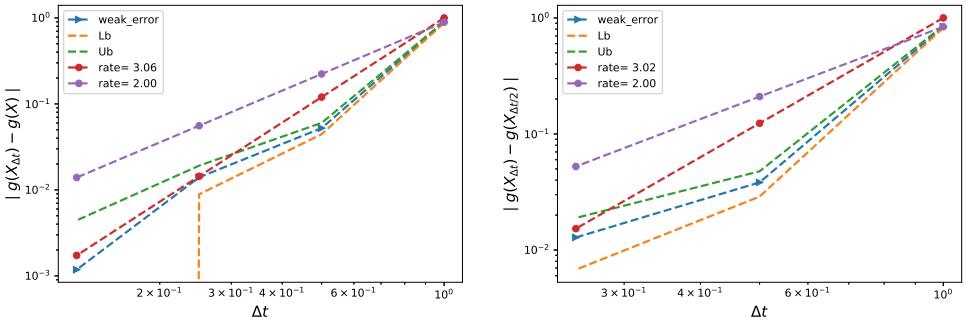
**With Richardson extrapolation (level 1)**

From figures (17 and 18), we see that for both cases $H \in \{0.43, 0.07\}$, we get a weak error of order $\Delta t^2$ (We can see this from the first points, however I think the last points are influenced by the statistical error). The upper and lower bounds are 95% confidence interval.




(a) (b)

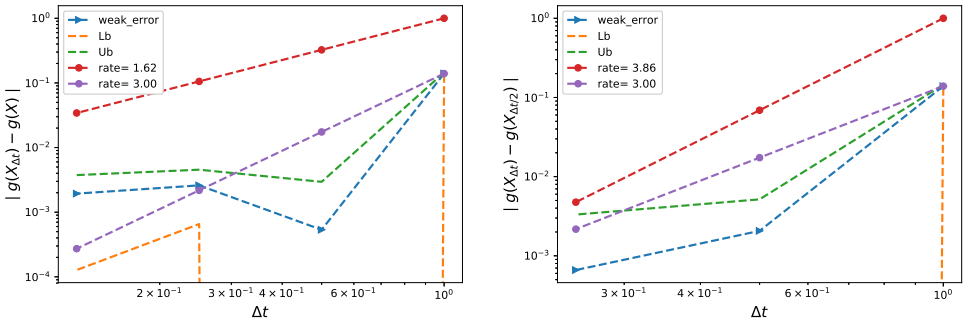
Figure 17: The rate of convergence of the weak error for $H = 0.43$ $K = 1$, with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$




(a) (b)

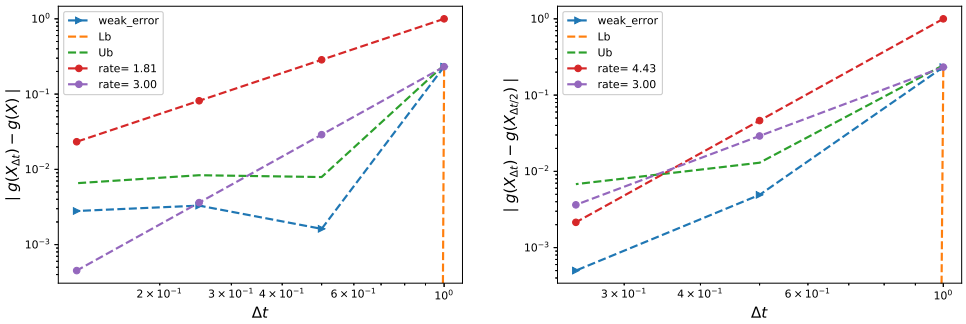
Figure 18: The rate of convergence of the weak error for $H = 0.07$ $K = 1$, with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$

**With Richardson extrapolation (level 2)**




(a) (b)

Figure 19: The rate of convergence of the weak error for $H = 0.43$ $K = 1$, with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\frac{1}{3}\mathrm{E}\left[8g(X_{\Delta t/4}) - 6g(X_{\Delta t/2}) + g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\frac{1}{3}\mathrm{E}\left[-8g(X_{\Delta t/8}) + 14g(X_{\Delta t/4}) - 7(X_{\Delta t/2}) + g(X_{\Delta t})\right]\right|$




(a) (b)

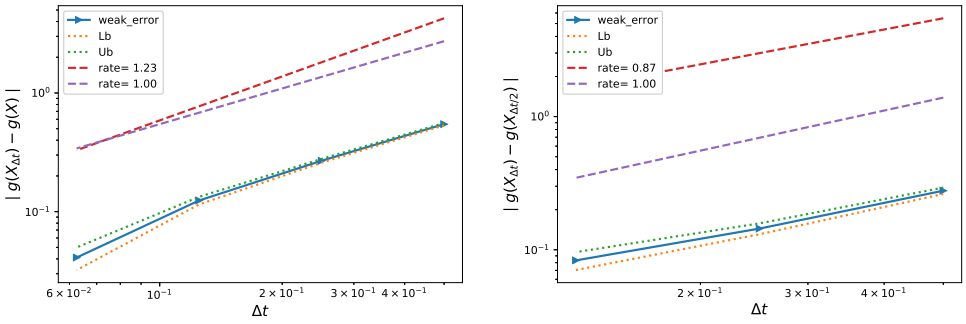
Figure 20: The rate of convergence of the weak error for $H = 0.07$ $K = 1$, with Richardson extraploation (level 2), using MC with $M = 10^6$: a) $\left|\frac{1}{3}\mathrm{E}\left[8g(X_{\Delta t/4}) - 6g(X_{\Delta t/2}) + g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\frac{1}{3}\mathrm{E}\left[-8g(X_{\Delta t/8}) + 14g(X_{\Delta t/4}) - 7(X_{\Delta t/2}) + g(X_{\Delta t})\right]\right|$

### 7.3.2 With change of measure

#### Without Richardson extrapolation

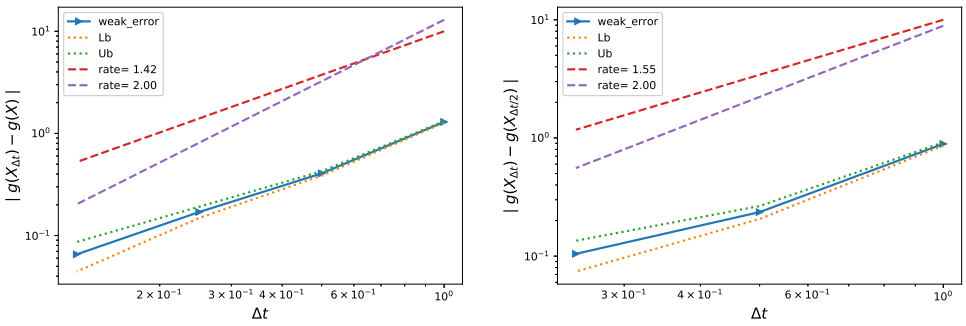
From figures 16), we see that for $H = 0.07$, we get a weak error of order $\Delta t$. The upper and lower bounds are 95% confidence interval.

#### With Richardson extrapolation (level 1)




(a) (b)

Figure 21: The rate of convergence of the weak error for $H = 0.07$ $K = 1$, without Richardson extraploation, using MC with $M = 10^5$: a) $|\mathrm{E}\left[g(X_{\Delta t})\right] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$




(a) (b)

Figure 22: The rate of convergence of the weak error for $H = 0.07$ $K = 1$, with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$

### 7.4 Comparing relative errors using hierarchical representation

The used parameters are $H = \{0.43, 0.07\}$, $\eta = 1.9$, $\rho = -0.9$, $T = 1, K = 1$ and $\xi_0 = 0.235^2$. The results were reported for number of time steps $N \in \{2, 4, 8, 16\}$. Also, we use $S_0 = 1$, so the options will be prices in terms of the moneyness $K$, where $K$ is the strike price.

In the following, we compare the relative errors for H $\in \{0.43, 0.07\}$ (see appendices B.3 and B.4 for the values of Call option prices). We note that for each case the reference solution was computed for $N = 500$ (number of time steps) using MC with $10^6$ samples. In each case we report the results for 3 scenarios: i) Without using Richardson extrapolation, ii) Using level 1 Richardson extrapoaltion and iii) Using level 2 Richardson extrapoaltion. Tables (1, 2, 3) coresspond to $H = 0.43$ and tables (4, 5, 6) coresspond to $H = 0.07$.

Given the normalized bias computed by MC method (See Section 7.3) (reported as bold values in the tables), we report in red in each table the smallest tolerance that MISC required to get below

that relative bias (I do not put values for smaller tolerances, once the required bias is reached).

From the tables below, we have the following observations:

- Using Richardson extrapoaltion, we got a significant improvement for the relative error with the use of minimal time steps. For instance, for $H = 0.43$, we achieved around 8% of relative error, with 16 time steps when not using Richardson extrapolation (see table 1). However, When using level 1 of Richardson extrapolation (see table 2), we achieved around 6% of relative error, with only 2 time steps in the coarse level, and we got around 1% of relative error, with 4 time steps in the coarse level. A more significant improvement is seen with level 2 of Richardson extrapolation, in fact, with just 1 step in the coarse level, we got around 3% percent of relative error.

- For $H = 0.07$, we achieved around 8% of relative error, with 16 time steps when not using Richardson extrapolation (see table 4). However, When using level 1 of Richardson extrapolation (see table 5), we achieved around 5% of relative error, with only 2 time steps in the coarse level, and we got below 1% of relative error, with 4 time steps in the coarse level. We observed a less significant improvement when using levle 2 of Ricchardson extrapolation, compared to the case of $H = 0.43$.

### 7.4.1 Case $H = 0.43$, Relative error for different methods

| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.6011 | 0.3497 | 0.1910 | 0.0969 |
| MISC ($TOl = 2.10^{-1}$) | 0.6011 | 0.3497 | 0.1910 | 0.0801 |
| MISC ($TOl = 10^{-1}$) | 0.6011 | 0.3497 | 0.2233 | 0.1236 |
| MISC ($TOl = 5.10^{-2}$) | 0.6011 | 0.3539 | 0.1882 | 0.1573 |
| MISC ($TOl = 10^{-2}$) | 0.5126 | 0.3258 | 0.1770 | 0.0829 |
| MISC ($TOl = 5.10^{-3}$) | 0.4930 | 0.3076 | 0.1671 | – |
| MISC ($TOl = 10^{-3}$) | 0.5126 | 0.2935 | 0.1503 | – |
| MISC ($TOl = 10^{-4}$) | 0.5154 | 0.2935 | – | – |
| MC method ($M = 10^6$) | **0.5154** | **0.2935** | **0.1545** | **0.0801** |

Table 1: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.43$, without Richardson extrapolation

| Method \Steps | $1-2$ | $2-4$ | $4-8$ | $8-16$ |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.9059 | 0.0997 | 0.0323 | 0.0028 |
| MISC ($TOl = 10^{-1}$) | 0.9059 | 0.0997 | 0.1025 | 0.0688 |
| MISC ($TOl = 5.10^{-2}$) | 0.9059 | 0.1671 | 0.0857 | 0.0646 |
| MISC ($TOl = 10^{-2}$) | 0.7374 | 0.0969 | 0.0463 | 0.0028 |
| MISC ($TOl = 5.10^{-3}$) | 0.7205 | 0.0941 | 0.0211 | – |
| MISC ($TOl = 10^{-3}$) | 0.7191 | 0.0758 | 0.0112 | – |
| MISC ($TOl = 5.10^{-4}$) | 0.7129 | 0.0609 | – | – |
| MC method ($M = 10^6$) | **0.7133** | **0.0698** | **0.0160** | **0.0035** |

Table 2: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.43$, using Richardson extrapolation (level 1)

| Method \Steps | $1-2-4$ | $2-4-8$ | $4-8-16$ |
|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1699 | 0.0098 | 0.0056 |
| MISC ($TOl = 2.10^{-1}$) | 0.1699 | – | 0.0014 |
| MISC ($TOl = 10^{-1}$) | 0.2037 | – | – |
| MISC ($TOl = 5.10^{-2}$) | 0.0295 | – | – |
| MC method ($M = 10^6$) | **0.1440** | **0.0180** | **0.0023** |

Table 3: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.43$, using Richardson extrapolation (level 2)

### 7.4.2 Case $H = 0.07$, Relative error for different methods

| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.3662 | 0.1578 | 0.1010 | 0.0758 |
| MISC ($TOl = 10^{-1}$) | 0.3662 | 0.1578 | – | – |
| MISC ($TOl = 5.10^{-2}$) | 0.3662 | – | – | – |
| MISC ($TOl = 10^{-2}$) | – | – | – | – |
| MC method ($M = 10^6$) | **0.5354** | **0.2879** | **0.1515** | **0.0783** |

Table 4: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.07$, without Richardson extrapolation

| Method \Steps | $1-2$ | $2-4$ | $4-8$ | $8-16$ |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.5682 | 0.0505 | 0.1389 | 0.1604 |
| MISC ($TOl = 16.10^{-2}$) | 0.5682 | 0.0505 | 0.1389 | 0.0038 |
| MISC ($TOl = 10^{-1}$) | 0.5682 | 0.0505 | 0.1692 | – |
| MISC ($TOl = 5.10^{-2}$) | 0.5682 | 0.1465 | 0.0088 | – |
| MISC ($TOl = 10^{-2}$) | – | 0.0669 | 0.0088 | – |
| MC method ($M = 10^6$) | **0.8915** | **0.0537** | **0.0129** | **0.0043** |

Table 5: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.07$, using Richardson extrapolation (level 1)

| Method \Steps | $1-2-4$ | $2-4-8$ | $4-8-16$ |
|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.2563 | 0.1692 | 0.1679 |
| MISC ($TOl = 10^{-1}$) | 0.2563 | 0.1566 | 0.0025 |
| MISC ($TOl = 7.10^{-2}$) | 0.3005 | 0.0227 | – |
| MISC ($TOl = 5.10^{-2}$) | 0.4874 | – | – |
| MISC ($TOl = 10^{-2}$) | 0.1742 | – | – |
| MC method ($M = 10^6$) | **0.2231** | **0.0279** | **0.0035** |

Table 6: Relative error of Call option price of the different tolerances for different number of time steps. Case $K = 1$, $H = 0.07$, using Richardson extrapolation (level 2)

# References Cited


[1] Pierre Bajgrowicz, Olivier Scaillet, and Adrien Treccani. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, 62(8):2198–2217, 2015.

[2] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[3] Christian Bayer, Peter K Friz, Paul Gassiat, Joerg Martin, and Benjamin Stemper. A regularity structure for rough volatility. *arXiv preprint arXiv:1710.07481*, 2017.

[4] Christian Bayer, Peter K Friz, Archil Gulisashvili, Blanka Horvath, and Benjamin Stemper. Short-time near-the-money skew in rough fractional volatility models. *arXiv preprint arXiv:1703.05132*, 2017.

[5] Mikkel Bennedsen et al. Rough electricity: a new fractal multi-factor model of electricity spot prices. 2015.

[6] Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Decoupling the short-and long-term behavior of stochastic volatility. *arXiv preprint arXiv:1610.00332*, 2016.

[7] Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Hybrid scheme for brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

[8] Lorenzo Bergomi. Smile dynamics ii. 2005.

[9] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Probability and Its Applications. Springer London, 2008.

[10] R. Carmona and M.R. Tehranchi. *Interest Rate Models: an Infinite Dimensional Stochastic Analysis Perspective*. Springer Finance. Springer Berlin Heidelberg, 2007.

[11] Lanouar Charfeddine. True or spurious long memory in volatility: Further evidence on the energy futures markets. *Energy policy*, 71:76–93, 2014.

[12] Kim Christensen, Roel CA Oomen, and Mark Podolskij. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599, 2014.

[13] Alexandra Chronopoulou and Frederi G Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2-3):379–403, 2012.

[14] Laure Coutin. An introduction to (stochastic) calculus with respect to fractional brownian motion. In *Séminaire de Probabilités XL*, pages 3–65. Springer, 2007.

[15] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 2016.

[16] Martin Forde and Hongzhong Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1):114–145, 2017.

[17] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[18] Jim Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[19] Jim Gatheral, Thibault Jaisson, Andrew Lesniewski, and Mathieu Rosenbaum. Volatility is rough, part 2: Pricing.

[20] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *arXiv preprint arXiv:1410.3394*, 2014.

[21] Blanka Horvath, Antoine Jacquier, and Aitor Muguruza. Functional central limit theorems for rough volatility. 2017.

[22] Antoine Jacquier, Claude Martini, and Aitor Muguruza. On vix futures in the rough bergomi model. *Quantitative Finance*, 18(1):45–61, 2018.

[23] Antoine Jacquier, Mikko S Pakkanen, and Henry Stone. Pathwise large deviations for the rough bergomi model. *arXiv preprint arXiv:1706.05291*, 2017.

[24] Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[25] Tina Marquardt et al. Fractional lévy processes with an application to long memory moving average processes. *Bernoulli*, 12(6):1099–1126, 2006.

[26] Ryan McCrickerd and Mikko S Pakkanen. Turbocharging monte carlo pricing for the rough bergomi model. *arXiv preprint arXiv:1708.02563*, 2017.

[27] David Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006.

[28] Marc Romano and Nizar Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997.

[29] Martin Rypdal and Ola Løvsletten. Modeling electricity spot prices using mean-reverting multifractal processes. *Physica A: Statistical Mechanics and its Applications*, 392(1):194–207, 2013.

[30] Peter Tankov. Pricing and hedging in exponential lévy models: review of recent results. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 319–359. Springer, 2011.

[31] Thomas Walther, Tony Klein, Hien Pham Thu, and Krzysztof Piontek. True or spurious long memory in european non-emu currencies. *Research in International Business and Finance*, 40:217–230, 2017.

# A Gaussian Hermite Quadrature with importance sampling

Let us call the integrand that we feed to MISC by $I(W_1, W_2)$, then

$$C_{RB}(T,K) = \int_{\mathbb{R}_+^{2N}} I(\mathbf{W}_1, \mathbf{W}_2)\rho(\mathbf{W}_1)\rho(\mathbf{W}_2)d\mathbf{W}_1 d\mathbf{W}_2, \tag{12}$$

where $N$ is the number of time steps. We can rewrite the previous expression as

$$C_{RB}(T,K) = \int_{\mathbb{R}_+^{2N}} \frac{I(\mathbf{W}_1, \mathbf{W}_2)\rho(\mathbf{W}_1)}{h(\mathbf{W}_1; \widehat{\mathbf{W}}_1, \Psi)} h(\mathbf{W}_1; \widehat{\mathbf{W}}_1, \Psi)\rho(\mathbf{W}_2)d\mathbf{W}_1 d\mathbf{W}_2, \tag{13}$$

where $h(\mathbf{W}_1; \widehat{\mathbf{W}}_1, \Psi)$ is a multivariate normal density with first and second order moments given by

$$\widehat{\mathbf{W}}_1 = \arg\max_{\mathbf{W}_1 \in \mathbb{R}^N} [\log I(\mathbf{W}_1; \mathbf{W}_2)] \tag{14}$$

$$\Psi = \left( -\frac{\partial^2 [\log I(\mathbf{W}_1; \mathbf{W}_2)]}{\partial \mathbf{W}_1^T \mathbf{W}_1} \right)^{-1}_{\mathbf{W}_1 = \widehat{\mathbf{W}}_1} \tag{15}$$

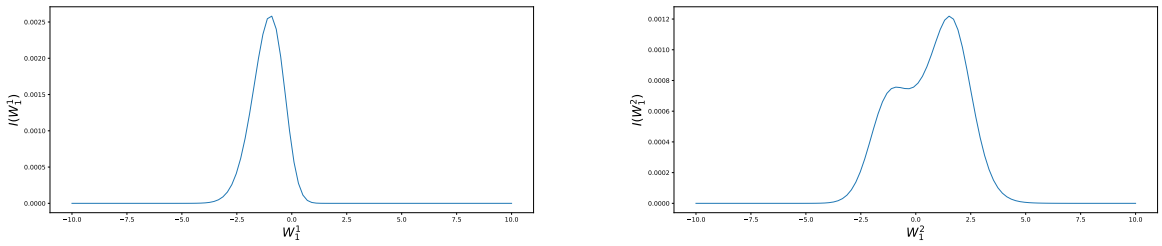
Let us define $\tilde{\mathbf{W}}_1$ as uncorrelated varaibles and the Cholesky factorization of $\Psi$ is given by $\Psi = LL^T$, and $\overline{\mathbf{W}}_1 = \sqrt{2}L\tilde{\mathbf{W}}_1 + \widehat{\mathbf{W}}_1$ then Eq 13 becomes

(16)

$$C_{RB}(T,K) = = 2^{N/2}.|L| \int_{\mathbb{R}_+^{2N}} \left( I(\overline{\mathbf{W}}_1, \mathbf{W}_2) \exp(-\frac{1}{2}\overline{\mathbf{W}}_1^T \overline{\mathbf{W}}_1) \exp(\frac{1}{2}\tilde{\mathbf{W}}^T \tilde{\mathbf{W}}) \right) \rho(\tilde{\mathbf{W}}_1)\rho(\mathbf{W}_2) d\tilde{\mathbf{W}}_1 d\mathbf{W}_2$$

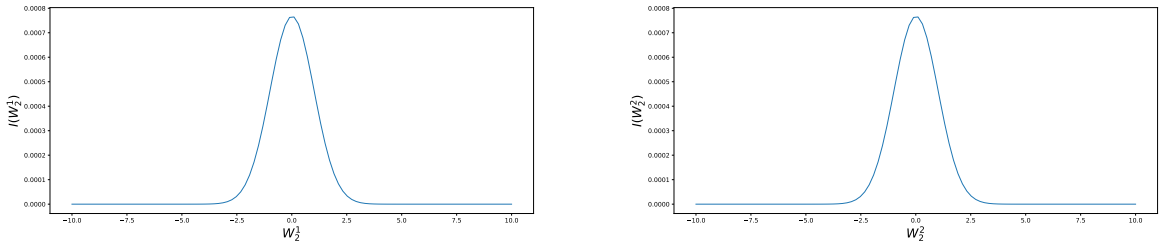
# B additional results

## B.1 Integrand plotting wrt different random inputs N=2, H=0.43



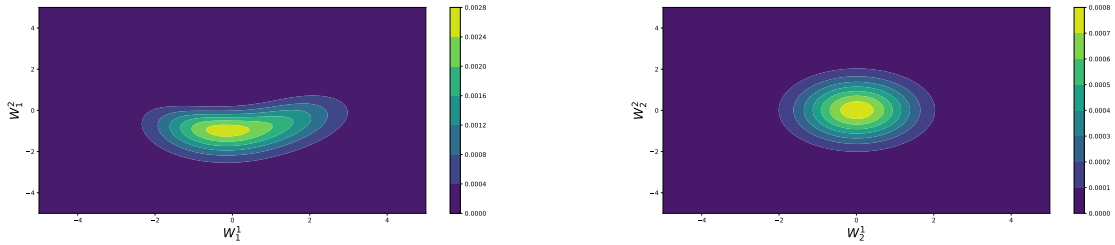
(a) (b)

Figure 23: Plotting the integrand $I$ (in (8)) as a function of $W_1$ coordinates for $H = 0.43$ and $N = 2$.



(a) (b)

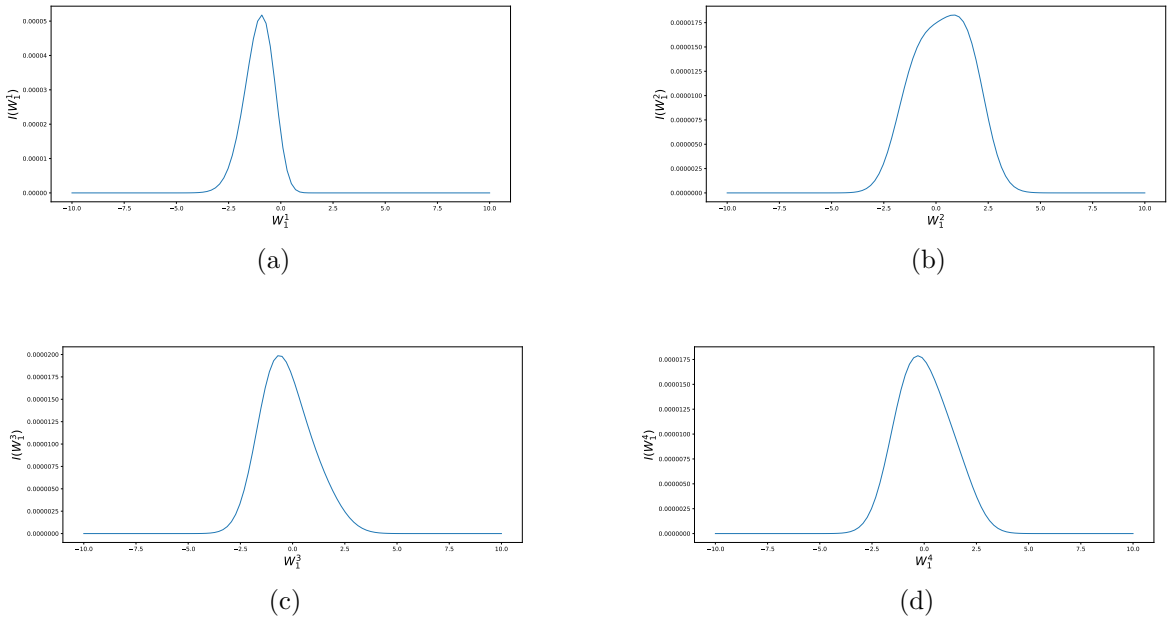
Figure 24: Plotting the integrand $I$ (in (8)) as a function of $W_2$ coordinates for $H = 0.43$ and $N = 2$.



(a) (b)

Figure 25: Two dimensional Plotting of the integrand $I$ (in (8)) for $H = 0.43$ and $N = 2$, a) function of $W_1$ coordinates, b) function of $W_2$ coordinates

## B.2 Integrand plotting wrt different random inputs: N=4, H=0.43




(a) (b)

(c) (d)

Figure 26: Plotting the integrand $I$ (in (8)) as a function of $W_1$ coordinates for $H = 0.43$ and $N = 4$.

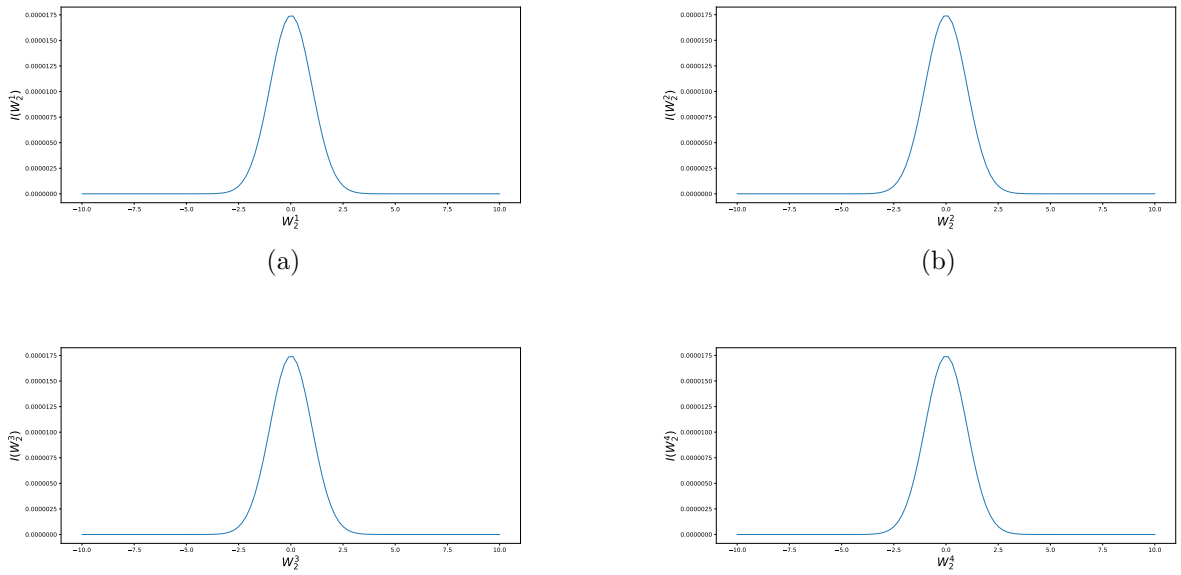


Figure 27: Plotting the integrand $I$ (in (8)) as a function of $W_2$ coordinates for $H = 0.43$ and $N = 4$.

### B.3 Case $H = 0.43$, Call prices for different methods

| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1140 | 0.0961 | 0.0848 | 0.0781 |
| MISC ($TOl = 2.10^{-1}$) | 0.1140 | 0.0961 | 0.0848 | 0.0769 |
| MISC ($TOl = 10^{-1}$) | 0.1140 | 0.0961 | 0.0871 | 0.0800 |
| MISC ($TOl = 5.10^{-2}$) | 0.1140 | 0.0964 | 0.0846 | 0.0824 |
| MISC ($TOl = 10^{-2}$) | 0.1077 | 0.0944 | 0.0838 | 0.0771 |
| MISC ($TOl = 5.10^{-3}$) | 0.1063 | 0.0931 | 0.0831 | – |
| MISC ($TOl = 10^{-3}$) | 0.1077 | 0.0921 | 0.0819 | – |
| MISC ($TOl = 10^{-4}$) | 0.1079 | 0.0921 | – | – |
| MC method ($M = 10^6$) | 0.1079 (1.55$e$−04) | 0.0921 (9.65$e$−05) | 0.0822 (7.61$e$−05) | 0.0769 (6.65$e$−05) |

Table 7: Call option price of the different methods for different number of time steps. Case $K = 1$, $H = 0.43$, without Richardson extrapolation. The values between parentheses in the tables are the standard errors for MC method

| Method \Steps | $1-2$ | $2-4$ | $4-8$ | $8-16$ |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1357 | 0.0783 | 0.0735 | 0.0714 |
| MISC ($TOl = 10^{-1}$) | 0.1357 | 0.0783 | 0.0785 | 0.0761 |
| MISC ($TOl = 5.10^{-2}$) | 0.1357 | 0.0831 | 0.0773 | 0.0758 |
| MISC ($TOl = 10^{-2}$) | 0.1237 | 0.0781 | 0.0745 | 0.0714 |
| MISC ($TOl = 5.10^{-3}$) | 0.1225 | 0.0779 | 0.0727 | – |
| MISC ($TOl = 10^{-3}$) | 0.1224 | 0.0766 | 0.0720 | – |
| MISC ($TOl = 5.10^{-4}$) | 0.1221 | 0.0763 | – | – |

Table 8: Call option price of the different methods for different number of time steps. Case $K = 1$, $H = 0.43$, using Richardson extrapolation (level 1)

| Method \Steps | $1-2-4$ | $2-4-8$ | $4-8-16$ |
|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.0591 | 0.0719 | 0.0708 |
| MISC ($TOl = 2.10^{-1}$) | 0.0591 | – | 0.0711 |
| MISC ($TOl = 10^{-1}$) | 0.0567 | – | – |
| MISC ($TOl = 5.10^{-2}$) | 0.0733 | – | – |

Table 9: Call option price of the different methods for different number of time steps. Case $K = 1$, $H = 0.43$, using Richardson extrapolation (level 2)

### B.4 Case $H = 0.07$, Call prices for different methods

| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1082 | 0.0917 | 0.0872 | 0.0732 |
| MISC ($TOl = 10^{-1}$) | 0.1082 | 0.0917 | – | – |
| MISC ($TOl = 5.10^{-2}$) | 0.1082 | – | – | – |
| MISC ($TOl = 10^{-2}$) | – | – | – | – |
| MC method ($M = 10^6$) | 0.1216 (1.05e−03) | 0.1020 (1.86e−04) | 0.0912 (1.35e−04) | 0.0854 (1.08e−04) |

Table 10: Call option price of the different methods for different number of time steps. Case $K = 1$, without Richardson extrapolation. The values between parentheses in the tables are the standard errors for MC method

| Method \Steps | $1-2$ | $2-4$ | $4-8$ | $8-16$ |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1242 | 0.0752 | 0.0682 | 0.0665 |
| MISC ($TOl = 16.10^{-2}$) | 0.1242 | 0.0752 | 0.0682 | 0.0795 |
| MISC ($TOl = 10^{-1}$) | 0.1242 | 0.0752 | 0.0658 | – |
| MISC ($TOl = 5.10^{-2}$) | 0.1242 | 0.0676 | 0.0799 | – |
| MISC ($TOl = 10^{-2}$) | – | 0.0845 | 0.0799 | – |

Table 11: Call option price of the different methods for different number of time steps. Case $K = 1$, $H = 0.07$, using Richardson extrapolation (level 1)

| Method \Steps | $1-2-4$ | $2-4-8$ | $4-8-16$ |
|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.0589 | 0.0658 | 0.0659 |
| MISC ($TOl = 10^{-1}$) | 0.0589 | 0.0668 | 0.079 |
| MISC ($TOl = 7.10^{-2}$) | 0.0554 | 0.0810 | – |
| MISC ($TOl = 5.10^{-2}$) | 0.0406 | – | – |
| MISC ($TOl = 10^{-2}$) | 0.0654 | – | – |

Table 12: Call option price of the different methods for different number of time steps. Case $K = 1$, $H = 0.07$, using Richardson extrapolation (level 2)

## B.5 Comparing call options prices

### B.5.1 Without Hierarchical representation

**Case** $H = 0.43$

| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1057 | 0.0988 | 0.0944 | 0.0921 |
| MISC ($TOl = 10^{-1}$) | 0.1057 | 0.0988 | 0.0836 | 0.0594 |
| MISC ($TOl = 5.10^{-2}$) | 0.1057 | 0.0976 | 0.0758 | 0.0781 |
| MISC ($TOl = 10^{-2}$) | 0.1113 | 0.0940 | 0.0820 | – |
| MC method ($M = 10^6$) | 0.1079 (1.55$e$−04) | 0.0921 (9.65$e$−05) | 0.0822 (7.61$e$−05) | 0.0769 (6.65$e$−05) |

Table 13: Call option price of the different methods for different number of time steps. Case $K = 1$

**Case** $H = 0.07$

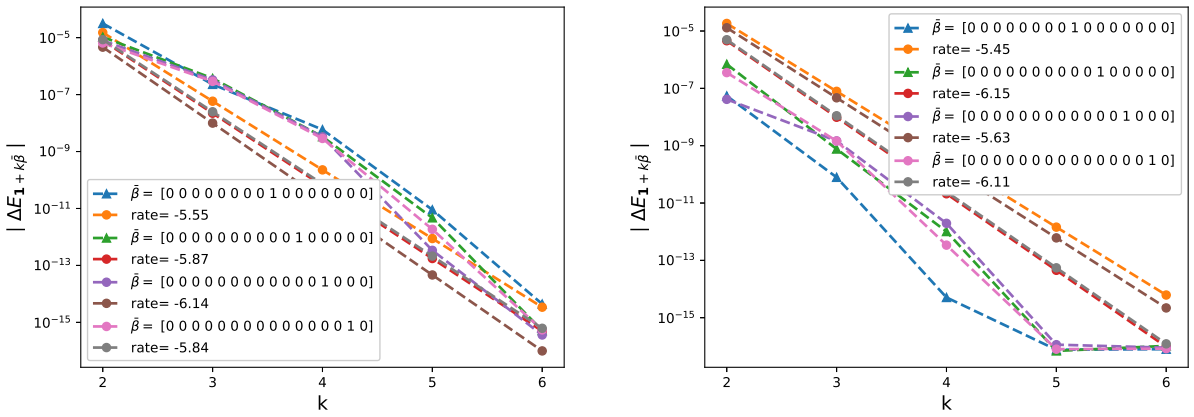
| Method \Steps | 2 | 4 | 8 | 16 |
|---|---|---|---|---|
| MISC ($TOl = 5.10^{-1}$) | 0.1065 | 0.0900 | 0.0809 | 0.0762 |
| MISC ($TOl = 10^{-1}$) | 0.1065 | 0.0900 | 0.0733 | 0.0956 |
| MISC ($TOl = 5.10^{-2}$) | 0.1065 | 0.0898 | 0.0881 | – |
| MISC ($TOl = 10^{-2}$) | 0.1226 | 0.1022 | 0.0933 | – |
| MC method ($M = 10^6$) | 0.1216 (1.05$e$−03) | 0.1020 (1.86$e$−04) | 0.0912 (1.35$e$−04) | 0.0854 (1.08$e$−04) |

Table 14: Call option price of the different methods for different number of time steps. Case $K = 1$

## B.6 Investigating differences wrt $H$

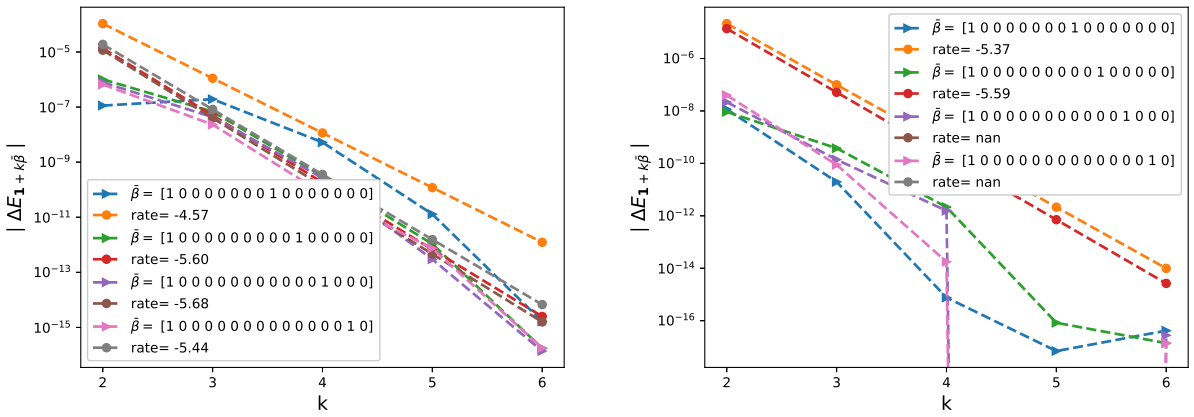
### B.6.1 Totally Hierarchical

In this section, we do both hierarchical transformation, based on brownian bridges, for both directions $W_1$ and $W_2$.




(a) (b)

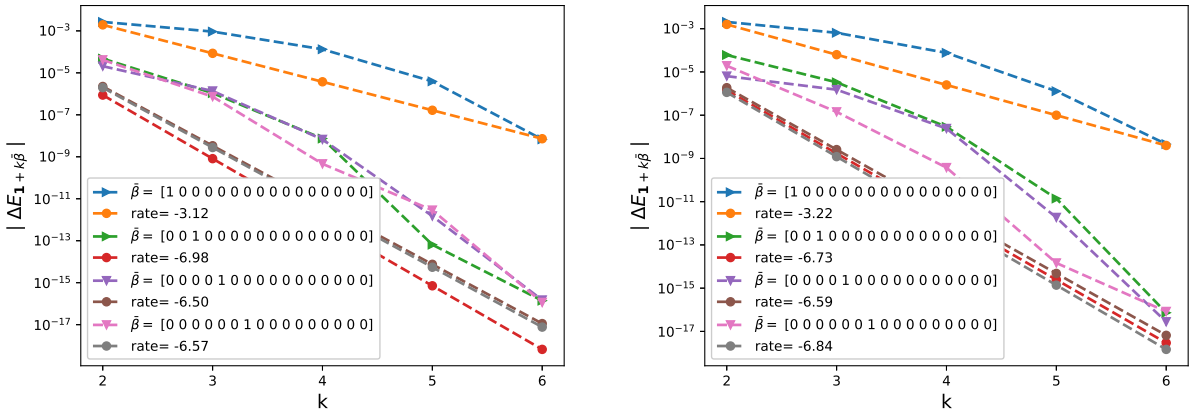
Figure 28: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) Without hierarchical for $W_2$ b) With hierarchical for $W_2$




(a) (b)

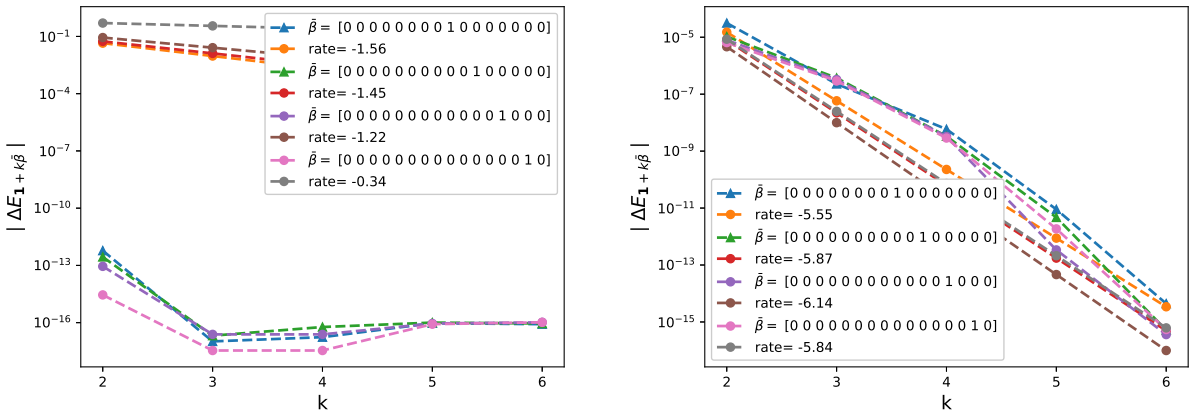
Figure 29: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) Without hierarchical for $W_2$ b) With hierarchical for $W_2$

### B.6.2 Hierarchical



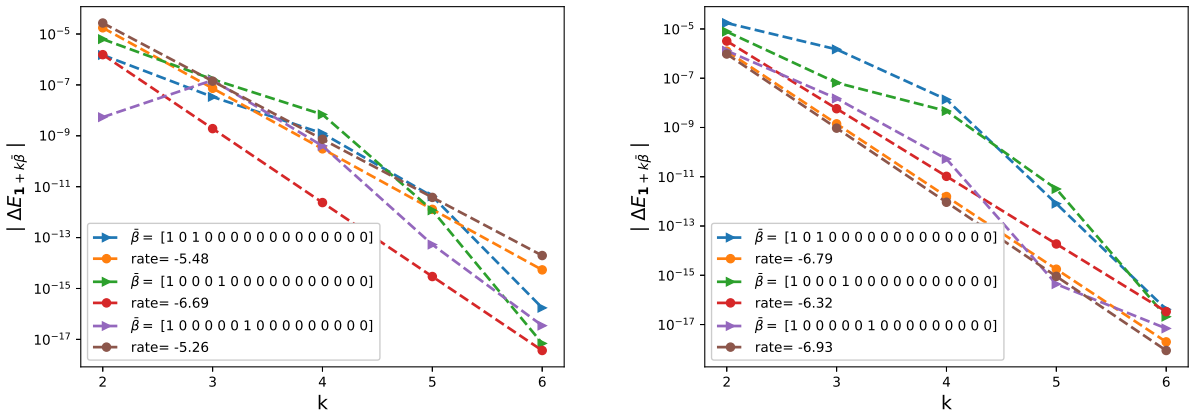

(a) (b)

Figure 30: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$



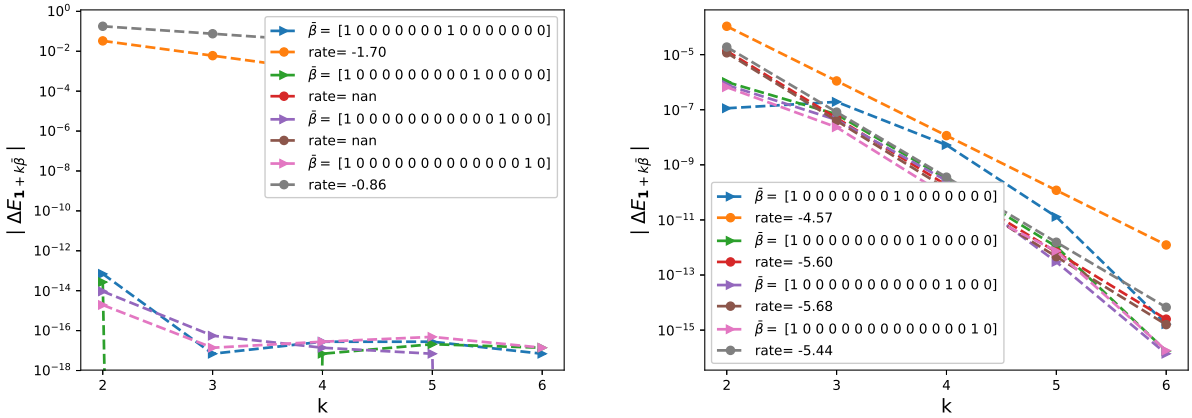

(a) (b)

Figure 31: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$




(a) (b)

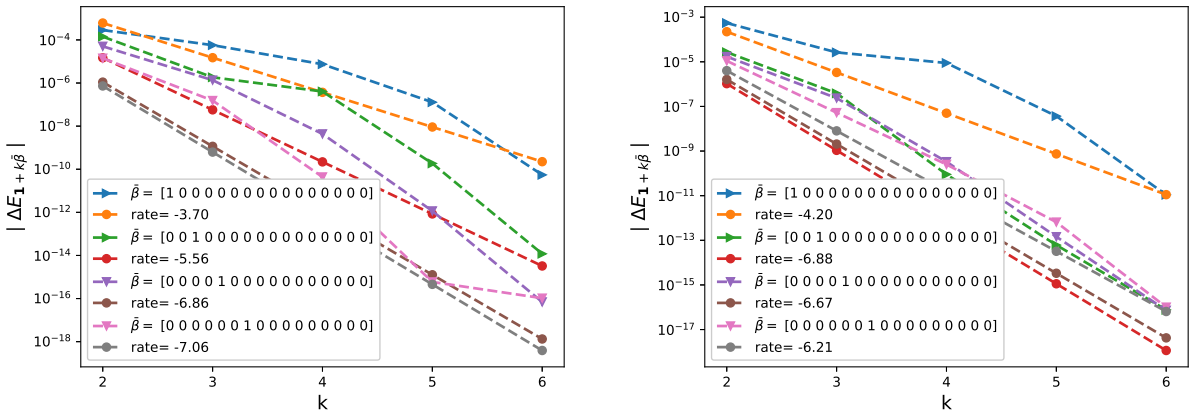
Figure 32: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$




(a) (b)

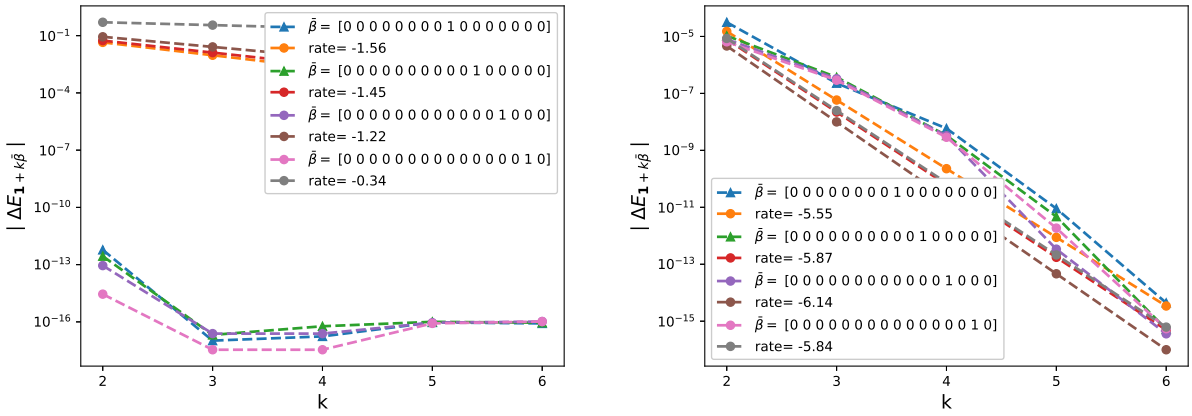
Figure 33: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$

### B.6.3 Non Hierarchical




(a) (b)

Figure 34: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta}))$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$

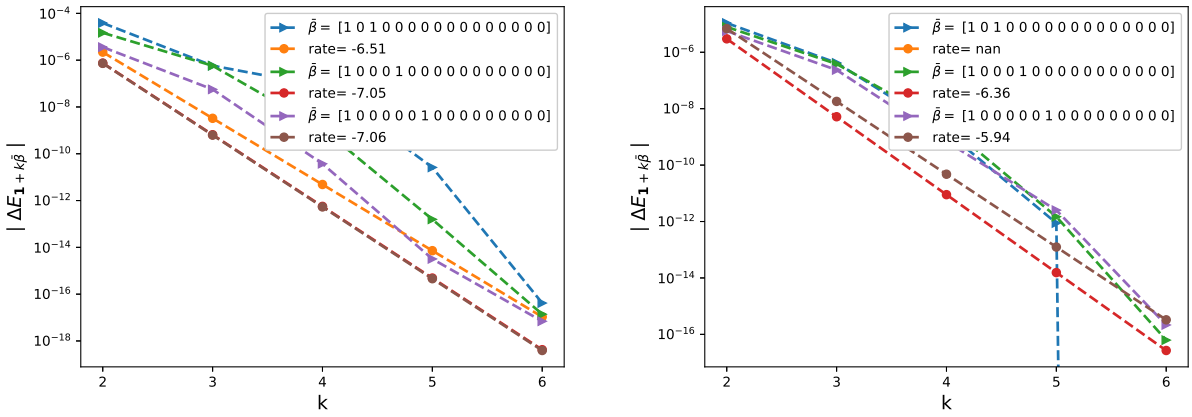



(a) (b)

Figure 35: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta}))$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$

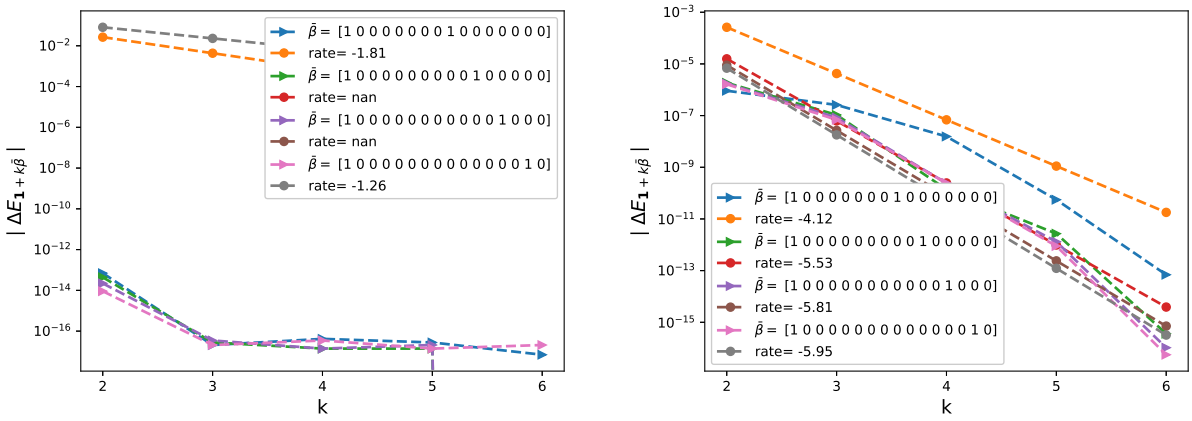
## B.7 Investigating mixed differences wrt $\rho$

**N=4, K=1**




(a) (b)

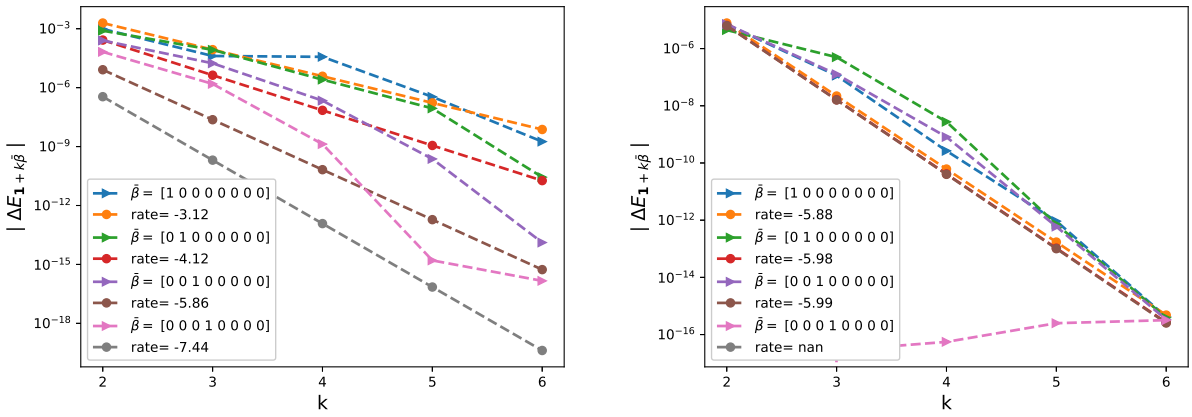
Figure 36: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\overline{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$




(a) (b)

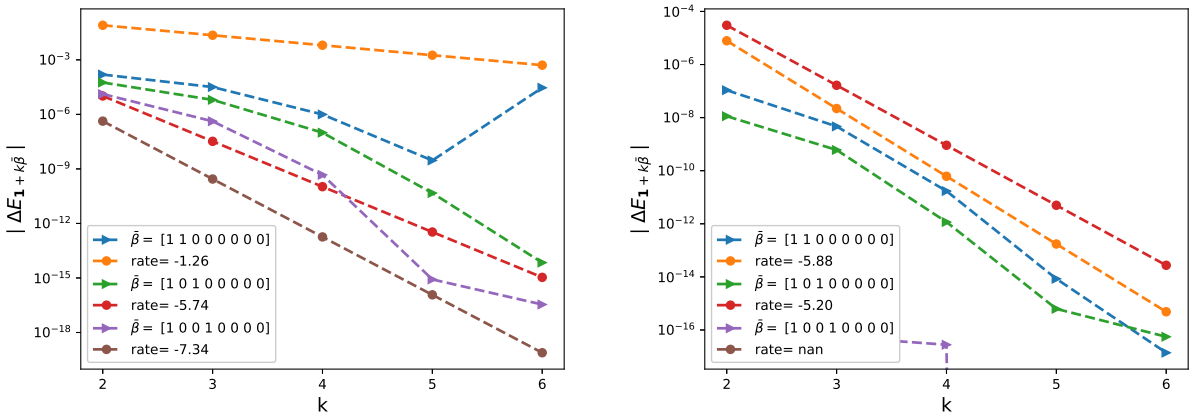
Figure 37: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\overline{\beta})$ for $K = 1$: a) $H = 0.43$ b) $H = 0.07$

**N=8, K=1**




(a) (b)

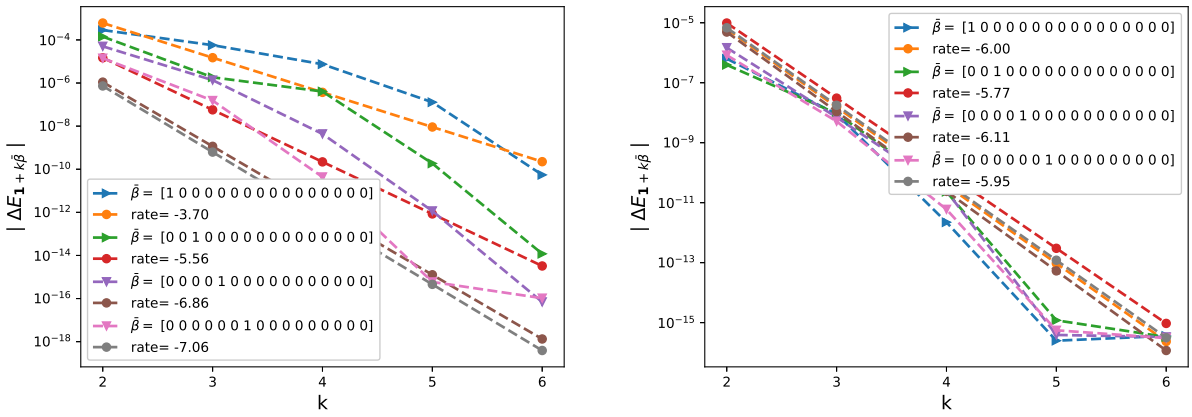
Figure 38: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$ for $K = 1$: a) $\rho = -0.9$ b) $\rho = 0$.




(a) (b)

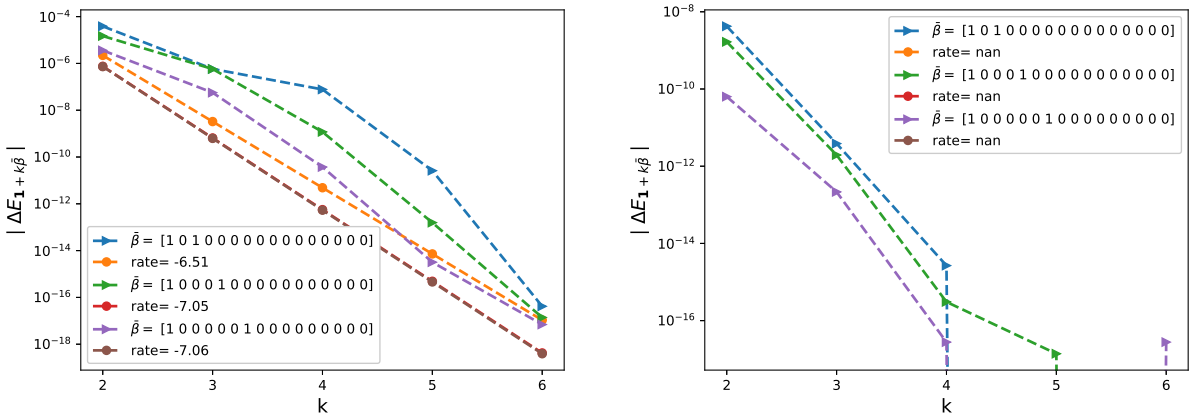
Figure 39: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$: a) $\rho = -0.9$ b) $\rho = 0$.

**N=4, K=0.8**




(a) (b)

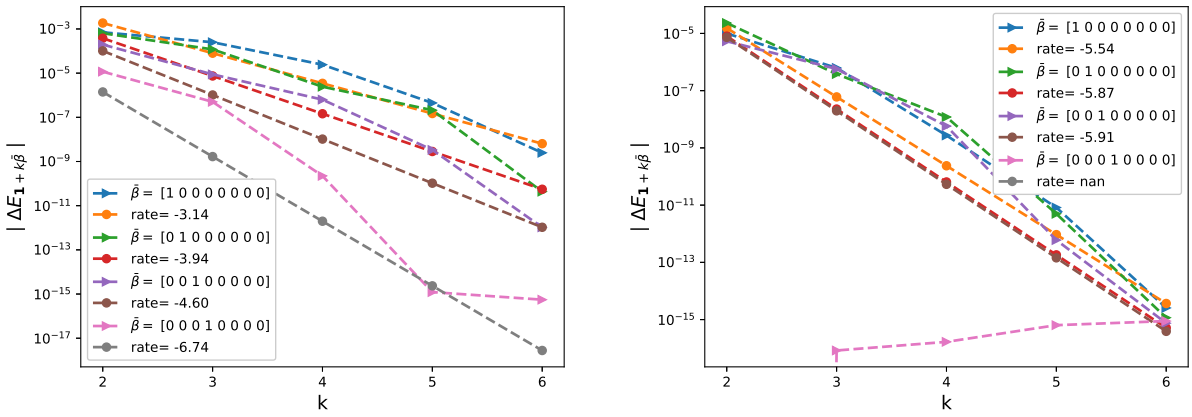
Figure 40: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.




(a) (b)

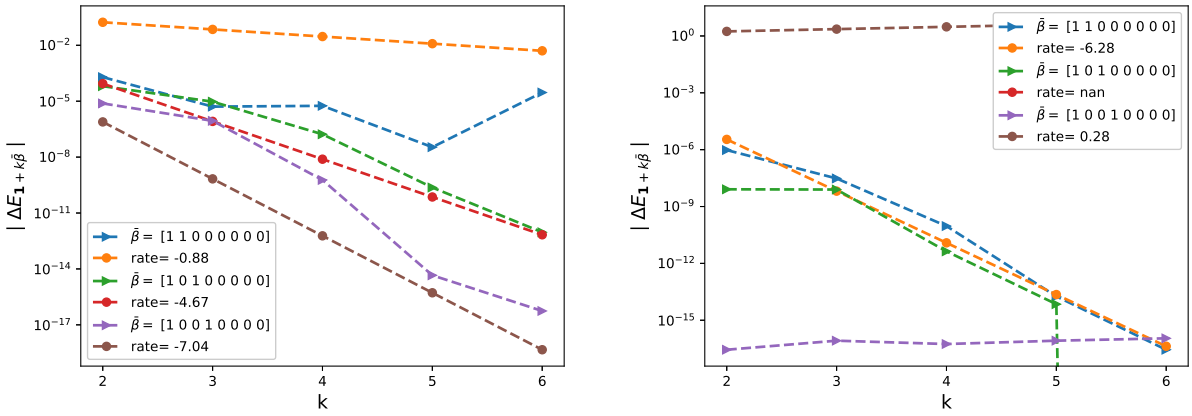
Figure 41: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.

**N=8, K=0.8**




(a) (b)

Figure 42: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.

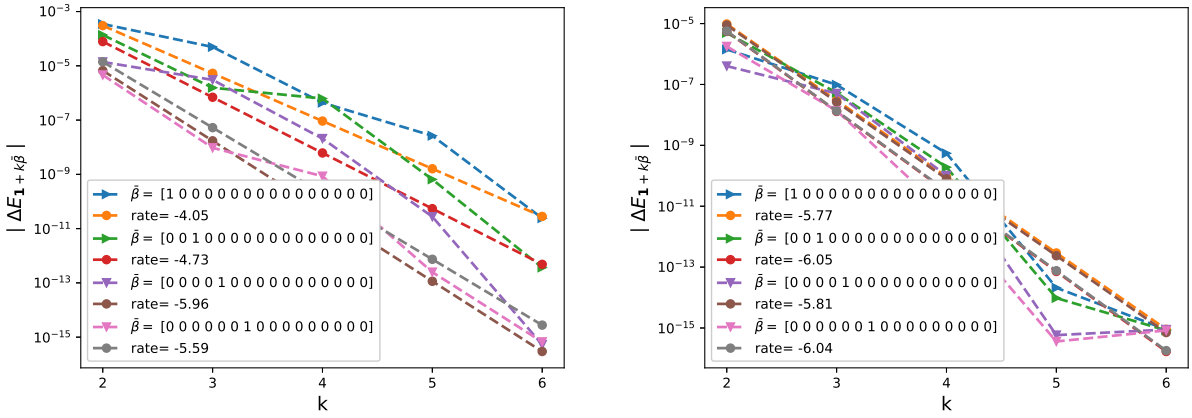



(a) (b)

Figure 43: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.

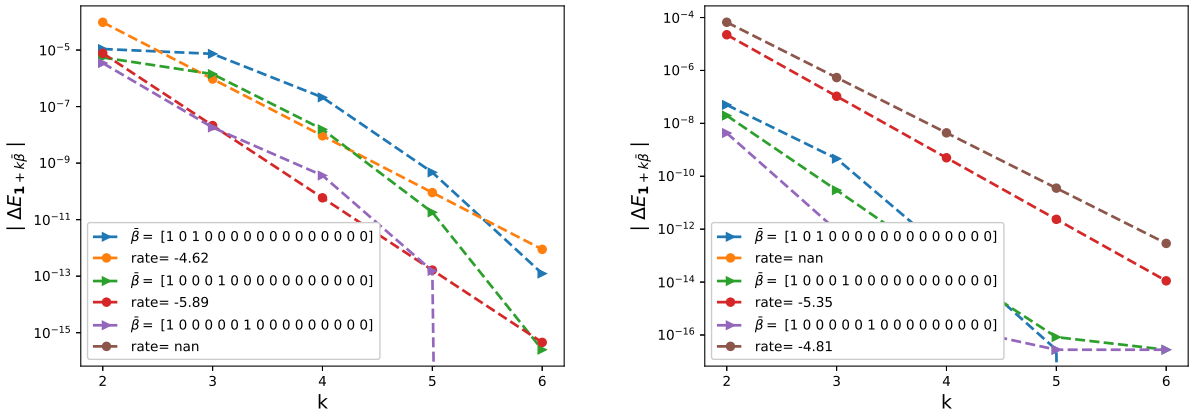
## B.8 Investigating mixed differences wrt $\xi$

**N=4, K=1**




(a) (b)

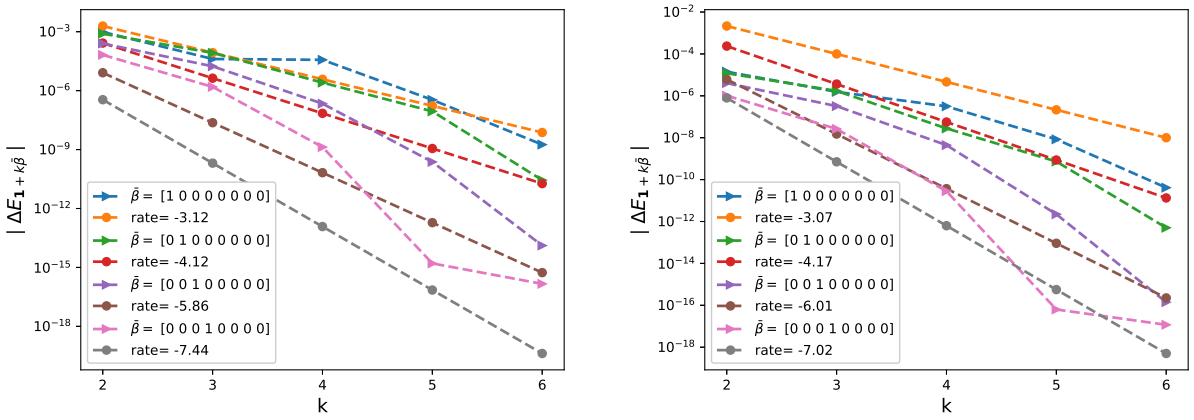
Figure 44: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.




(a) (b)

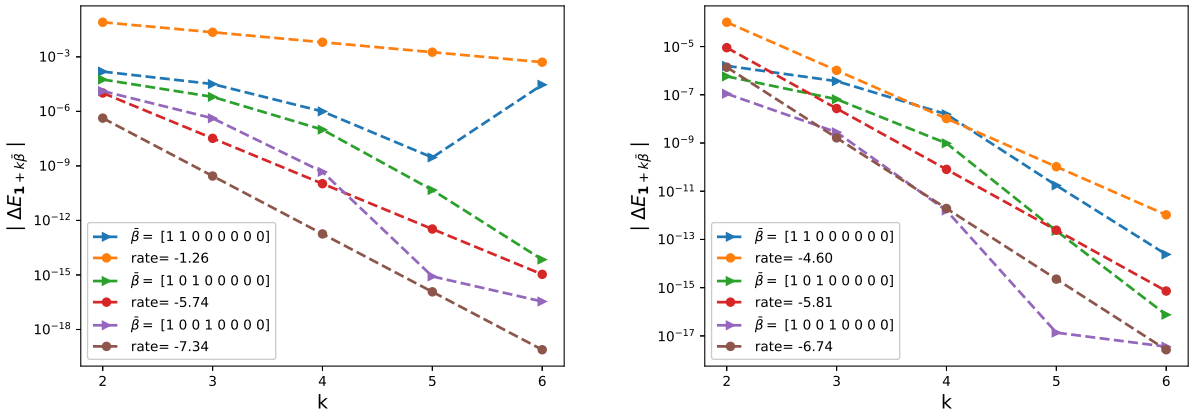
Figure 45: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\rho = -0.9$ b) $\rho = 0$.

**N=8, K=1**




(a) (b)

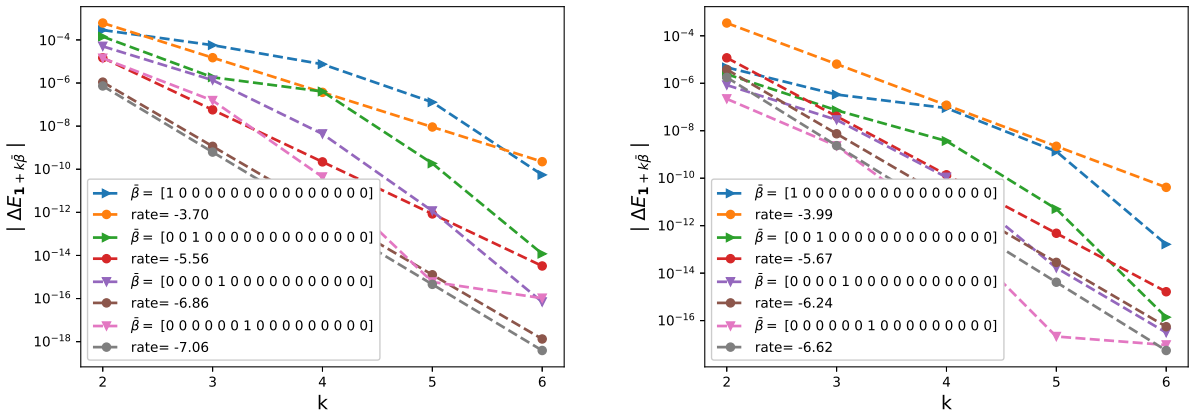
Figure 46: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$




(a) (b)

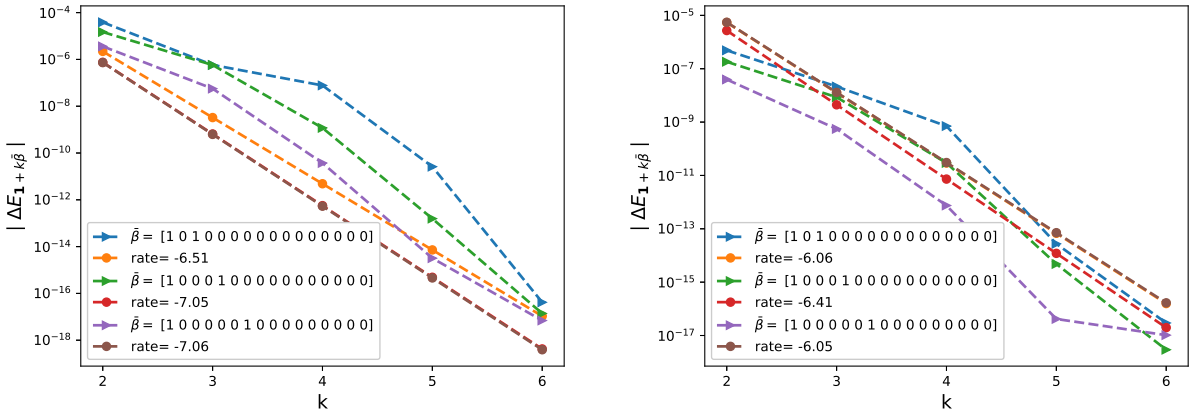
Figure 47: The rate of convergence of mixed order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$

**N=4, K=0.8**




(a) (b)

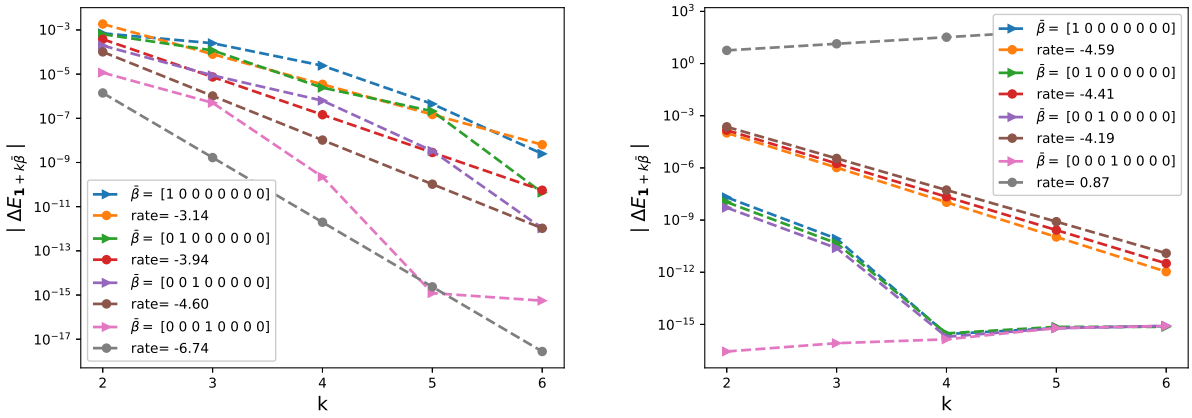
Figure 48: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$




(a) (b)

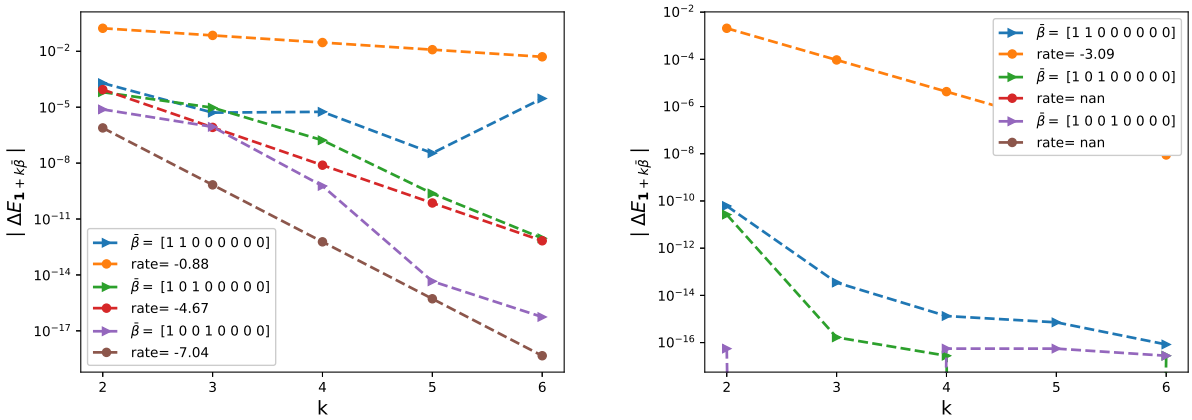
Figure 49: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$

**N=8, K=0.8**




(a) (b)

Figure 50: The rate of convergence of first order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$



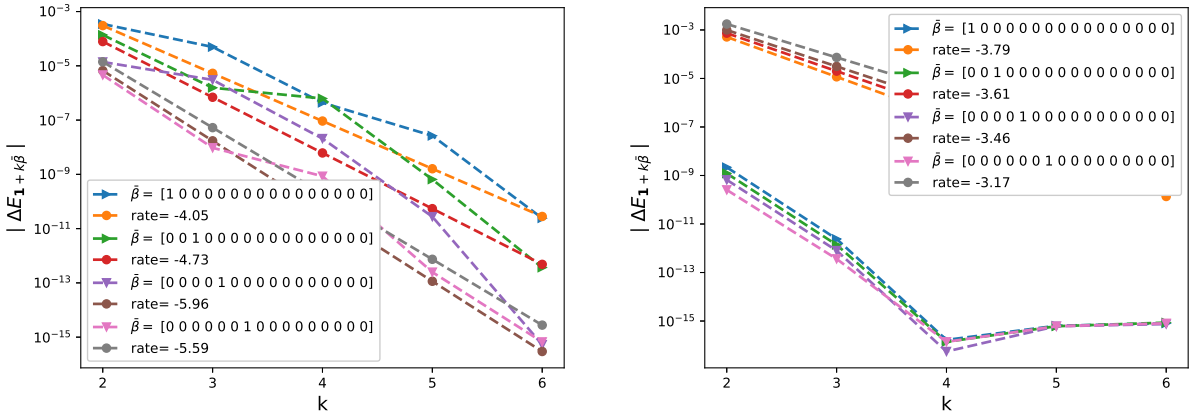

(a) (b)

Figure 51: The rate of convergence of mixed order differences $|\Delta E_\beta|$ $(\beta = \mathbf{1} + k\bar{\beta})$): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$

## B.9 Investigating mixed differences wrt moneyness $K$

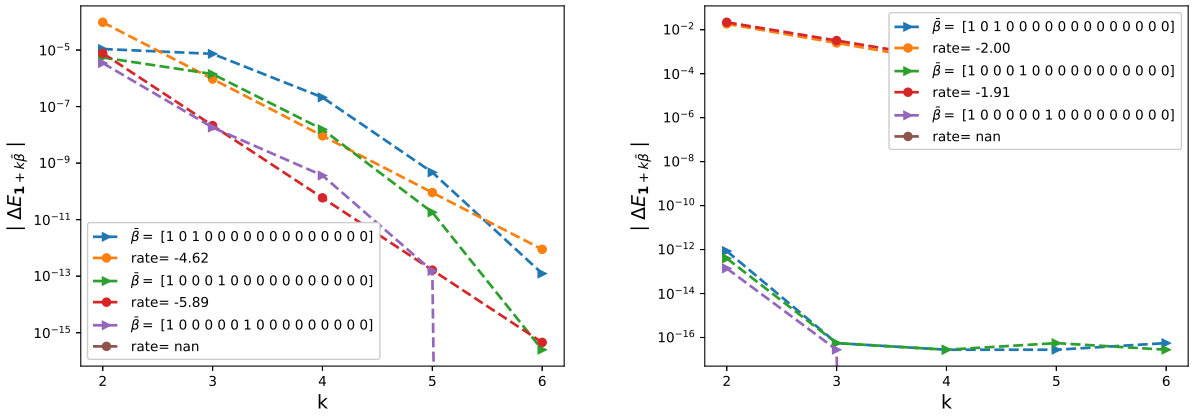
**Case** $H = 0.43$

$N = 8$




(a) (b)

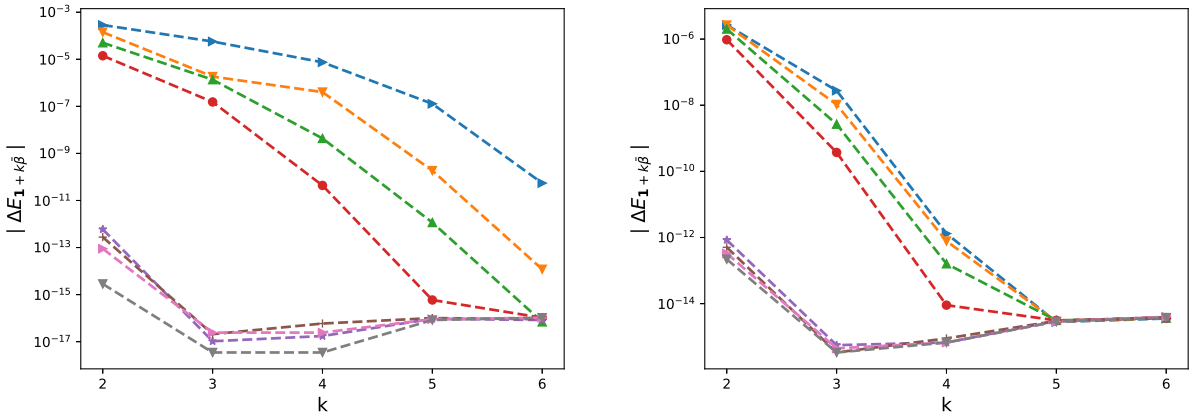
Figure 52: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$




(a) (b)

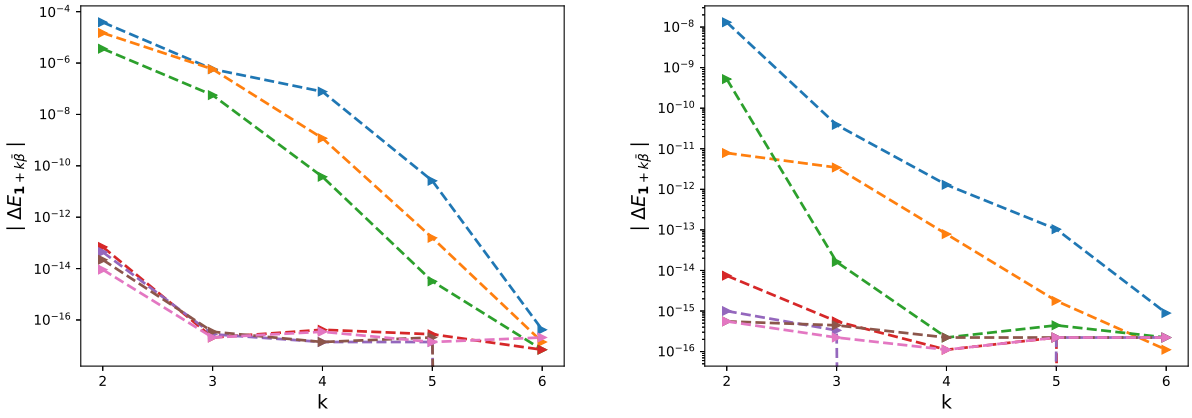
Figure 53: The rate of convergence of mixed order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $\xi = 0.235^2$ b) $\xi = 10^{-5}$

$N = 16$




(a) (b)

Figure 54: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.

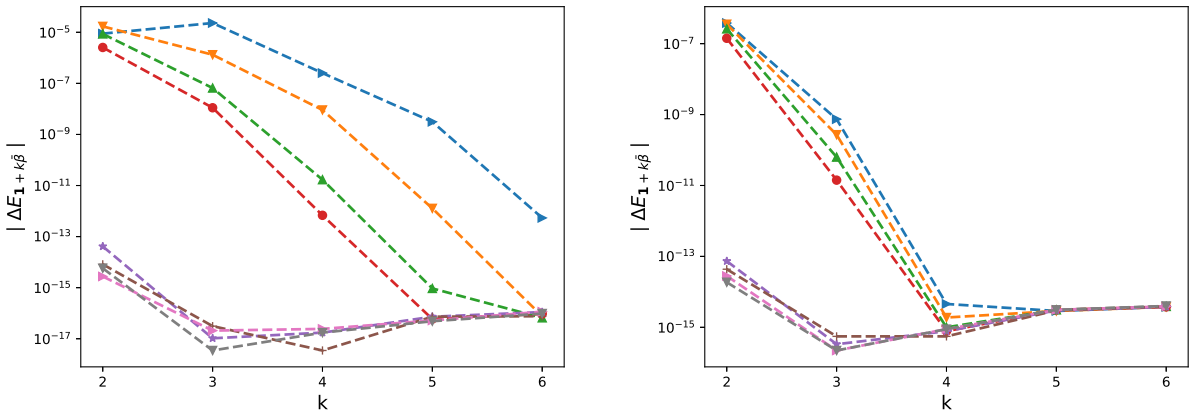



(a) (b)

Figure 55: The rate of convergence of second order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.

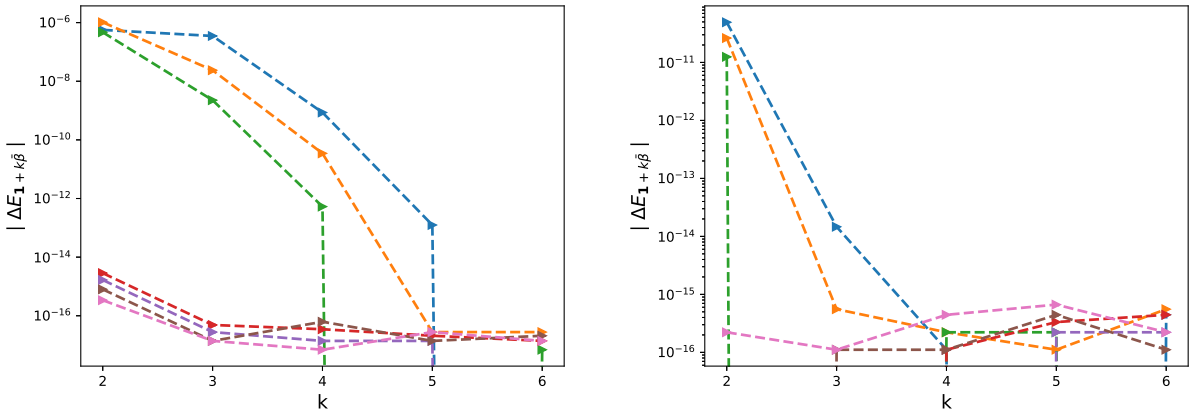
**Case** $H = 0.07$

$N = 8$




(a) (b)

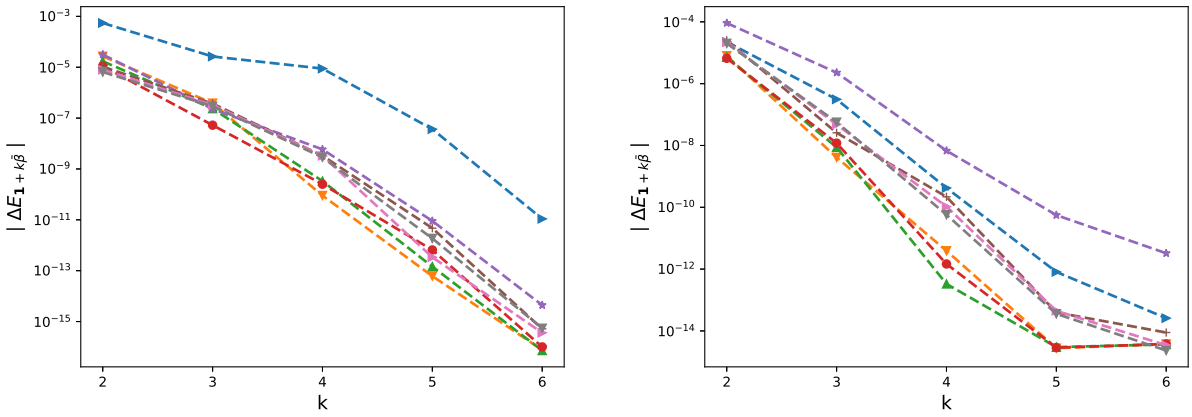
Figure 56: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.




(a) (b)

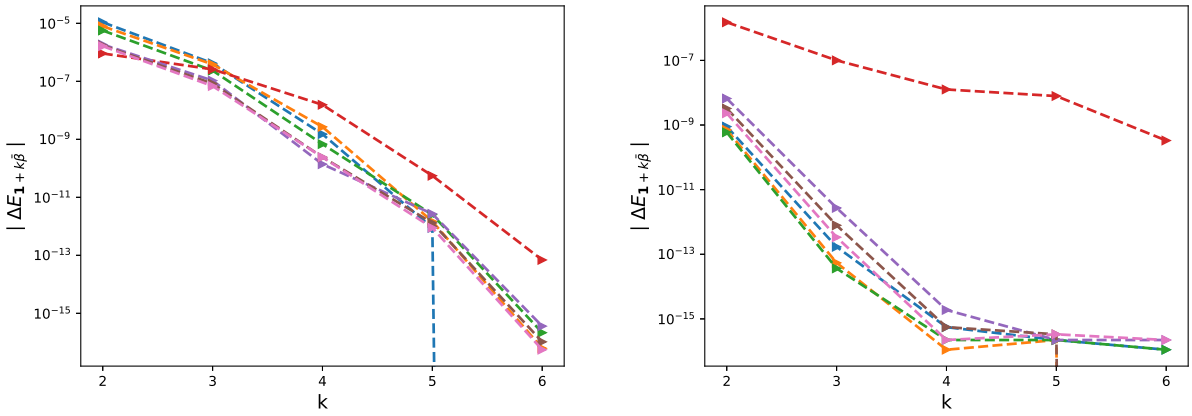
Figure 57: The rate of convergence of second order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.

$N = 16$




(a) (b)

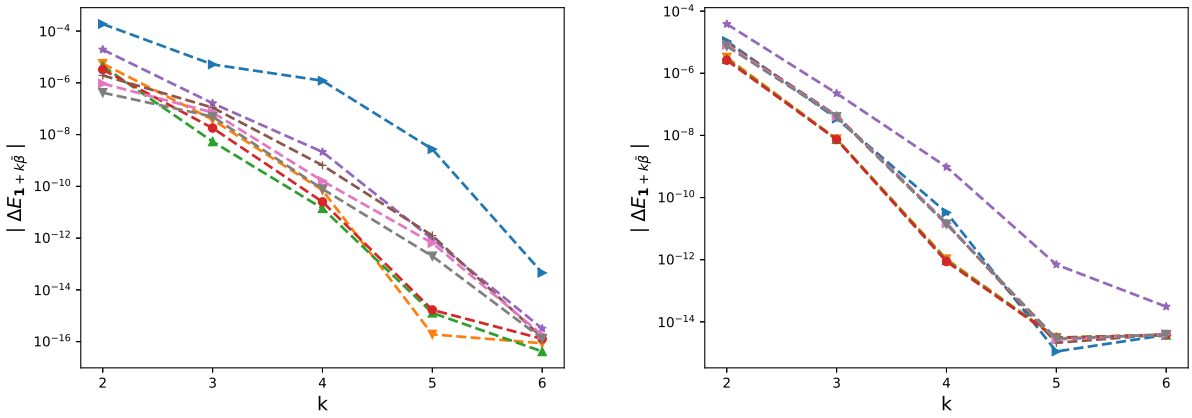
Figure 58: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.




(a) (b)

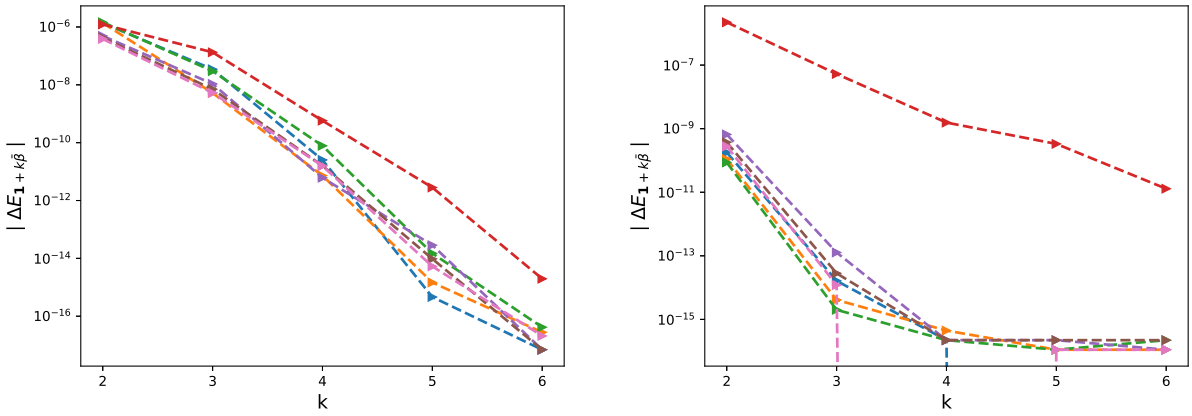
Figure 59: The rate of convergence of second order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.

## B.10 Convergence plots using MISC ($H = 0.43$)




(a) (b)

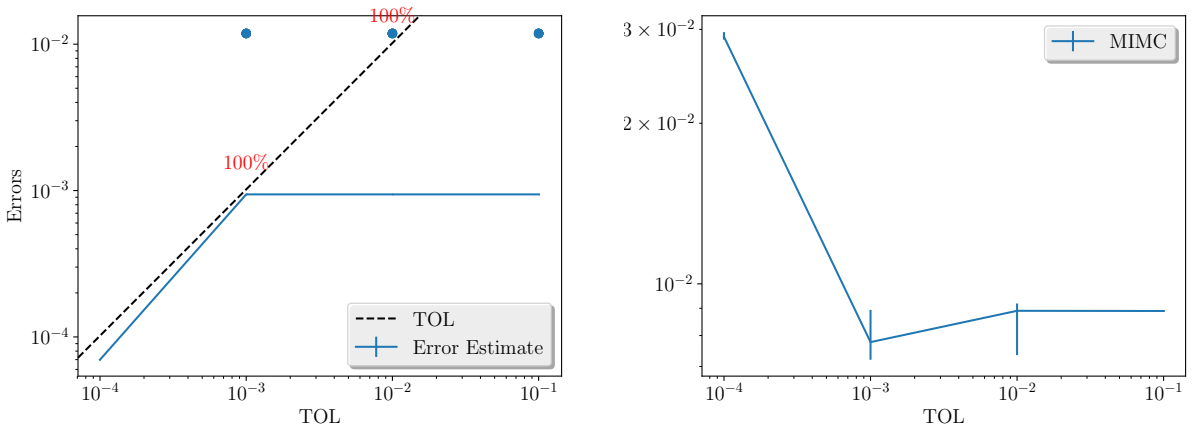
Figure 60: The rate of convergence of first order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.




(a) (b)

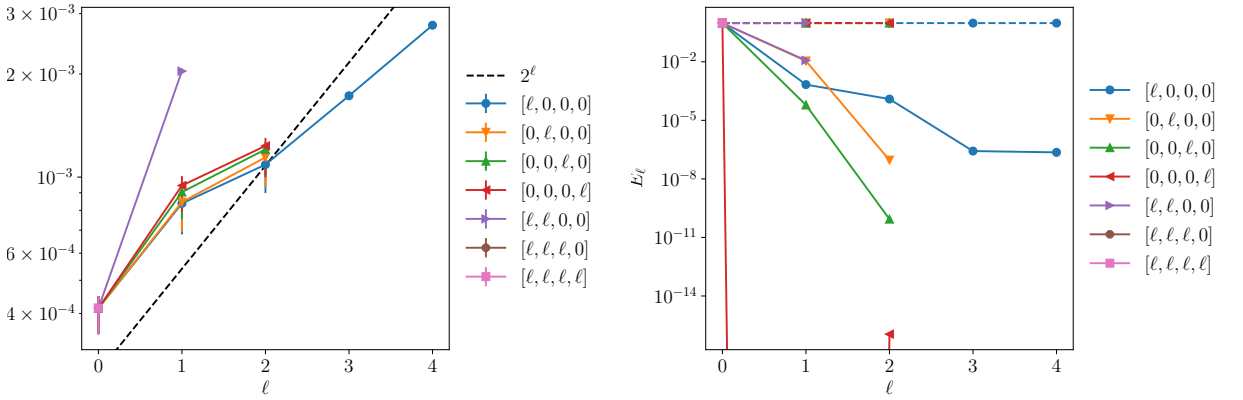
Figure 61: The rate of convergence of second order differences $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\bar{\beta}$)): a) $K = 1$ b) $K = \exp(-4)$.

**Case of 2 time steps, $K = e^{-4}$**




(a) Error estimate (b) Average running time as a function of $TOL$

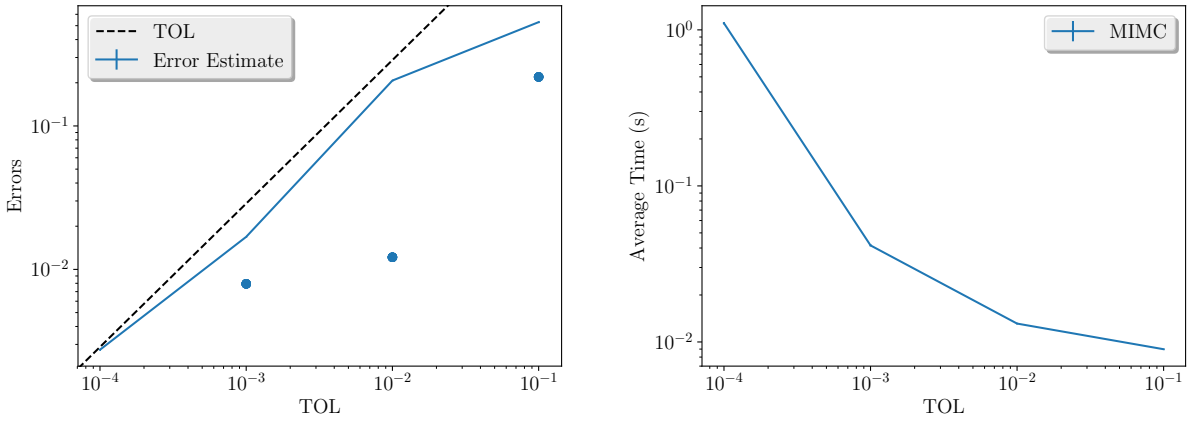
Figure 62: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

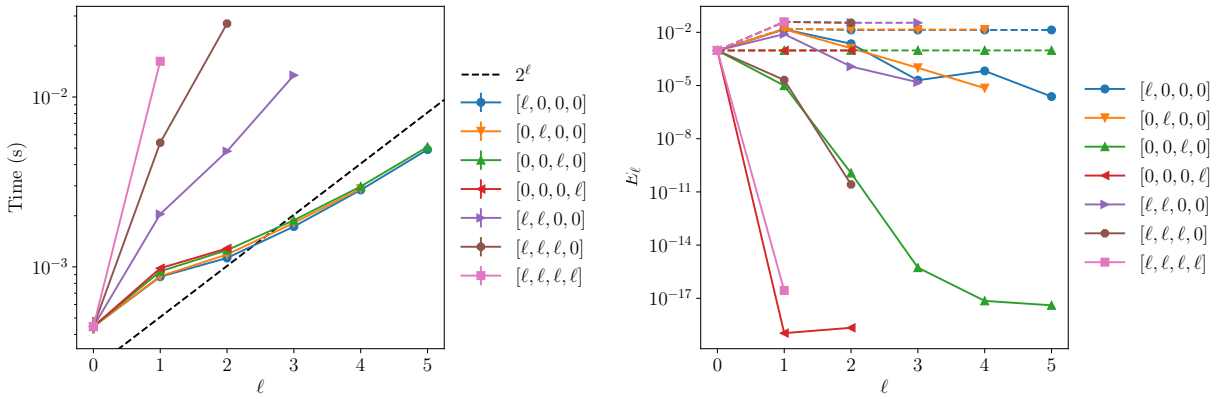
Figure 63: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of 2 time steps,** $K = 1.2$




(a) Error estimate (b) Average running time as a function of $TOL$

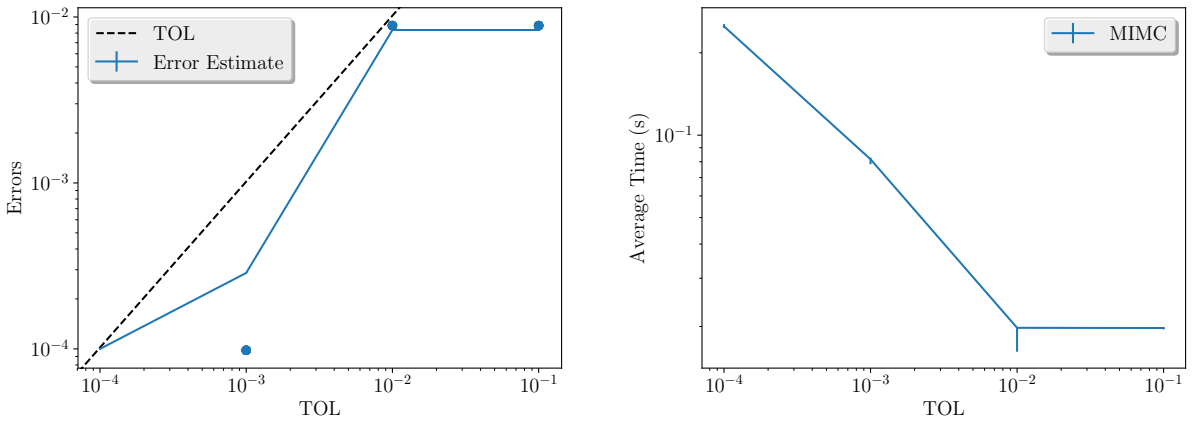
Figure 64: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

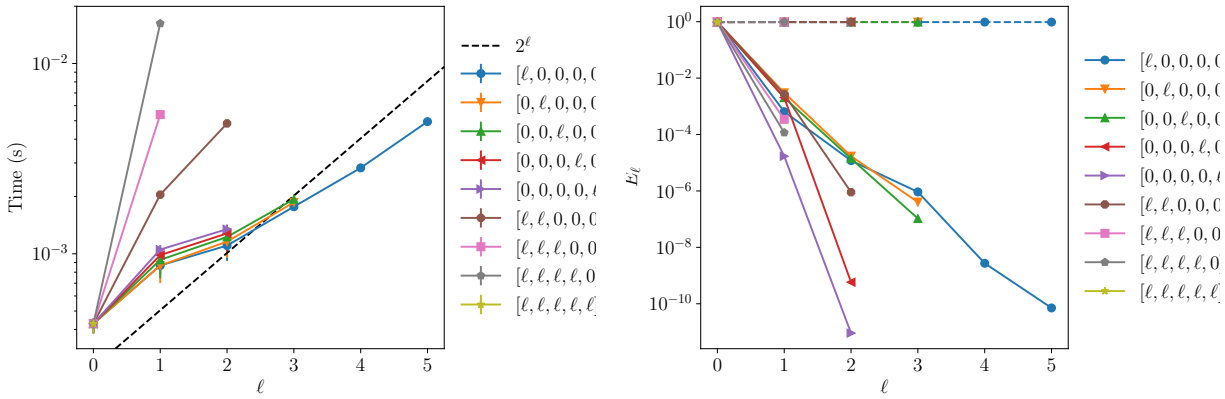
Figure 65: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of 4 time steps, $K = e^{-4}$**




(a) Error estimate (b) Average running time as a function of $TOL$

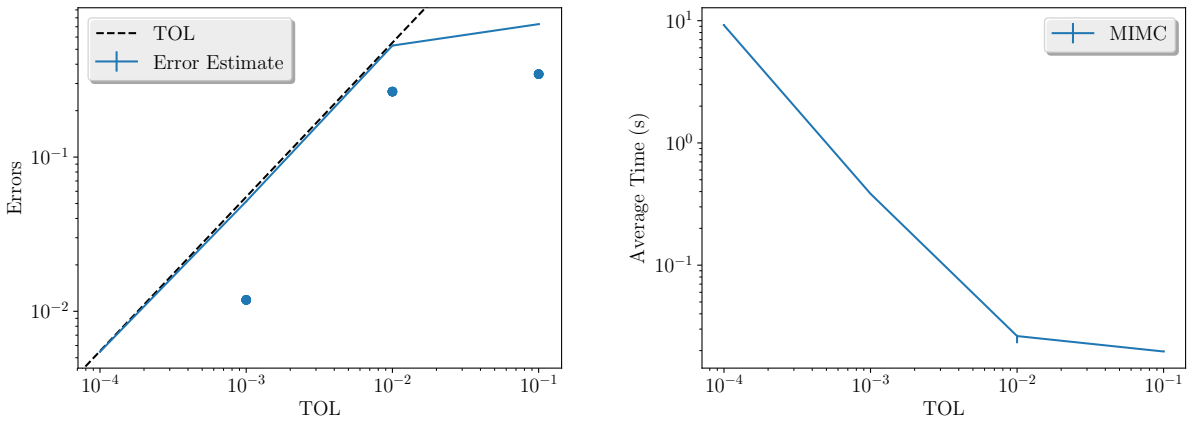
Figure 66: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

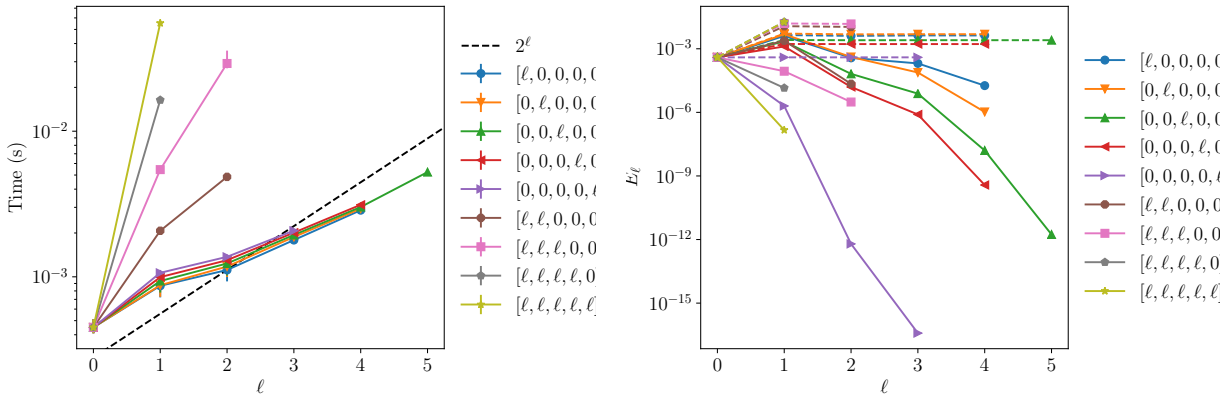
Figure 67: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of 4 time steps,** $K = 1.2$




(a) Error estimate (b) Average running time as a function of $TOL$

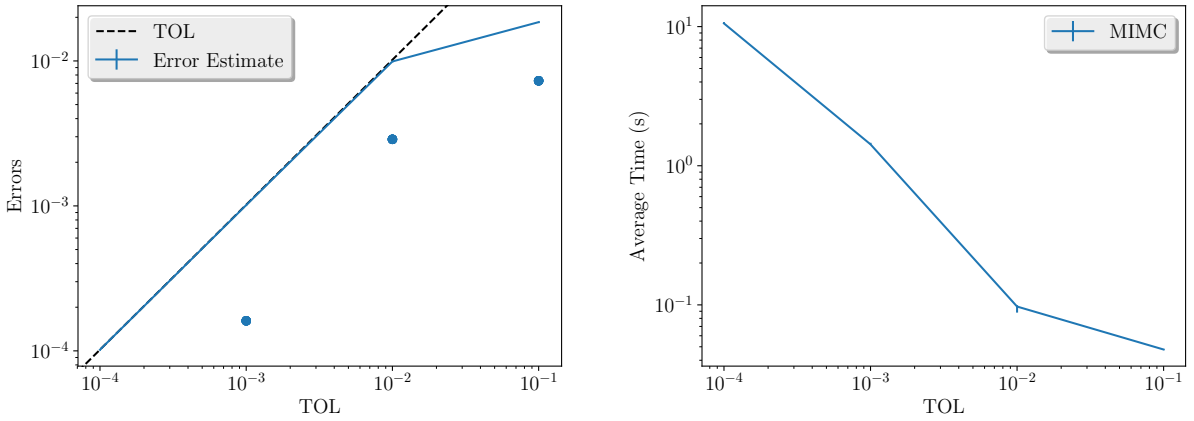
Figure 68: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

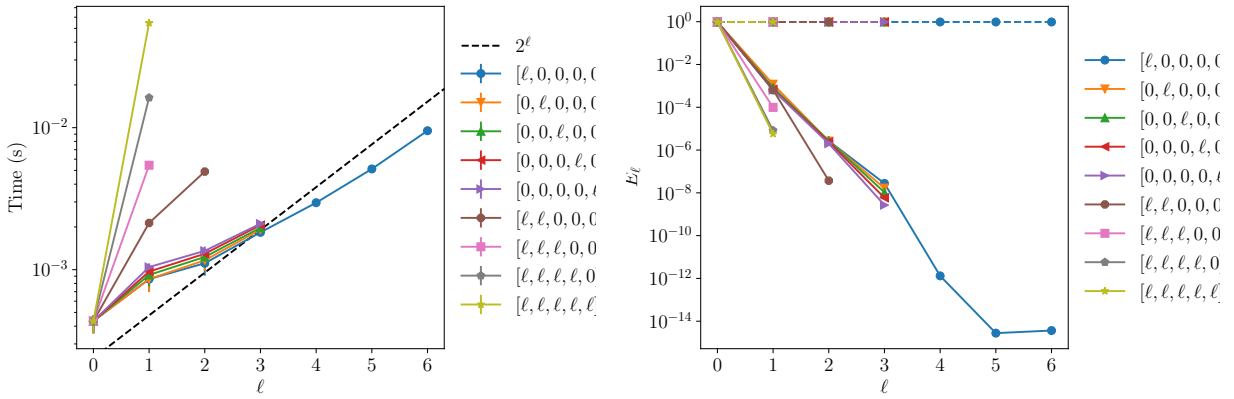
Figure 69: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of 8 time steps, $K = e^{-4}$**




(a) Error estimate (b) Average running time as a function of $TOL$

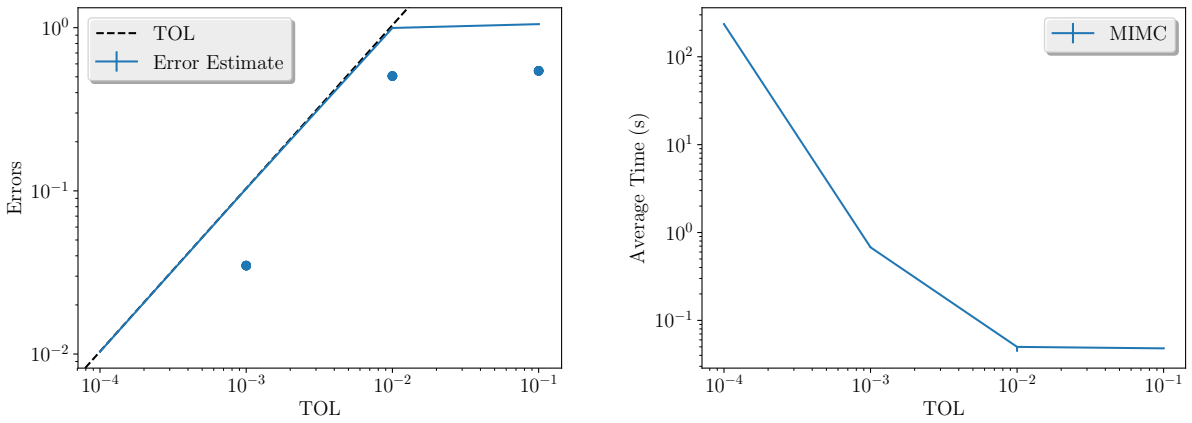
Figure 70: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

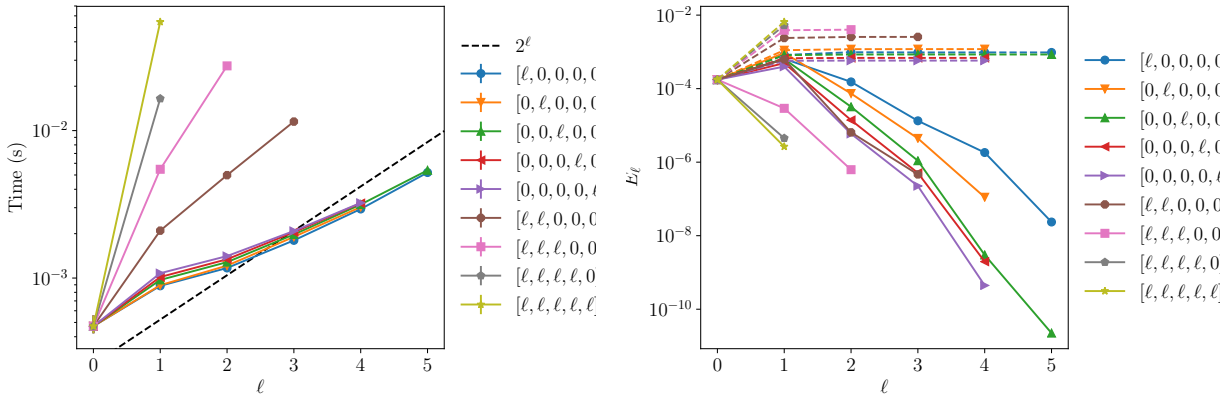
Figure 71: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of 8 time steps, $K = 1.2$**




(a) Error estimate (b) Average running time as a function of $TOL$

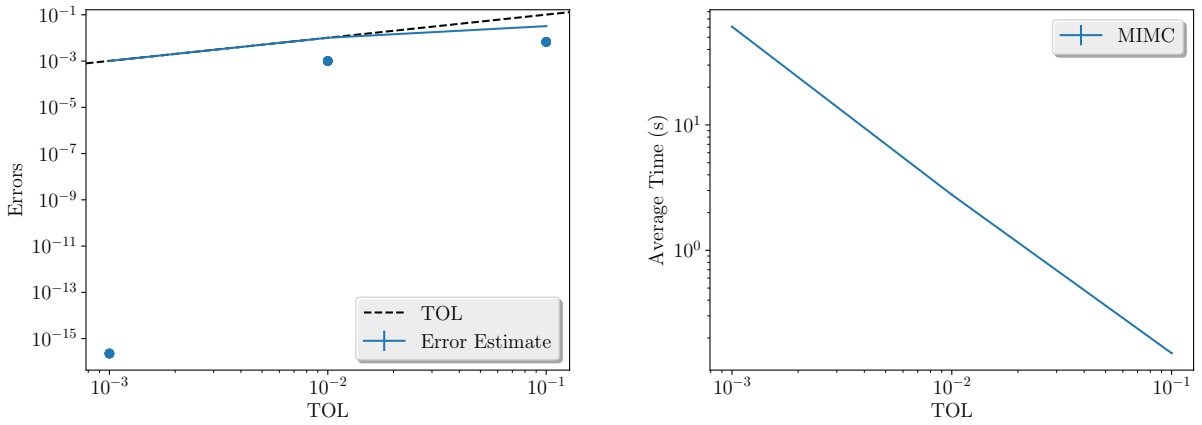
Figure 72: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

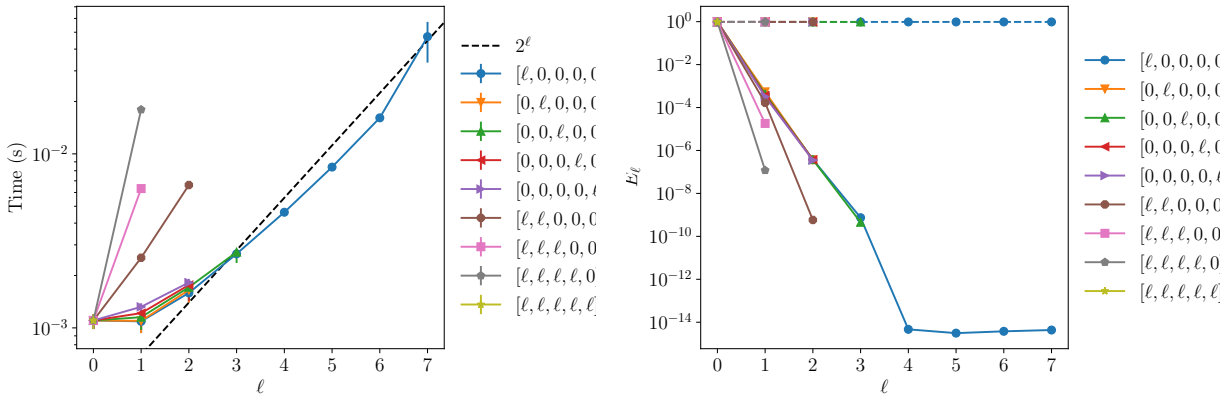
Figure 73: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of** 16 **time steps,** $K = e^{-4}$




(a) Error estimate (b) Average running time as a function of $TOL$

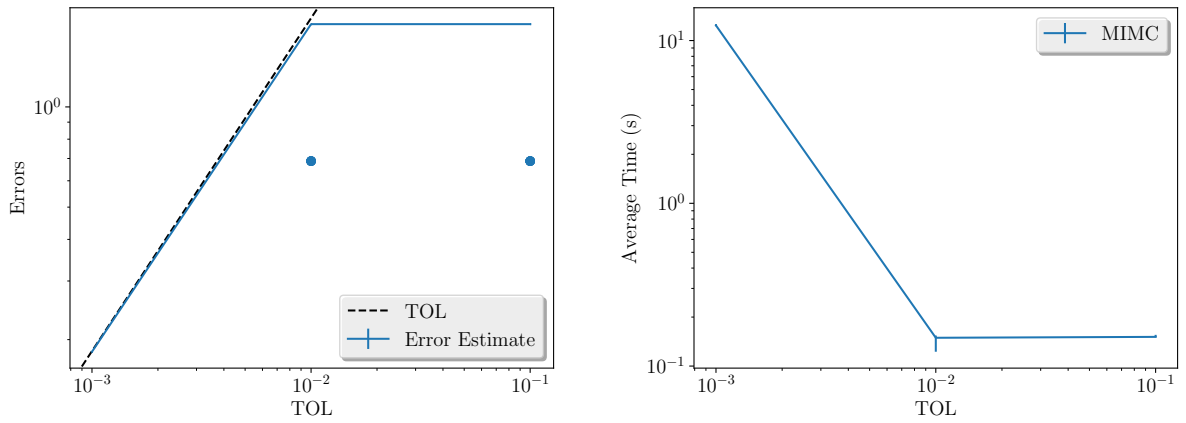
Figure 74: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

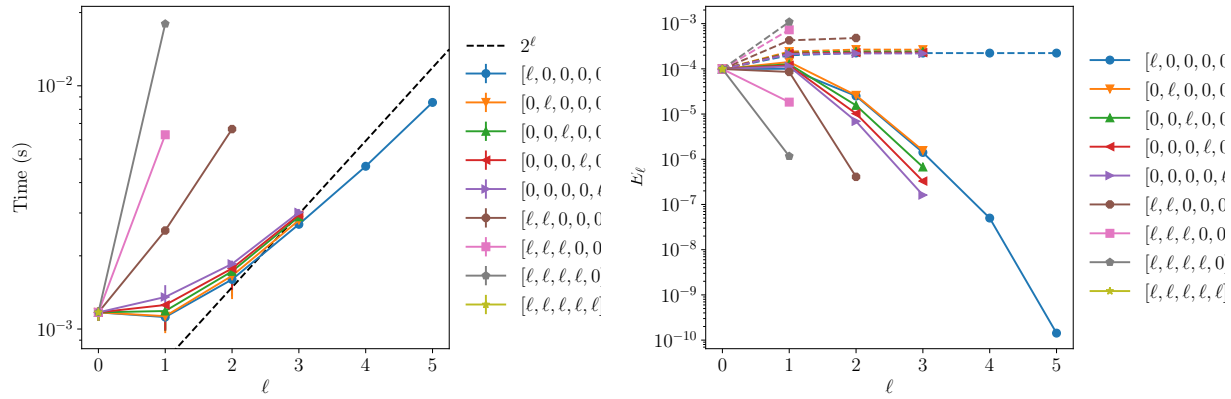
Figure 75: Convergence and work rates for discretization levels the call payoff with rBergomi model.

**Case of** 16 **time steps,** $K = 1.2$




(a) Error estimate (b) Average running time as a function of $TOL$

Figure 76: Convergence and complexity results for the call payoff with rBergomi model.




(a) Average Computational time per level (b) The convergence rate of mixed differences per level

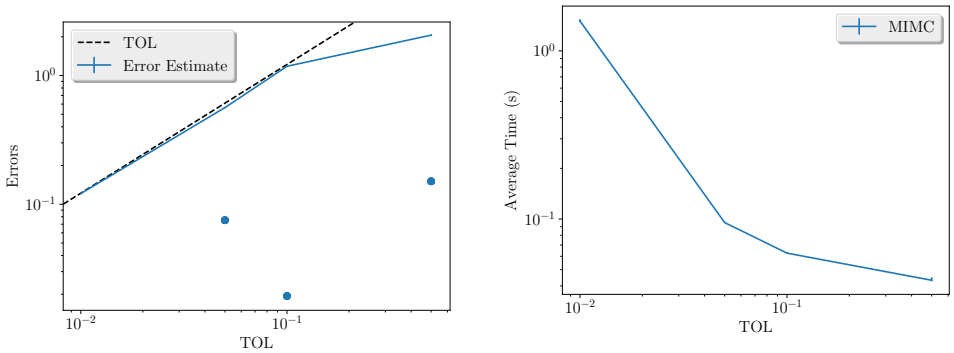
Figure 77: Convergence and work rates for discretization levels the call payoff with rBergomi model.

## B.11 MISC plots

## B.12 Non Hierarchical

### H=0.43

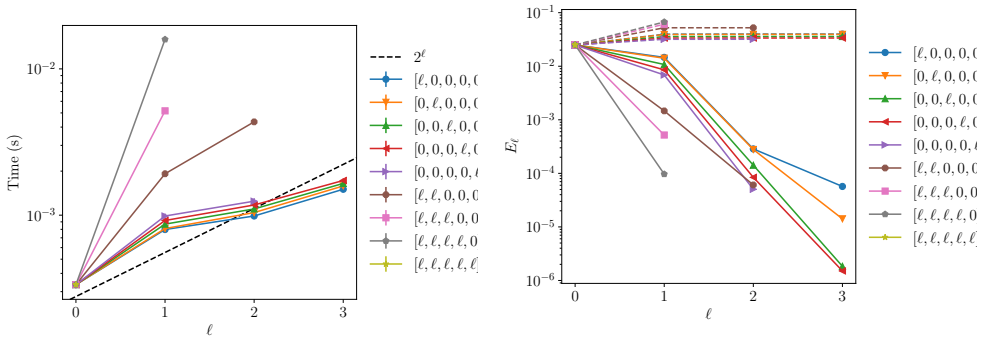
**Case of 8 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 78: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 8$.

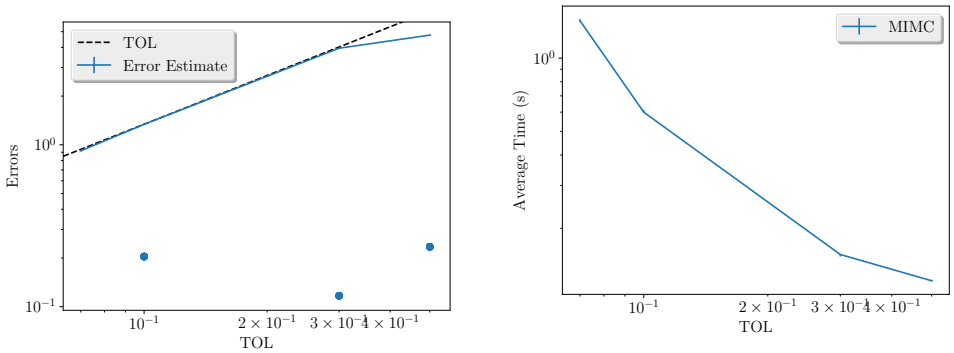



(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 79: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 8$.

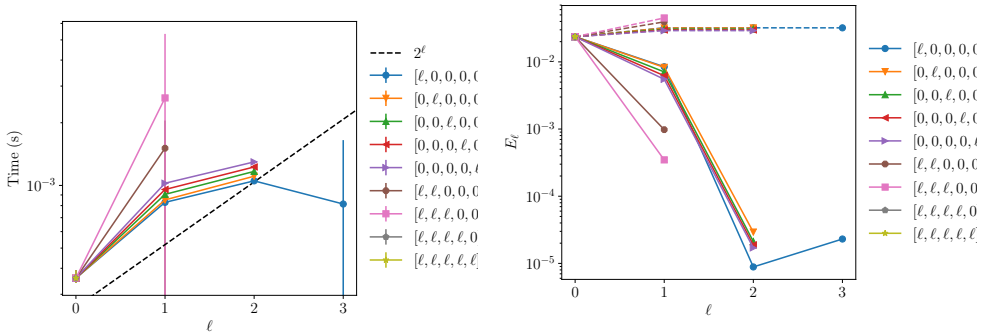
**Case of 16 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 80: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 16$.



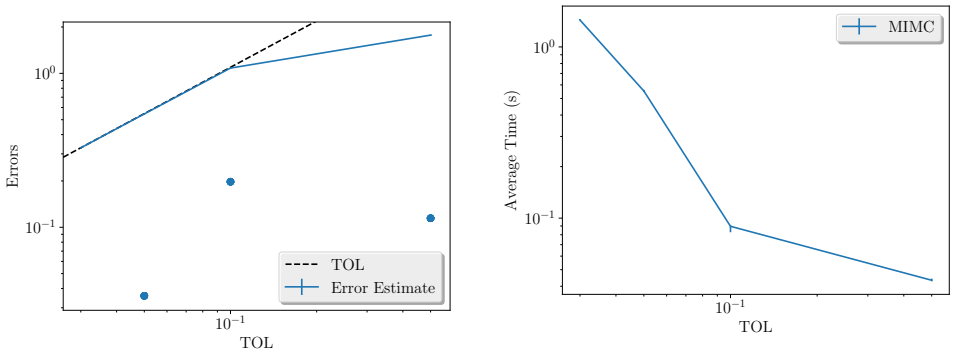

(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 81: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 16$.

**H=0.07**

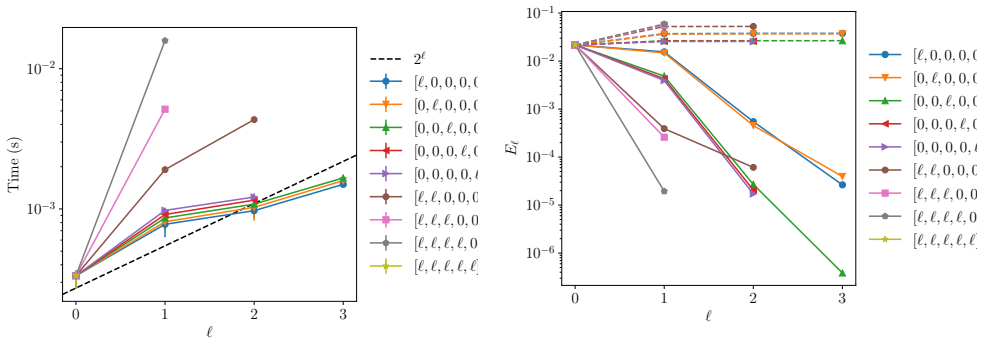
**Case of 8 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 82: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 8$.

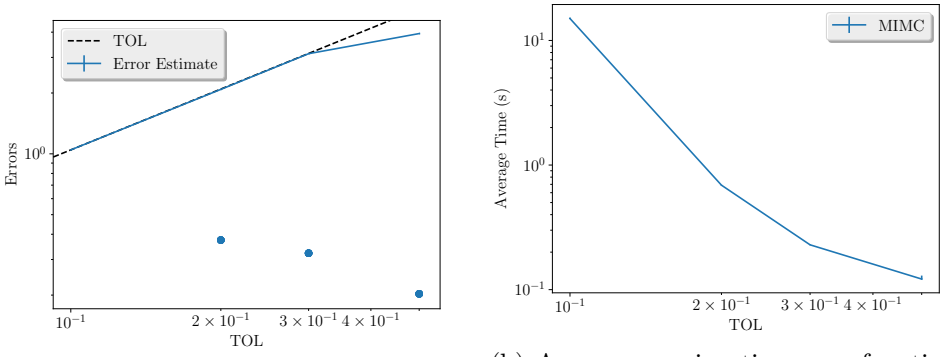



(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 83: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 8$.

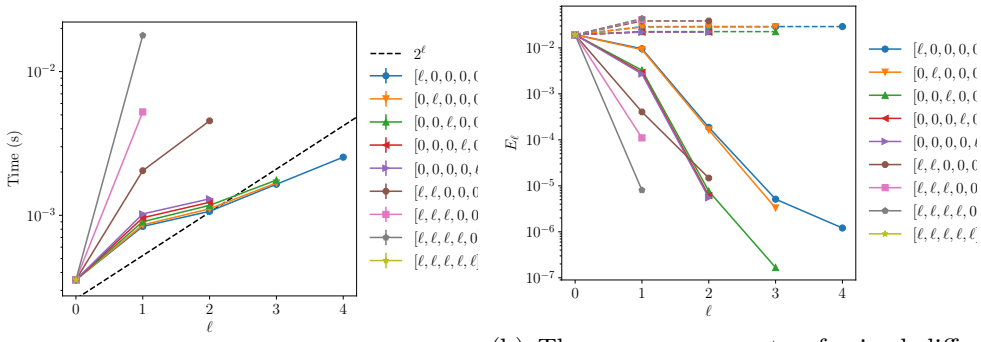
**Case of 16 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 84: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 16$.




(a) Average Computational time per level

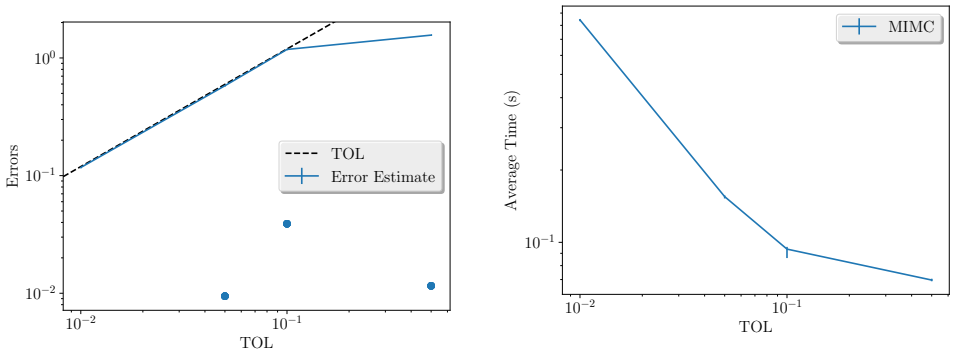
(b) The convergence rate of mixed differences per level

Figure 85: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 16$.

## B.13 Hierarchical

**H=0.43**

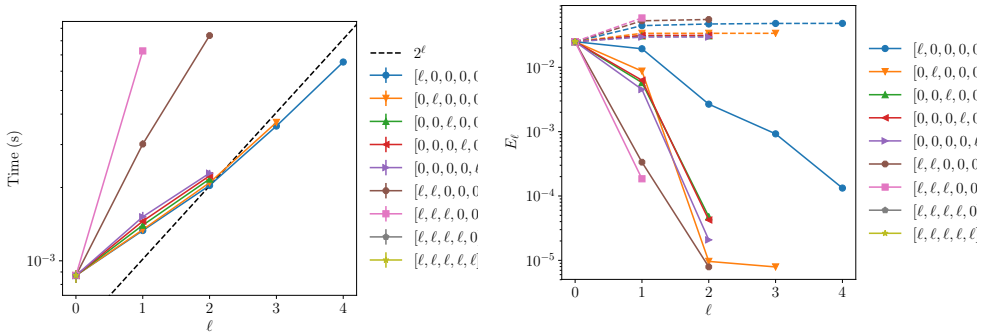
**Case of 8 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 86: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 8$.

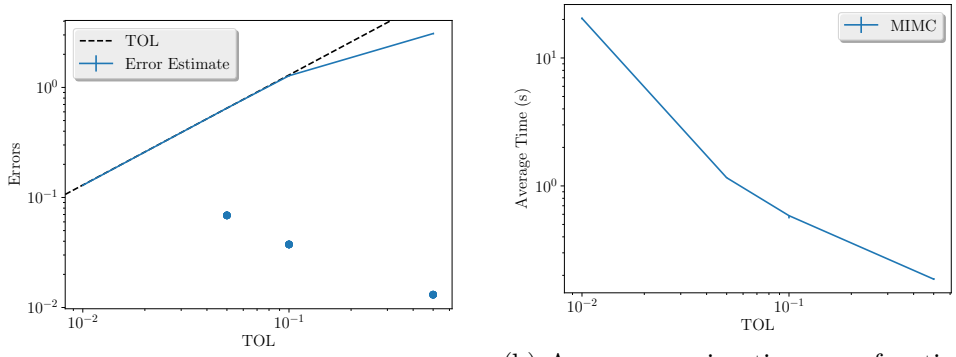



(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 87: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 8$.

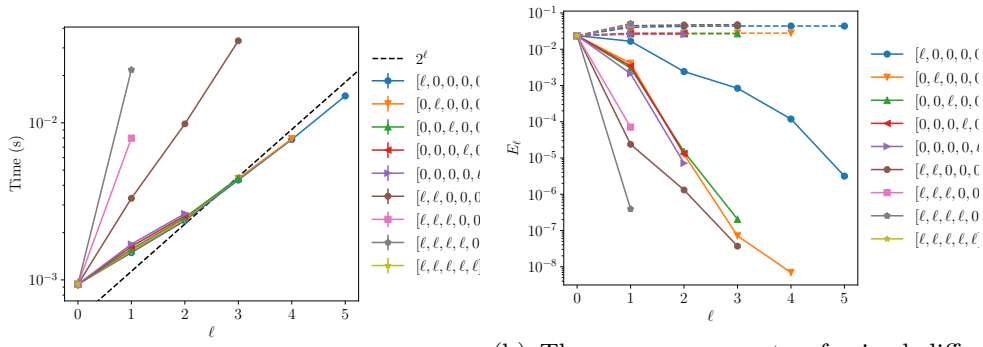
**Case of 16 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 88: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 16$.



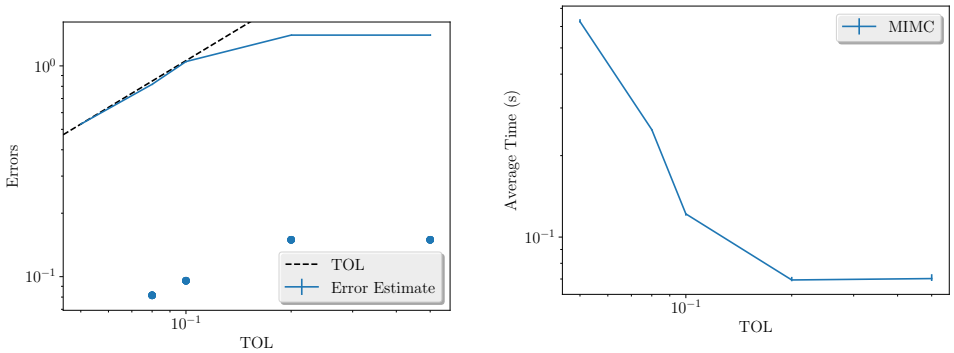

(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 89: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.43$ and $N = 16$.

**H=0.07**

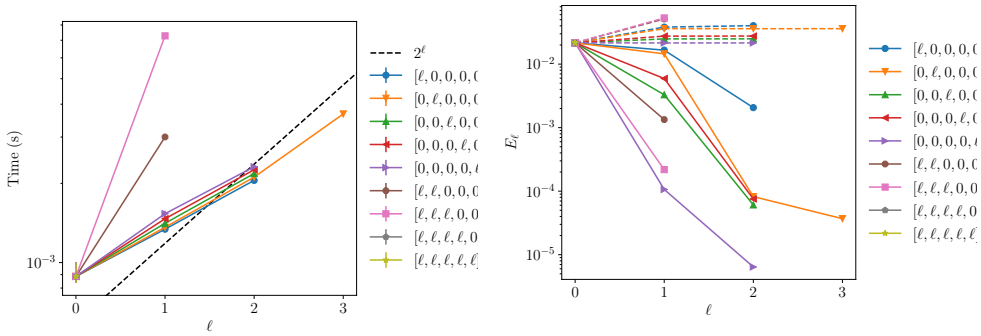
**Case of 8 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 90: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 8$.

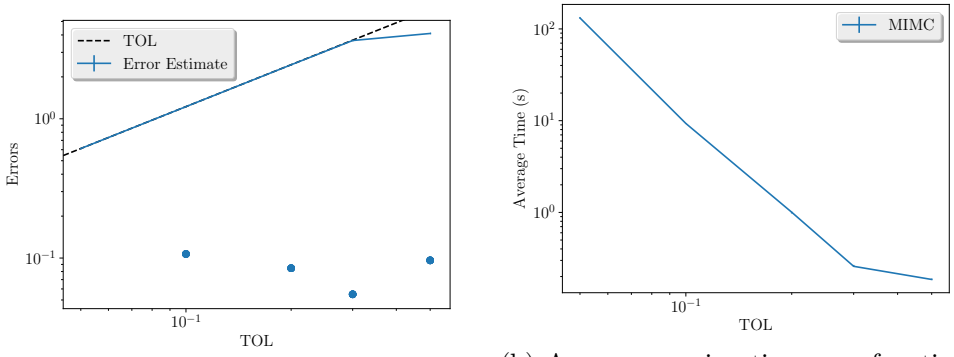



(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 91: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 8$.

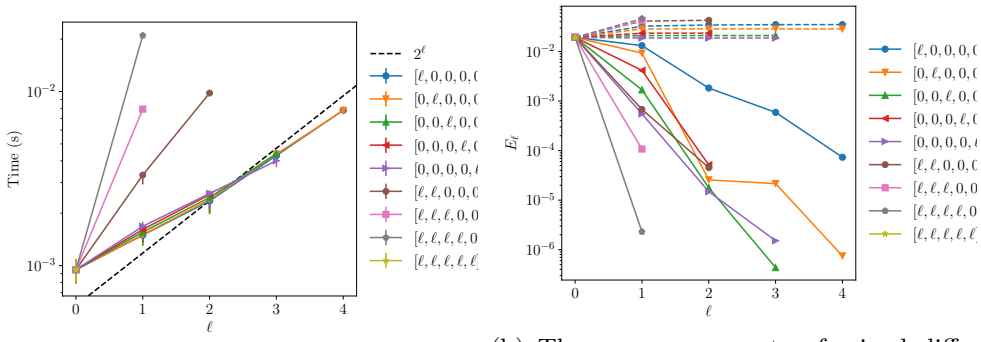
**Case of 16 time steps**




(a) Error estimate

(b) Average running time as a function of $TOL$

Figure 92: Convergence and complexity results for the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 16$.




(a) Average Computational time per level

(b) The convergence rate of mixed differences per level

Figure 93: Convergence and work rates for discretization levels the call payoff with rBergomi model for $K = 1$, $H = 0.07$ and $N = 16$.