Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

DISCRETIZATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS, L^p WASSERSTEIN METRICS, AND ECONOMETRICAL MODELS

BY MATTHIAS GELBRICH* AND SVETLOZAR T. RACHEV†
Humboldt-Universität, Berlin
and University of California, Santa Barbara

The paper deals with weak approximations of stochastic differential equations of Itô type, where convergence rates of the approximate solutions are obtained using $\mathrm{E}\|\cdot\|^p_{C[t_0,T]}$, $p \in [2,\infty)$. The rates can also be interpreted as rates for the L^p Wasserstein metrics, $p \in [1,\infty)$, between the distributions of exact and approximate solutions. This metric is a minimal distance of two r.v.'s with fixed distributions, and, thus, it is the optimal value of a marginal problem. The approximation scheme considered is a combination of the time discretization based on the stochastic Euler method with a chance discretization based on the invariance principle, and it works on a grid constructed to tune both discretizations. The schemes are adapted to treat econometric ARCH/GARCH models.

1. Introduction. This paper is designed to approximate the solution of a multi-dimensional stochastic differential equation (*sde*) of Itô type, following the lines in Gelbrich (1995) and adapting the results in order to deal with approximate solutions known in econometrical models. That means, drift and diffusion may depend not only on the present, but also on past time points. The methods investigated here are based on the evaluation of the drift and diffusion coefficients at grid points, and they combine the time discretization of the sde – as done for instance by the stochastic analogue of Euler's method – with the discretization of the stochastic input, the Wiener process. This combination of time and chance discretization is necessary for a computer simulation of the solution of the Itô sde.

---

* This research is supported by a grant from Deutscher Akademischer Austauschdienst and by Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuering" of Deutsche Forschungsgemeinschaft

† This work was supported in part by NSF Grand DMS-9103452 and NATO Scientific Affairs Division Grant CRG 900798.

AMS 1991 Subject Classification: Primary 60H10; Secondary 60F17

Key words and phrases: stochastic differential equations, discrete approximation, convergence rates, invariance principle.

A broad survey of various approximations to solutions of sde's is given in the monograph by Kloeden and Platen (1992). Platen (1981) gives convergence orders of time discrete approximations – constructed via the stochastic Taylor expansion – with respect to the mean square of the supremum norm. The method of order one considered there is the stochastic Euler method (introduced by Maruyama (1955)) and will be the basis of the method considered in the present paper. Together with these time discretizations we will discretize the Wiener process and estimate the distance between the distribution of the exact solution and the distributions of the approximate solutions – all solutions being referred to as random variables with values in a space of continuous functions.

Kanagawa (1986) uses a method derived from the stochastic Euler method by replacing the increments of the Wiener process by other "simpler" i.i.d. random variables: He uses L^p Wasserstein metrics ($p \geq 2$) between the distributions of exact and approximate solutions, thus achieving convergence rates. (For a broad survey of probability metrics see Rachev (1991), of L^p Wasserstein metrics see, e.g., Givens and Shortt (1984) and Gelbrich (1990). We use the same metrics, but make the method of Kanagawa more flexible, so that it will converge faster.

On an interval $[t_0, T]$ let an equidistant grid H with grid points $t_0 = \hat{t}_0 < \hat{t}_1 < \cdots < \hat{t}_{\hat{n}} = T$ with step size $\hat{h}$ be given. H will be the minimal set of time points at which values are available for the method, and $\hat{h}$ will be the period between two neighbouring observations in the past which influence the present drift and diffusion coefficients at any time. For any $t \in [t_0, T]$ we define $i_H(t) := \max\{i : \hat{t}_i \leq t\}$ as the number of time steps $\hat{h}$ one can go back into the past from t.

We consider a stochastic differential equation in integral form where drift and diffusion coefficients depend on the present state as well as on the states at times reached by going from the present back into the past by multiples of $\hat{h}$:

$$\begin{aligned} x(t) - x_0 &= \int_{t_0}^{t} b(x,s)ds + \int_{t_0}^{t} \sigma(x,s)dw(s) \\ &= \int_{t_0}^{t} b(x,s)ds + \sum_{j=1}^{q} \int_{t_0}^{t} \sigma_j(x,s)dw_j(s), \qquad (I) \\ & t \in [t_0, T], x_0 \in I\!R^d, \end{aligned}$$

where $w = (w_1, \ldots, w_q)^T$ is a q-dimensional standard Wiener process, and where we use the notations

$$\begin{aligned}
b(x,s) &:= b^{i_H(s)}(x(s), x(s-\hat{h}), x(s-2\hat{h}), \ldots, x(s-i_H(s)\hat{h})),\\
\sigma(x,s) &= (\sigma_1(x,s), \ldots, \sigma_q(x,s))\\
&:= \sigma^{i_H(s)}(x(s), x(s-\hat{h}), x(s-2\hat{h}), \ldots, x(s-i_H(s)\hat{h}))
\end{aligned}$$

with $b^\nu \in C(\mathbb{R}^{(\nu+1)d}; \mathbb{R}^d)$ and $\sigma^\nu \in C(\mathbb{R}^{(\nu+1)d}; \mathcal{L}(\mathbb{R}^q; \mathbb{R}^d))$, $\nu = 0, \ldots, i_H(T)$, where $\sigma_j^\nu \in C(\mathbb{R}^{(\nu+1)d}; \mathbb{R}^d)$, $j = 1, \ldots, q$, denote the columns of the matrix function $\sigma^\nu = (\sigma_1^\nu, \ldots, \sigma_q^\nu)$. Here and in the sequel we denote by C spaces of continuous functions, by $\mathcal{L}$ spaces of linear mappings, and by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^n$ ($n \in \mathbb{N}$) and the corresponding induced norm on a space $\mathcal{L}$.

For any random variable ζ mapping a probability space $(\Omega, \mathcal{A}, P)$ into a separable metric space (X, d) with the Borel σ-algebra $\mathcal{B}(X)$, the notation $D(\zeta)$ shall mean the distribution $P \circ \zeta^{-1}$ induced on X by ζ. $\mathcal{P}(X)$ shall be the set of all Borel probability measures on X.

The case in which b and σ explicitly depend on the time t can be written in the form (I) by taking t as another component of x. A direct treatment of this case – carried out in Gelbrich (1989) for equidistant grids and bounded b and σ depending only on the present state – follows the same lines as in this paper, but permits relaxation of second order differentiability w.r.t. t, that would be required for using the results in the present paper, to first order differentiability w.r.t. t.

For $p \in [1, \infty)$ we define a metric W_p on the set $\mathcal{M}_p(X) := \{\mu \in \mathcal{P}(X) : \int_X (d(x,\theta))^p d\mu(x) < \infty, \theta \in X\}$ by

$$W_p(\mu, \nu) := \left[\inf \int_{X \times X} (d(x,y))^p d\eta(x,y)\right]^{1/p} \qquad (\mu, \nu \in \mathcal{M}_p(X))$$

where the infimum is taken over all measures $\eta \in \mathcal{P}(X \times X)$ with marginal distributions μ and ν. Thus, computing W_p is equivalent to solving a marginal problem. W_p is called the L^p Wasserstein metric or L^p Kantorovich metric (see Rachev (1991)) and has the properties of a metric on $\mathcal{M}_p(X)$ (see Givens and Shortt (1984)). With respect to these metrics Kanagawa (1986) states a convergence result for a sequence of approximations to the solution x of (I) which are constructed over equidistant grids using both the stochastic Euler method and a substitution of the Wiener process increments between grid points by other i.i.d. r.v.'s (which are for instance easier to generate on a computer). This idea of joint discretization w.r.t. time and chance (earlier considered also by Janssen (1984)) gave rise to a certain construction leading to the definition of the approximate solution (E3); we shall describe them both in the rest of this section.

The approximate solution (E3) can be seen as a framework for dealing, e.g., with certain models known in econometrics literature as *Autoregressive Conditional Heteroscedasticity (ARCH)* or *Generalized ARCH (GARCH)* models which use discrete time processes in order to model stock price changes. We give a definition of these concepts:

Considering the equidistant grid $t_0 = \hat{t}_0 < \hat{t}_1 < \cdots < \hat{t}_{\hat{n}} = T$ with step size $\hat{h}$ on the time interval $[t_0, T]$, we follow Engle (1982) and Bollerslev, Chou, and Kroner (1992) in defining a univariate ARCH model as a discrete time stochastic process $(\epsilon_{\hat{t}_i})_{i=0,\ldots,\hat{n}}$ of the form

$$\epsilon_{\hat{t}_{i+1}} = \hat{\sigma}_{\hat{t}_i}\delta_{\hat{t}_i}$$

where $\hat{\sigma}_{\hat{t}_i}$ is a positive measurable function of the time points $\hat{t}_0, \hat{t}_1, \ldots, \hat{t}_i$ and the $\delta_{\hat{t}_i}$ are i.i.d. r.v.'s with zero mean and variance one. In a *linear ARCH(ψ)* the variances $\sigma_{\hat{t}_i}$ depend on the squares of the past ψ values of the process:

$$\hat{\sigma}^2_{\hat{t}_i} := \omega + \sum_{r=0}^{\psi-1} \alpha_r \epsilon^2_{\hat{t}_{i-r}}$$

whereas in the more general *linear GARCH(ϕ,ψ)* they may also depend on the ϕ recent variances:

$$\hat{\sigma}^2_{\hat{t}_i} := \omega + \sum_{r=0}^{\psi-1} \alpha_r \epsilon^2_{\hat{t}_{i-r}} + \sum_{r=1}^{\phi} \beta_r \hat{\sigma}^2_{\hat{t}_{i-r}}. \tag{1}$$

In these models it is assumed that $\omega > 0$, $\alpha_r \geq 0$, $\beta_r \geq 0$ for all r. Later we will embed this model (slightly modified) into the constructed approximation for the sde (I).

The corresponding multivariate model reflects price changes in portfolios of d assets and is a process $(\epsilon_{\hat{t}_i})_{i=0,\ldots,\hat{n}} \subset I\!R^d$ with

$$\epsilon_{\hat{t}_{i+1}} = \Omega^{1/2}_{\hat{t}_i}\delta_{\hat{t}_i}$$

where the $\Omega_{\hat{t}_i}$ are positive definite $d \times d$ matrices and measurable functions of $\hat{t}_0, \ldots, \hat{t}_i$, and where the $\delta_{\hat{t}_i}$ are i.i.d. r.v.'s with zero mean and have the d-dimensional unit matrix as covariance matrix. For the the multivariate *linear GARCH(ϕ,ψ)* one sets

$$\mathrm{vech}(\Omega_{\hat{t}_{i+1}}) := W + \sum_{r=0}^{\psi-1} A_r \mathrm{vech}(\epsilon_{\hat{t}_{i-r}}\epsilon^T_{\hat{t}_{i-r}}) + \sum_{r=1}^{\phi} B_r \mathrm{vech}(\Omega_{\hat{t}_{i-r}})$$

where vech($\cdot$) puts the lower right triangle of a symmetric $d \times d$ matrix in the form of a vector in $I\!R^{1/2 \cdot d(d+1)}$, and where $W \in I\!R^{1/2 \cdot d(d+1)}$ and the A_r and B_r

are $(1/2 \cdot d(d+1)) \times (1/2 \cdot d(d+1))$ matrices. The process $(\epsilon_{\hat{t}_i})$ is designed to model stock price changes, and a model $(S_{\hat{t}_i})_{1,\ldots,\hat{n}}$ of a d-dimensional portfolio price process is obtained by setting

$$\epsilon_{\hat{t}_i} = \ln(S_{\hat{t}_i}) - \ln(S_{\hat{t}_{i-1}}),$$

the logarithms taken componentwise. Then the process

$$\Lambda_{\hat{t}_i} := \ln(S_{\hat{t}_i}) = \ln(S_{\hat{t}_0}) + \sum_{r=1}^{i} \epsilon_{\hat{t}_r}$$

has independent increments whose covariance matrices $\Omega_{\hat{t}_i}$ depend on the prices at the times $\hat{t}_0, \ldots, \hat{t}_{i-1}$. The method (E3), defined in the sequel, will also produce processes with independent increments and will provide convergence rates w.r.t. the W_p metrics towards the solution of (I). Compared with $(\ln(S_{\hat{t}_i}))_{i=0,\ldots,\hat{n}}$, the convergence result for (E3) will require a bounded diffusion and allow for a drift, both as functions of the present state and of the states reached by going any number of time intervals $\hat{h}$ back into the past. Moreover, the discretization will go beyond the grid mentioned above and use (possibly) a finer grid for a better time discretization, using the stochastic Euler method (E1), and an even finer grid to construct invariance principle approximations – with a rate fitted to the time discretization rate – of the Wiener process between neighbouring grid points for the Euler method (E1). The method (E3) is given – following the lines in Gelbrich (1995) – together with the two "intermediate" methods (E1) and (E2) which will facilitate the proof of the main convergence result by allowing us to divide it into three steps.

In the sequel we shall use the following general assumptions concerning (I):

(V1) There exists a constant $M > 0$ such that
for all $j = 1, \ldots, q$; $\nu = 0, \ldots, i_H(T)$ and $x_0, \ldots, x_\nu \in \mathbb{R}^d$
$\|b^\nu(x_0, \ldots, x_\nu)\| \leq M(1 + \max_{0 \leq \rho \leq \nu} \|x_\rho\|)$ and
$\|\sigma_j^\nu(x_0, \ldots, x_\nu)\| \leq M$.

(V2) There exists a constant $L > 0$ such that
for all $j = 1, \ldots, q$; $\nu = 0, \ldots, i_H(T)$ and $x_0, \ldots, x_\nu, y_0, \ldots, y_\nu \in \mathbb{R}^d$
$\|b^\nu(x_0, \ldots, x_\nu) - b^\nu(y_0, \ldots, y_\nu)\| \leq L \max_{0 \leq \rho \leq \nu} \|x_\rho - y_\rho\|$ and
$\|\sigma_j^\nu(x_0, \ldots, x_\nu) - \sigma_j^\nu(y_0, \ldots, y_\nu)\| \leq L \max_{0 \leq \rho \leq \nu} \|x_\rho - y_\rho\|$.

(V1) and (V2) assure the existence and uniqueness of the solution of (I), both in the strong sense (see Gikhman and Skorokhod (1977)). The boundedness of σ_j in (V1) seems to be essential for the proof of Theorem 2.3.

As mentioned above, the approximate solutions in this paper are each based on a "double grid" – a *coarse* grid for the time discretization and a *fine* grid, being a refinement of the former, for the chance discretization via the invariance principle which yields a lower convergence speed than the time discretization. That is why we consider a grid class $\mathcal{G}(m,\alpha,\beta)$. Here let $m:(0,T-t_0]\to[1,\infty)$ be a monotone decreasing function, and let $\alpha,\beta>0$ be constants. Then each element G of $\mathcal{G}(m,\alpha,\beta)$ is constructed in the following way and has the following properties:

G consists of two kinds of grid points:

- the time discretization points $t_k, k=0\dots,n$, with $t_0<t_1<\dots<t_n=T$ and
- the chance discretization points $u_i^k, i=0,\dots,m_k, k=0,\dots,n-1$, with $t_k=u_0^k<u_1^k<\dots<u_{m_k}^k=t_{k+1},\quad k=0,\dots,n-1.$

Hence, G is a combination of a coarse subgrid consisting of all points t_k relevant for the pure time discretization and of a fine grid consisting of all points u_i^k needed for the discretization of the Wiener process. Now G is required to satisfy the following assumptions:

(G1) $t_k-t_{k-1}=\frac{T-t_0}{n}=:h\le 1$ for all $k=1,\dots,n$ and $\hat{h}/h\in \mathbb{N}$,

(G2) $1\le m_k\le m(h)^\alpha$ for all $k=0,\dots,n-1$,

(G3) $u_i^k-u_{i-1}^k=\frac{h}{m_k}\le\beta\frac{h}{m(h)}$ for all $k=0,\dots,n-1,\ i=1,\dots,m_k$.

Here (G1) means that the coarse grid is equidistant with step size h (required to be bounded by 1 only for convenience, in order to write simpler upper bounds later) and contains the master grid H. (G2) and (G3) say that each interval of the coarse subgrid is subdivided in an equidistant way by the points u_i^k, both the number of the subdivisions and the step size of the full grid being bounded by functions of h. As an example, it is easy to see that all equidistant grids which satisfy $m_k=[m(h)]$, $k=0,\dots,n-1$, belong to $\mathcal{G}(m,1,2)$.

For a grid G of $\mathcal{G}(m,\alpha,\beta)$ we define

$$[t]_G:=t_k \quad\text{and}\quad i_G(t):=k,\quad\text{if}\quad t\in[t_k,t_{k+1}),\ k=0,\dots,n-1,\text{ and}$$
$$[t]_G^*:=u_i^k\quad\text{if}\quad t\in[u_i^k,u_{i+1}^k),\ i=0,\dots,m_k-1,\ k=0,\dots,n-1.$$

We construct the approximate solution in (E3) in three steps. The first step is a pure time discretization using the stochastic Euler method (E1) (see Maruyama (1955)). Here only the coarse subgrid is involved.

$$\text{(E1)}\quad y^E(t)=x_0+\int_{t_0}^t b(y^E,[s]G)ds+\sum_{j=1}^q\int_{t_0}^t\sigma_j(y^E,[s]_G)dw_j(s),\quad t\in[t_0,T].$$

In the second step, a continuous and piecewise linear interpolation of the trajectories in (E1) between the points of the whole fine grid yields the method (E2).

$$\text{(E2)}\begin{cases}\tilde{y}^E \text{ is continuous, and linear in the intervals } [u_{i-1}^k, u_k^k],\\ \quad i=1,\dots,m_k,\ \ k=0,\dots,n-1,\\ \text{with } \tilde{y}^E(u_i^k)=y^E(u_i^k),\ \ i=0,\dots,m_k,\ \ k=0,\dots,n-1.\end{cases}$$

In the third step, the Wiener process increments over the fine grid are replaced by other i.i.d. r.v.'s: Let $\mu \in \mathcal{P}(I\!R)$ be a measure with mean value 0 and variance 1, and let

$$\{\xi_{js}^k :\ j=1,\dots,q;\ s=1,\dots,m_k;\ k=0,\dots,n-1\}$$

be a family of i.i.d. r.v.'s with distribution $D(\xi_{11}^0)=\mu$. Then we can define the following method (E3) yielding continuous trajectories which are linear between neighbouring grid points:

$$\text{(E3)}\begin{cases} z^E(u_0^0)=x_0, \text{ and}\\ z^E(u_i^k)=x_0+\sum\limits_{r=0}^{k-1} hb(z^E,t_r)+h\cdot\frac{i}{m_k}b(z^E,t_k)\\ \quad +\sum\limits_{j=1}^{q}\left[\sum\limits_{r=0}^{k-1}\sqrt{\frac{h}{m_r}}\sigma_j(z^E,t_r)\sum\limits_{s=1}^{m_r}\xi_{js}^r+\sqrt{\frac{h}{m_k}}\sigma_j(z^E,t_k)\sum\limits_{s=1}^{i}\xi_{js}^k\right]\\ \quad \text{for all } i=1,\dots,m_k,\ k=0,\dots,n-1.\end{cases}$$

For this last step, the Wiener process w and the r.v.'s ξ_{ji}^k will have to be defined anew on a *common* probability space. The following section investigates the convergence rates w.r.t. the norm $\mathrm{E}\sup\limits_{t_0\le t\le T}\|\cdot\|^p$ for $C([t_0,T];I\!R^d)$-valued r.v.'s in each of the three steps.

But first we will explain how (E3) looks for the univariate GARCH(ψ,ϕ) model (1). For this we note that we can write the diffusion $\hat{\sigma}_{\hat{t}_\nu}$, $\nu=0,\dots,\hat{n}$, as a function

$$\hat{\sigma}_{\hat{t}_\nu}=\hat{\sigma}^\nu(\Lambda_{\hat{t}_\nu},\Lambda_{\hat{t}_\nu-\hat{h}},\dots,\Lambda_{\hat{t}_\nu-\nu\hat{h}})=(\omega_\nu+\sum_{k=0}^{\nu-1}\theta_{\nu,k}(\Lambda_{\hat{t}_\nu-k\hat{h}}-\Lambda_{\hat{t}_\nu-(k+1)\hat{h}})^2)^{1/2},$$

for some positive numbers ω_ν and $\theta_{\nu,k}$. This becomes clear by recursive substitution using (1). It is obvious that all $\hat{\sigma}^\nu$ satisfy (V2). In order to fulfill the boundedness condition (V1), instead of $\hat{\sigma}^\nu$, we use a function bounded by some number $B>0$, namely

$$\bar{\sigma}^\nu := (\hat{\sigma}^\nu \wedge B)\vee(-B).$$

This slight modification should not be significant in practical applications. Now, for the univariate linear GARCH(ψ,ϕ), our model (E3) has the following form which we will call *refined bounded GARCH (RBGARCH)*:

$$\text{(RBGARCH)}\begin{cases} \Lambda^E(u_0^0) = \Lambda_{t_0}, \text{ and} \\ \Lambda^E(u_i^k) = \Lambda^E(t_k) + \sqrt{\frac{h}{m_k}}\bar{\sigma}^{i_H(t_k)}\Big(\Lambda^E(t_k), \Lambda^E(t_k-\hat{h}), \\ \quad \ldots, \qquad \Lambda^E(t_k - i_H(t_k)\hat{h})\Big)\sum_{s=1}^{i}\xi_{js}^k \\ \quad \text{for all} \quad i=1,\ldots,m_k,\ k=0,\ldots,n-1, \end{cases}$$

where we define the values of Λ^E between the grid points by linear interpolation. Here the increments of Λ^E may depend on all the so far passed time points in the coarse time discretization grid which is usually finer than the master grid H (but contains it) on which the scheme (1) operates. That means, with a finer time discretization more observations are needed for the whole approximation, but at each time point the diffusion coefficient still needs a restricted number of observations back in the past in time intervals with fixed length $\hat{h}$.

The matching sde (I) for the method (RBGARCH) is

$$\Lambda(t) = \Lambda_{t_0} + \int_{t_0}^{t} \bar{\sigma}^{i_H(s)}(\Lambda(s), \Lambda(s-\hat{h}), \ldots, \Lambda(s-i_H(s)\hat{h}))dw(s), \qquad (2)$$

for $t \in [t_0, T]$. Later we will see that Λ^E converges towards Λ in W_p-sense. We call (2) *continuous bounded GARCH.* It is our hope that the L^p estimates of the closeness between the discrete model (RBGARCH) and the continuous bounded GARCH will provide us with the necessary tools to construct a contingency claim valuation theory and capital asset pricing models based on (2) which is certainly a much more realistic model for asset pricing than the log-Gaussian model (see Kariya (1993), Mittnik and Rachev (1993) for further discussion).

2. Convergence results. According to the evolution of the method (E3) via (E1) and (E2), each step will be represented by one convergence theorem, yielding then immediately the main result given in two forms – one using the W_p metrics. The proofs of these three theorems can be found in Section 3. The theorems in the sequel will be formulated for an arbitrary fixed grid G of the grid class $\mathcal{G}(m,\alpha,\beta)$. Therefore G fulfills (G1)–(G3) with the construction in the previous section.

For convenience, throughout the whole paper, we shall denote by K any constant depending only on p, the considered grid class, and on the data of the original sde (I). This means, K does not depend on the particular grid. Moreover, K may have different values at different occurrences.

The first theorem gives rates for the convergence of the approximate solutions in (E1) to the solution of (I). For $p = 2$ and drift and diffusion dependent only on the present state, it was proved by Platen (1981) – and it was generalized to the case $p \in [2, \infty)$ in Gelbrich (1995), Theorem 2.6, using quite similar techniques.

THEOREM 2.1. *Let $p \in [2, \infty)$. Then, (V1) and (V2) imply*

$$\mathrm{E} \sup_{t_0 \le t \le T} \left\| x(t) - y^E(t) \right\|^p \le K \cdot h^{p/2}.$$

Whereas the solution in (E1) behaves like the Wiener process between two neighbouring points t_{k-1} and t_k of the coarse subgrid of G, the method (E2) provides a solution smoothened by linear interpolation with vertices in *all* grid points of G, that means in all u_i^k. The next theorem gives estimates for the L^p-norm of the difference between the approximate solutions in (E1) and (E2):

THEOREM 2.2. *Let $p \in [2, \infty)$. Then (V1) and (V2) imply*

$$\mathrm{E} \sup_{t_0 \le t \le T} \left\| y^E(t) - \tilde{y}^E(t) \right\|^p \le K \left(\tfrac{h}{m(h)} \right)^{p/2} \left(1 + \ln \left(\tfrac{m(h)}{h} \right) \right)^{p/2}.$$

In the last discretization step the Wiener process increments shall be replaced by i.i.d. r.v.'s with a given distribution μ on $\mathbb{R}$. But the corresponding results in Theorem 2.3 hold only in the weak sense, i.e. the Wiener process (and its increments between the points of G) and i.i.d. r.v.'s ξ_{ji}^k can be defined on a common probability space such that the estimates hold.

THEOREM 2.3. *Let $p \in [2, \infty)$ and $\mu \in \mathcal{P}(\mathbb{R})$ have the following properties:*

$$\int_{-\infty}^{\infty} x \, d\mu(x) = 0, \int_{-\infty}^{\infty} x^2 d\mu(x) = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} e^{tx} d\mu(x) < \infty \qquad (3)$$

$$\textit{for all } t \textit{ with } \|t\| \le \tau, \ \tau > 0.$$

Then we can define a q-dimensional standard Wiener process $\left(w(t)\right)_{t \in [t_0, T]}$ and a set of i.i.d. r.v.'s $\{\xi_{ji}^k : j = 1, \dots, q; \ i = 1, \dots, m_k; \ k = 0, \dots, n-1\}$ with distribution $D(\xi_{11}^0) = \mu$ on a common probability space, such that for the solutions in (E2) and (E3) constructed with them we have, under the *assumptions (V1) and (V2), that*

$$\mathrm{E} \sup_{t_0 \le t \le T} \left\| \tilde{y}^E(t) - z^E(t) \right\|^p \le K \left(\frac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p .$$

The preceding three Theorems 2.1, 2.2, and 2.3 yield the following theorem which gives bounds for the L^p-norm of the differences between the exact solution x of (I) and the approximate solution z^E defined in (E3). Again, as in Theorem 2.3, this is a result in the weak sense.

THEOREM 2.4. *Let $p \in [2, \infty)$ and $\mu \in \mathcal{P}(\mathbb{R})$ have the properties (3). Then we can define a q-dimensional standard Wiener process $\big(w(t)\big)_{t\in[t_0,T]}$ and a set of i.i.d. r.v.'s $\{\xi_{ji}^k : j=1,\dots,q;\ i=1,\dots,m_k;\ k=0,\dots,n-1\}$ with distribution $D(\xi_{11}^0) = \mu$ on a common probability space, such that for (I) and the method (E3) constructed with them we have, under the assumptions (V1) and (V2), that*

$$\mathrm{E} \sup_{t_0 \le t \le T} \left\| x(t) - z^E(t) \right\|^p \le K \left\{ h^{p/2} + \left(\frac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p \right\} .$$

PROOF. show that the assertion follows from the Theorems 2.1, 2.2, and 2.3 it suffices to verify that

$$\frac{h}{m(h)} \left(1 + \ln \left(\frac{m(h)}{h} \right) \right) \le K \left(\frac{1+\ln m(h)}{\sqrt{m(h)}} \right)^2 .$$

But this follows easily from (26) for $\gamma = \frac{1}{h} \ge 1$ (because of (G1)) and $\delta = m(h) \ge 1$. ∎

Since Theorem 2.4 provides a result in the weak sense, it is appropriate to formulate it as an estimate for the L^p Wasserstein metric between the *distributions* of the exact solution and the approximate solution:

COROLLARY 2.5. *Let $p \in [1, \infty)$ and $\mu \in \mathcal{P}(\mathbb{R})$ have the properties (3). Moreover, let $\big(w(t)\big)_{t\in[t_0,T]}$ be a q-dimensional standard Wiener process and $\{\xi_{ji}^k : j = 1,\dots,q; i = 1,\dots,m_k;\ k = 0,\dots,n-1\}$ a set of i.i.d. r.v.'s with distribution $D(\xi_{11}^0) = \mu$. Then for (I) and the method (E3) constructed with them we have, under the assumptions (V1) and (V2), that*

$$W_p(D(x), D(z^E)) \le K \left\{ h^{1/2} + \frac{1+\ln m(h)}{\sqrt{m(h)}} \right\} .$$

PROOF. For $p \in [2, \infty)$ the assertion follows directly from Theorem 2.4 and after applying to the right-hand side the inequality $a_1^p + a_2^p \le (a_1 + a_2)^p$

for $a_1, a_2 \geq 0$ and $p \in [2, \infty)$. (This inequality becomes obvious by dividing a_1 and a_2 by $a_1 + a_2$ in case this sum is positive.) The assertions are also true for $p \in [1, 2)$, since $W_{p_1} \leq W_{p_2}$ for $1 \leq p_1 \leq p_2 < \infty$ (see Givens and Shortt (1984)). ∎

The estimates in Theorem 2.4 and Corollary 2.5 give convergence rates w.r.t. h for the method (E3) and for any grid sequence in $\mathcal{G}(m, \alpha, \beta)$. These rates consist of two summands, one depending on h and the other depending on $m(h)$, representing the rates of time and chance discretization, respectively. Obviously it is not desirable that one of both summands converges faster than the other for this would only increase the costs in relation to the effect. Namely, if the second summand converges faster than the first, this would mean that $m(h)$ increases too fast and consequently – because of (G3) – the whole *fine* grid has too small a step size, i.e. there are too many points u_i^k in relation to the t_k in each grid and therefore a random number generator would be used too often. If the first summand converged faster than the second, then $m(h)$ would increase too slowly, i.e. the intervals $[t_k, t_{k+1}]$ would not have enough intermediate grid points u_i^k, so that the chance discretization would not keep up with the time discretization. Therefore, it is desirable to tune the rates of both summands, i.e. to equal the powers of h in both summands. This means to choose $m(h)$ to be increasing like $1/h$. In this way we get the following two corollaries immediately from Theorem 2.4 and Corollary 2.5.

Corollary 2.6. *Let $p \in [2, \infty)$ and $\mu \in \mathcal{P}(I\!R)$ have the properties (3). Then we can construct the solutions in (I) and (E3) on a common probability space (as in Theorem 2.4) so that under the assumptions (V1), (V2), and* $\max\left\{\sup_{0<s\leq 1} sm(s), \sup_{0<s\leq 1} \frac{1}{sm(s)}\right\} \leq K$ we have that

$$\mathrm{E} \sup_{t_0 \leq t \leq T} \left\|x(t) - z^E(t)\right\|^p \leq K \cdot h^{p/2}(1 - \ln h)^p.$$

Corollary 2.7. *Under the assumptions in Corollary 2.5 and with* $\max\left\{\sup_{0<s\leq 1} sm(s), \sup_{0<s\leq 1} \frac{1}{sm(s)}\right\} \leq K$ *we have:*

$$W_p(D(x), D(z^E)) \leq K \cdot h^{1/2}(1 - \ln h).$$

Thus, given a grid sequence in $\mathcal{G}(m, \alpha, \beta)$ with $h \to 0$ and using the metric W_p, we have, under the assumptions of Corollary 2.7, for the method (E3) the convergence rate $O(h^{1/2}(1 - \ln h))$ w.r.t. the maximal step sizes h of the coarse subgrids and the convergence rate $O((\frac{h}{m(h)})^{1/4}(1 - \ln \frac{h}{m(h)}))$ w.r.t. the maximal step sizes $\frac{h}{m(h)}$ of the whole fine grids and the convergence rate $O(N^{-1/4}(1 + \ln N))$ w.r.t. the number N of all gridpoints of the whole fine grids.

Kanagawa (1986) deals with the method (E3) in the case of $m(h) \equiv 1$ and gets at most – assuming the existence of the third moment of ξ_{11}^0 – the convergence rate $O(N^{-1/6}(\ln N)^\epsilon)$, $(\epsilon > 1/2)$. Kanagawa's result does not follow from the results proved here and was proved using different tools and different assumptions. Our method (E3) yields a better order (essentially $N^{-1/4}$) than Kanagawa's method – we call it (K) – (essentially $N^{-1/6}$). Moreover, (E3) needs to compute the coefficients b and σ only in a small part of the N grid points, namely the points t_k of the coarse subgrids, whereas (K) requires the computation of the coefficients in all N grid points. This shows that (E3) has also lower costs than (K) for the same N. If we take in the grids for (E3) and (K) the same numbers n of "expensive" grid points (i.e. points where b and σ have to be computed) or the same corresponding step sizes h, then the orders of (E3) and (K) are essentially $n^{-1/2}$ and $n^{-1/6} = N^{-1/6}$ (or $h^{1/2}$ and $h^{1/6}$), which makes the difference between both methods more significant.

Under the assumptions of Corollary 2.7 we get for the discrete model (RBGARCH) and the solution of (2) (the continuous bounded GARCH) the convergence result

$$W_p(D(\Lambda), D(\Lambda^E)) = O(h^{1/2}(1 - \ln h)),$$

i.e., the distribution of the process Λ^E – obtained by (RBGARCH) – and the solution Λ of the sde (2) are closely related, where Λ can be referred to as the ideal, continuous model and Λ^E as its discrete approximation, and for the model (RBGARCH) we can estimate the approximation error by means of the W_p metrics. The link of this model to the sde (2) and its solution, the continuous martingale Λ, immediately allows the use of stochastic calculus and martingale theory to investigate (2) and Λ and draw conclusions for the time discrete process Λ^E defined by the method (RBGARCH).

3. Proofs. The proof of Theorem 2.1 shall use three lemmas which are stated below and proved in Gelbrich (1995). The first one provides the multi-dimensional Hölder inequality in both continuous and discrete form:

LEMMA 3.1 (HÖLDER'S INEQUALITY). *a) Let $p \in [1, \infty)$, $s < t$, and let $g : [s,t] \to \mathbb{R}^d$, $g(u) = (g_1(u), \ldots, g_d(u))^T$ $(u \in [s,t])$, be a Borel measurable function such that $|g_i|^p$ is Lebesgue integrable over $[s,t]$ for $i = 1, \ldots, d$. Then*

$$\left\| \int_s^t g(u)du \right\|^p \leq (t-s)^{p-1} \int_s^t \|g(u)\|^p du.$$

b) Let $p \in [1, \infty)$ and $a_i \in \mathbb{R}^d$ for all $i = 1, \ldots, r$. Then

$$\left\| \sum_{i=1}^r a_i \right\|^p \leq r^{p-1} \sum_{i=1}^r \|a_i\|^p.$$

The main tools for the proof of Theorem 2.1 are the multi-dimensional martingale inequalities which the following lemma contains in both continuous and discrete form. It is a consequence of a generalization of results in Ikeda and Watanabe (1981) and Shiryaev (1984) to the multi-dimensional case, combined with Lemma 3.1.

LEMMA 3.2. *Let $p \in [2,\infty)$. Then there exist constants $C_p, A_p > 0$ such that the following assertions hold:*

a) Let $(w(t),\mathcal{F}(t))_{t\in[\gamma,\delta]}$ be a one-dimensional standard Wiener process over the probability space $(\Omega,\mathcal{A},P)$. Then for every function $g=(g_1,\ldots,g_d)$: $[\gamma,\delta]\times\Omega\to\mathbb{R}^d$ with

(i) $g(\cdot,\omega)$ is square-integrable over $[\gamma,\delta]$ for almost all $\omega\in\Omega$, and

(ii) $g(u)=g(u,\cdot)$ is $\mathcal{F}(u)$-measurable for all $u\in[\gamma,\delta]$, we have

$$\mathrm{E}\sup_{\gamma\le s\le t}\left\|\int_\gamma^s g(u)dw(u)\right\|^p\le[d(\delta-\gamma)]^{p/2-1}C_p\int_\gamma^t \mathrm{E}\|g(u)\|^p du,$$

for all $t\in[\gamma,\delta]$.

b) *Let $(M_s,\mathcal{F}_s)_{s=0,\ldots,r}$ be an $\mathbb{R}^d$-valued martingale (i.e. each component is a martingale), and let $p\in[2,\infty)$. Then with $\Delta M_s := M_s - M_{s-1}$ we have*

$$\mathrm{E}\max_{0\le s\le r}\|M_s\|^p\le A_p(dr)^{p/2-1}\mathrm{E}\sum_{s=1}^r\|\Delta M_s\|^p.$$

Also for Gronwall's lemma we need – besides its original form – a discrete analogue:

LEMMA 3.3 (GRONWALL'S LEMMA). a) *Let $f:[t_0,T]\to[0,\infty)$ be a continuous function and c_1, c_2 be positive constants. If for all $t\in[t_0,T]$*

$$f(t)\le c_1+c_2\int_{t_0}^t f(s)ds$$

then

$$\sup_{t_0\le t\le T} f(t)\le c_1e^{c_2(T-t_0)}.$$

b) Let $a_0,\ldots,a_n$ and c_1, c_2 be non-negative real numbers. If for all $k=0,\ldots,n$

$$a_k\le c_1+c_2\frac{1}{n}\sum_{i=0}^{k-1}a_i$$

then

$$\max_{0\le i\le n}a_i\le c_1e^{c_2}.$$

Now we prove Theorem 2.1, following the lines of the proof of Theorem 2.6 in Gelbrich (1995):

PROOF OF THEOREM 2.1. First, we observe the boundedness of the pth moment of the solution in (I): From Lemma 3.2a) and Lemma 3.1a), b) and (V1) we get for all $t \in [t_0, T]$

$$\begin{aligned}
\mathrm{E} \sup_{t_0 \le s \le t} \|x(s)\|^p &\le K\left(\|x_0\|^p + \mathrm{E} \sup_{t_0 \le s \le t} (T-t_0)^{p-1} \int_{t_0}^{s} \|b(x,u)\|^p du \right. \\
&\quad \left. + \sum_{j=1}^{q} \mathrm{E} \sup_{t_0 \le s \le t} \left\| \int_{t_0}^{s} \sigma_j(x,u) dw_j(u) \right\|^p \right) \\
&\le K\left(1 + \int_{t_0}^{t} \mathrm{E}\|b(x,u)\|^p du + \sum_{j=1}^{q} \int_{t_0}^{t} \mathrm{E}\|\sigma_j(x,u)\|^p du\right) \\
&\le K\left(1 + \int_{t_0}^{t} \mathrm{E} \sup_{t_0 \le s \le u} \|x(s)\|^p du\right)
\end{aligned}$$

and from Lemma 3.3a)

$$\mathrm{E} \sup_{t_0 \le t \le T} \|x(t)\|^p \le K. \tag{4}$$

Using the definitions (I) and (E1), we split the following difference for $t \in [t_0, T]$:

$$\begin{aligned}
x(t) - y^E(t) &= \int_{t_0}^{t} [b(x,s) - b(x,[s]_G)] ds + \int_{t_0}^{t} [b(x,[s]_G) - b(y^E,[s]_G)] ds \\
&\quad + \sum_{j=1}^{q} \left\{ \int_{t_0}^{t} [\sigma_j(x,s) - \sigma_j(x,[s]_G)] dw_j(s) \right. \\
&\quad \left. + \int_{t_0}^{t} [\sigma_j(x,[s]_G) - \sigma_j(y^E,[s]_G)] dw_j(s) \right\} \\
&=: J_1(t) + J_2(t) + \sum_{j=1}^{q} \{J_{3j}(t) + J_{4j}(t)\}.
\end{aligned} \tag{5}$$

Now for all $t \in [t_0, T]$ Lemma 3.1a) and (V2) imply

$$\mathrm{E} \sup_{t_0 \le r \le t} \|J_1(r)\|^p \le (T-t_0)^{p-1} L^p \int_{t_0}^{t} \mathrm{E} \sup_{t_0 \le u \le s} \|x(u) - x([u]_G)\|^p ds, \tag{6}$$

$$\begin{aligned}
\mathrm{E} \sup_{t_0 \le r \le t} \|J_2(r)\|^p &\le (T-t_0)^{p-1} L^p \int_{t_0}^{t} \mathrm{E} \sup_{t_0 \le u \le s} \|x([u]_G) - y^E([u]_G)\|^p ds \\
&\le K \int_{t_0}^{t} \mathrm{E} \sup_{t_0 \le u \le s} \|x(u) - y^E(u)\|^p ds,
\end{aligned} \tag{7}$$

while Lemma 3.2a) and (V2) imply

$$\mathrm{E}\sup_{t_0\le r\le t}\|J_{3j}(r)\|^p \le K\int_{t_0}^{t}\mathrm{E}\sup_{t_0\le u\le s}\|x(u)-x([u]_G)\|^p ds, \tag{8}$$

$$\begin{aligned}\mathrm{E}\sup_{t_0\le r\le t}\|J_{4j}(r)\|^p &\le K\int_{t_0}^{t}\mathrm{E}\sup_{t_0\le u\le s}\|x([u]_G)-y^E([u]_G)\|^p ds\\ &\le K\int_{t_0}^{t}\mathrm{E}\sup_{t_0\le u\le s}\|x(u)-y^E(u)\|^p ds.\end{aligned} \tag{9}$$

Here, by Lemma 3.1b), a), Lemma 3,2a), (V1) and (4), it holds for all $s\in[t_0,T]$ that

$$\begin{aligned}&\mathrm{E}\sup_{t_0\le u\le s}\|x(u)-x([u]_G)\|^p\\ &\le K\cdot\mathrm{E}\sup_{t_0\le u\le s}\left\{\left\|\int_{[u]_G}^{u}b(x,v)dv\right\|^p+\sum_{j=1}^{q}\left\|\int_{[u]_G}^{u}\sigma_j(x,v)dw_j(v)\right\|^p\right\}\\ &\le K\cdot\mathrm{E}\sup_{t_0\le u\le s}\left\{h^{p-1}\int_{[u]_G}^{u}\|b(x,v)\|^p dv+\sum_{j=1}^{q}h^{p/2-1}\int_{[u]_G}^{u}\|\sigma_j(x,v)\|^p dv\right\}\\ &\le K\left\{h^{p/2}\left(1+\mathrm{E}\sup_{t_0\le t\le T}\|x(t)\|^p\right)\right\}\le K\cdot h^{p/2}.\end{aligned} \tag{10}$$

Summarizing (5)–(10), we get for all $t\in[t_0,T]$ that

$$\begin{aligned}\mathrm{E}\sup_{t_0\le r\le t}\|x(r)-y^E(r)\|^p &\le K\Bigg[\mathrm{E}\sup_{t_0\le r\le t}\|J_1(r)\|^p+\mathrm{E}\sup_{t_0\le r\le t}\|J_2(r)\|^p\\ &+\sum_{j=1}^{q}\left\{\mathrm{E}\sup_{t_0\le r\le t}\|J_{3j}(r)\|^p+\mathrm{E}\sup_{t_0\le r\le t}\|J_{4j}(r)\|^p\right\}\Bigg]\\ &\le K\left\{\int_{t_0}^{t}\mathrm{E}\sup_{t_0\le u\le s}\|x(u)-y^E(u)\|^p ds+h^{p/2}\right\},\end{aligned}$$

and the assertion follows from Lemma 3.3a). ∎

For the proof of Theorem 2.2 we need the following lemma which is proved in Gelbrich (1995) (Lemma 3.3).

LEMMA 3.4. *Let $a_0<a_1<\ldots<a_r$ be a partition of $[a_0,a_r]$ with maximal step size $\Delta:=\max_{0\le i\le r-1}(a_{i+1}-a_i)$ and $(\tilde w(t))_{t\in[a_0,a_r]}$ a one-dimensional standard Wiener process. Then*

$$\mathrm{E}\max_{0\le i\le r-1}\sup_{a_i\le t\le a_{i+1}}|\tilde w(t)-\tilde w(a_i)|^p\le K\cdot\Delta^{p/2}(1+\ln r)^{p/2}.$$

The proof of Theorem 2.2 again follows the lines of the proof of Theorem 3.4 in Gelbrich (1995):

PROOF OF THEOREM 2.2. First we consider the process $\bar{y}^E$ with $\bar{y}^E(t_0) = x_0$, $\bar{y}^E(u_i^k) = \tilde{y}^E(u_i^k)$, $\bar{y}^E(t) = \bar{y}^E(u_{i-1}^k)$ for $t \in [u_{i-1}^k, u_i^k)(k=0,\dots,n-1; i=1,\dots,m_k)$. Then, by Lemma 3.1b), (V1), Lemma 3.4, (G2) and (G3), we have

$$\begin{aligned}
&\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\bar{y}^E(t)\right\|^p \\
&\le K\left\{\mathrm{E}\sup_{t_0\le t\le T}\left\|\int_{[t]_G^*}^t b(y^E,[t]_G)ds\right\|^p+\sum_{j=1}^q\mathrm{E}\sup_{t_0\le t\le T}\left\|\sigma_j(y^E,[t]_G)\int_{[t]_G^*}^t dw_j(s)\right\|^p\right\} \\
&\le K\left\{\mathrm{E}\sup_{t_0\le t\le T}\left[(t-[t]_G^*)^pM^p(1+\sup_{t_0\le s\le t}\|y^E([s]_G)\|^p)\right]\right. \\
&\quad\left.+\ M^p\sum_{j=1}^q\mathrm{E}\max_{\substack{0\le k\le n-1\\0\le i\le m_k-1}}\ \sup_{u_i^k\le t\le u_{i+1}^k}|w_j(t)-w_j(u_i^k)|^p\right\} \\
&\le K\left\{\max_{0\le k\le n-1}\left(\tfrac{h}{m_k}\right)^p\left(1+\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)\right\|^p\right)\right. \\
&\quad\left.+\sum_{j=1}^q\max_{0\le k\le n-1}\left(\tfrac{h}{m_k}\right)^{p/2}(1+\ln(n\cdot m(h)^\alpha))^{p/2}\right\} \\
&\le K\left\{1+\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)\right\|^p\right\}\left(\tfrac{h}{m(h)}\right)^{p/2}(1+\ln n+\ln m(h))^{p/2}.
\end{aligned}\tag{11}$$

Since we have by Minkowski's inequality that

$$\left(\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)\right\|^p\right)^{1/p}\le\left(\mathrm{E}\sup_{t_0\le t\le T}\left\|x(t)-y^E(t)\right\|^p\right)^{1/p}+\left(\mathrm{E}\sup_{t_0\le t\le T}\|x(t)\|^p\right)^{1/p}$$

where the right-hand side is bounded because of Theorem 2.1 and (4), it holds that

$$\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)\right\|^p\le K. \tag{12}$$

Hence, by (11) and (G1),

$$\begin{aligned}
\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\bar{y}^E(t)\right\|^p &\le K\left(\tfrac{h}{m(h)}\right)^{p/2}(1+\ln n+\ln m(h))^{p/2} \\
&\le K\left(\tfrac{h}{m(h)}\right)^{p/2}\left(1+\ln\left(\tfrac{m(h)}{h}\right)\right)^{p/2}.
\end{aligned}\tag{13}$$

On the other hand,

$$
\begin{aligned}
\mathrm{E}\sup_{t_0\le t\le T}\left\|\bar{y}^E(t)-\tilde{y}^E(t)\right\|^p &= \mathrm{E}\max_{\substack{0\le k\le n-1\\ 0\le i\le m_k-1}}\sup_{u_i^k\le t\le u_{i+1}^k}\|\bar{y}^E(t)-\tilde{y}^E(t)\|^p\\
&= \mathrm{E}\max_{\substack{0\le k\le n-1\\ 0\le i\le m_k-1}}\|\bar{y}^E(u_{i+1}^k)-\tilde{y}^E(u_i^k)\|^p\\
&\le \mathrm{E}\max_{\substack{0\le k\le n-1\\ 0\le i\le m_k-1}}\sup_{u_i^k\le t\le u_{i+1}^k}\|y^E(t)-y^E(u_i^k)\|^p\\
&= \mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\bar{y}^E(t)\right\|^p.
\end{aligned}
\tag{14}
$$

Now, by (13) and (14) we have

$$
\begin{aligned}
&\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\tilde{y}^E(t)\right\|^p\\
&\le K\left\{\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\bar{y}^E(t)\right\|^p+\mathrm{E}\sup_{t_0\le t\le T}\left\|\bar{y}^E(t)-\tilde{y}^E(t)\right\|^p\right\}\\
&\le K\cdot\mathrm{E}\sup_{t_0\le t\le T}\left\|y^E(t)-\bar{y}^E(t)\right\|^p\le K\left(\tfrac{h}{m(h)}\right)^{p/2}\left(1+\ln\left(\tfrac{m(h)}{h}\right)\right)^{p/2}. \quad \blacksquare
\end{aligned}
$$

As the main tool in the proof of Theorem 2.3 we shall use the following lemma. The proof of the lemma can be found in Gelbrich (1995) (Theorem 4.1, Lemmas 4.2 and 4.3 and the beginning of Theorem 4.4 up to the formulas (32)–(34)) and essentially uses results by Komlós, Major, and Tusnády (1975, 1976).

LEMMA 3.5. *Let $\mu\in\mathcal{P}(I\!R)$ have the following properties:*

$$
\int_{-\infty}^{\infty}x\,d\mu(x)=0,\quad \int_{-\infty}^{\infty}x^2d\mu(x)=1 \text{ and } \int_{-\infty}^{\infty}e^{tx}d\mu(x)<\infty
$$
$$
\text{for all } t \text{ with } \|t\|\le\tau,\ \tau>0.
$$

Then there exist a q-dimensional standard Wiener process $(w(t))_{t\in[t_0,T]}$ and a set $\{\xi_{ji}^k : j=1,\dots,q; k=0,\dots,n-1; i=1,\dots,m_k\}$ of i.i.d. random variables with distribution $D(\xi_{11}^0)=\mu$, both on the same probability space, such that with the notation $\Delta_i^k w_j := w_j(u_i^k)-w_j(u_{i-1}^k)$, $j=1,\dots,q$; $i=1,\dots,m_k$; $k=0,\dots,n-1$, the following three assertions hold:

a) *For $k=0,\dots,n-1$, $j=1,\dots,q$, and $p\in[2,\infty)$,*

$$
\mathrm{E}\max_{1\le i\le m_k}\left|\sum_{s=1}^{i}\xi_{js}^k-\sum_{s=1}^{i}\sqrt{\tfrac{m_k}{h}}\Delta_s^k w_j\right|^p\le K(1+\ln m(h))^p.
$$

b) *For* $j=1,\ldots,q$ *and* $p \in [2,\infty)$,

$$\mathrm{E} \max_{0\le k\le n-1} \max_{1\le i\le m_k} \left| \sum_{s=1}^{i} \xi_{js}^{k} - \sum_{s=1}^{i} \sqrt{\tfrac{m_k}{h}} \Delta_s^k w_j \right|^p \le K(1+\ln n+\ln m(h))^p.$$

c) *For* $k = 1,\ldots,n-1$,

> ξ_{js}^{r} and $\Delta_s^r w_j$, $j=1,\ldots,q$; $s=1,\ldots,m_r$; $r=k,\ldots,n-1$, are independent of the σ-algebra $\mathcal{A}_k$ generated by $\{\xi_{js}^{r},\ \Delta_s^r w_j :\ j=1,\ldots,q;\ s=1,\ldots,m_r;\ r=0,\ldots,k-1\}$.

The estimate for the chance discretization step is proved along the lines of the proof of Theorem 4.4 in Gelbrich (1995):

PROOF OF THEOREM 2.3. We consider those w and ξ that were asserted to exist in Lemma 3.5, as well as the approximation methods (E2) and (E3) defined on the basis of w and ξ. According to the definitions, for the estimate only the values of the approximate solutions in the grid points of G have to be taken into account:

First we consider the approximate solutions (E2) and (E3) only in the grid points t_k of the coarse subgrid of G. Then, with the notation

$$\Delta_k w_j := w_j(t_{k+1}) - w_j(t_k), \quad j=1,\ldots,q; \quad k=0,\ldots,n-1,$$

the definitions of (E2) and (E3) yield, with Lemma 3.1b), for $k=0,\ldots,n$

$$\begin{aligned} &\mathrm{E} \max_{0\le f\le k} \|\tilde{y}^E(t_f) - z^E(t_f)\|^p \\ &\le K\left\{ \mathrm{E} \max_{1\le f\le k} \left\| \sum_{r=0}^{f-1} h[b(\tilde{y}^E,t_r) - b(z^E,t_r)] \right\|^p \right. \\ &\left. + \sum_{j=1}^{q} \mathrm{E} \max_{1\le f\le k} \left\| \sum_{r=0}^{f-1} \left[\sigma_j(\tilde{y}^E,t_r)\Delta_r w_j - \sigma_j(z^E,t_r)\sqrt{\tfrac{h}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^{r} \right] \right\|^p \right\} \\ &=: K\left\{ D_1^E(k) + \sum_{j=1}^{q} D_{2j}^E(k) \right\}. \end{aligned} \tag{15}$$

Now, from Lemma 3.1b), (V2), and (G1), it follows that for $k=0,\dots,n$

$$\begin{aligned} D_1^E(k) &\le k^{p-1}L^p \mathrm{E} \max_{1\le f\le k} \sum_{r=0}^{f-1} h^p \max_{0\le\rho\le i_H(t_r)} \|\tilde{y}^E(t_r-\rho\hat{h}) - z^E(t_r-\rho\hat{h})\|^p \\ &\le K(nh)^p \frac{1}{n} \mathrm{E} \sum_{r=0}^{k-1} \max_{0\le\rho\le i_H(t_r)} \|\tilde{y}^E(t_r-\rho\hat{h}) - z^E(t_r-\rho\hat{h})\|^p \\ &\le K\cdot\frac{1}{n}\sum_{r=0}^{k-1} \mathrm{E} \max_{0\le s\le r} \|\tilde{y}^E(t_s) - z^E(t_s)\|^p, \end{aligned} \tag{16}$$

and, for $j=1,\dots,q$,

$$\begin{aligned} D_{2j}^E(k) &\le K\Bigg\{ \mathrm{E} \max_{1\le f\le k} \left\| \sum_{r=0}^{f-1} \left[\sigma_j(\tilde{y}^E,t_r) - \sigma_j(z^E,t_r)\right] \Delta_r w_j \right\|^p \\ &\quad + \mathrm{E} \max_{1\le f\le k} \left\| \sum_{r=0}^{f-1} \sigma_j(z^E,t_r) \left[\Delta_r w_j - \sqrt{\tfrac{h}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r\right] \right\|^p \Bigg\} \\ &=: K\{D_{2j1}^E(k) + D_{2j2}^E(k)\}. \end{aligned} \tag{17}$$

Because of Lemma 3.5c),

$$\begin{aligned} M_{j1}(f) &:= \sum_{r=0}^{f-1} [\sigma_j(\tilde{y}^E,t_r) - \sigma_j(z^E,t_r)]\Delta_r w_j \quad \text{and} \\ M_{j2}(f) &:= \sum_{r=0}^{f-1} \sigma_j(z^E,t_r) \left[\Delta_r w_j - \sqrt{\tfrac{h}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r\right], \quad f=0,\dots,n, \end{aligned}$$

are d-dimensional martingales w.r.t. $(\mathcal{A}_f)_{f=0,\dots,n}$, and that is why, using Lemma 3.2b), they can be estimated in the following way for all $j=1,\dots,q$ and $k=0,\dots,n$:

$$D_{2j1}^E(k) \le K(dk)^{p/2-1} \sum_{r=0}^{k-1} \mathrm{E}\{\|\sigma_j(\tilde{y}^E,t_r) - \sigma_j(z^E,t_r)\|^p |\Delta_r w_j|^p\}.$$

Since both factors in the braces are independent (because of Lemma 3.5c)), from (V2) and (G1) it follows that

$$\begin{aligned}
D^E_{2j1}(k) &\le K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \left\{ \mathrm{E}\|\sigma_j(\tilde{y}^E, t_r) - \sigma_j(z^E, t_r)\|^p \mathrm{E}|\Delta_r w_j|^p \right\} \\
&\le K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \Bigg\{ h^{p/2} \mathrm{E}\left(\tfrac{1}{\sqrt{h}} |\Delta_r w_j| \right)^p \\
&\qquad \times \mathrm{E} \max_{0 \le \rho \le i_H(t_r)} \|\tilde{y}^E(t_r - \rho\hat{h}) - z^E(t_r - \rho\hat{h})\|^p \Bigg\} \qquad (18) \\
&\le K \cdot n^{p/2-1} h^{p/2} \sum_{r=0}^{k-1} \mathrm{E} \max_{0 \le \rho \le i_H(t_r)} \|\tilde{y}^E(t_r - \rho\hat{h}) - z^E(t_r - \rho\hat{h})\|^p \\
&\le K \cdot \frac{1}{n} \sum_{r=0}^{k-1} \mathrm{E} \max_{0 \le s \le r} \|\tilde{y}^E(t_s) - z^E(t_s)\|^p.
\end{aligned}$$

Here we used that, since all $\frac{1}{\sqrt{h}}\Delta_r w_j$ are standard-normally distributed, all $\mathrm{E}\left|\frac{1}{\sqrt{h}}\Delta_r w_j\right|^p$ are equal to the same constant only depending on p.

For the other summand Lemma 3.2b), (V1), Lemma 3.5a), (G1), and (G3) yield

$$\begin{aligned}
D^E_{2j2}(k) &\le K \cdot (dk)^{p/2-1} \sum_{r=0}^{k-1} \mathrm{E} \left\{ \|\sigma_j(z^E, t_r)\|^p \left| \Delta_r w_j - \sqrt{\tfrac{h}{m_r}} \sum_{s=1}^{m_r} \xi^r_{js} \right|^p \right\} \\
&\le K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \left(\sqrt{\tfrac{h}{m_r}} \right)^p \mathrm{E} \left| \sqrt{\tfrac{m_r}{h}} \Delta_r w_j - \sum_{s=1}^{m_r} \xi^r_{js} \right|^p \\
&\le K \cdot n^{p/2-1} \left(\frac{h}{m(h)} \right)^{p/2} k(1 + \ln m(h))^p \qquad (19) \\
&\le K \left(\frac{nh}{m(h)} \right)^{p/2} (1 + \ln m(h))^p \le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p .
\end{aligned}$$

Now, considering (15) to (19), we get for all $k = 1, \ldots, n$ that

$$\begin{aligned}
&\mathrm{E} \max_{0 \le f \le k} \|\tilde{y}^E(t_f) - z^E(t_f)\|^p \\
&\qquad \le K \left\{ \frac{1}{n} \sum_{r=0}^{k-1} \mathrm{E} \max_{0 \le s \le r} \|\tilde{y}^E(t_s) - z^E(t_s)\|^p + \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p \right\},
\end{aligned}$$

and by Lemma 3.3b) we have

$$\mathrm{E} \max_{0 \le f \le n} \|\tilde{y}^E(t_f) - z^E(t_f)\|^p \le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p . \qquad (20)$$

In the next step we extend this estimate to the intermediate grid points

u_i^k, and we use the notation $\Delta_{i,0}^k w_j := w_j(u_i^k) - w_j(u_0^k)$, $j = 1, \ldots, q; k = 0, \ldots, n-1; i=1, \ldots, m_k$:

$$\begin{aligned}
&\mathrm{E} \max_{0\le k\le n-1} \max_{0\le i\le m_k} \|\tilde{y}^E(u_i^k) - z^E(u_i^k)\|^p \le K \Big\{ \mathrm{E} \max_{0\le k\le n-1} \|\tilde{y}^E(t_k) - z^E(t_k)\|^p \\
&+ \mathrm{E} \max_{0\le k\le n-1} \max_{0\le i\le m_k} \left\| h \cdot \tfrac{i}{m_k} \left(b(\tilde{y}^E, t_k) - b(z^E, t_k) \right) \right\|^p \\
&+ \sum_{j=1}^{q} \mathrm{E} \max_{0\le k\le n-1} \max_{1\le i\le m_k} \left\| \sigma_j(\tilde{y}^E, t_k) \Delta_{i,0}^k w_j - \sigma_j(z^E, t_k) \sqrt{\tfrac{h}{m_k}} \sum_{s=1}^{i} \xi_{js}^k \right\|^p \Big\} \\
&=: K \left\{ \mathrm{E} \max_{0\le k\le n-1} \|\tilde{y}^E(t_k) - z^E(t_k)\|^p + D_3^E + \sum_{j=1}^{q} D_{4j}^E \right\}.
\end{aligned} \tag{21}$$

Here, (V2) and (G1) imply

$$D_3^E \le K \cdot h^p \mathrm{E} \max_{0\le k\le n-1} \|\tilde{y}^E(t_k) - z^E(t_k)\|^p. \tag{22}$$

On the other hand, we have for all $j=1, \ldots, q$ that

$$\begin{aligned}
D_{4j}^E \le K \Big\{ &\mathrm{E} \max_{0\le k\le n-1} \max_{1\le i\le m_k} \|[\sigma_j(\tilde{y}^E, t_k) - \sigma_j(z^E, t_k)] \Delta_{i,0}^k w_j\|^p \\
&+ \mathrm{E} \max_{0\le k\le n-1} \max_{1\le i\le m_k} \left\| \sigma_j(z^E, t_k) \left[\Delta_{i,0}^k w_j - \sqrt{\tfrac{h}{m_k}} \sum_{s=1}^{i} \xi_{js}^k \right] \right\|^p \Big\} \\
&=: K\{D_{4j1}^E + D_{4j2}^E\}.
\end{aligned} \tag{23}$$

Further, using (V2), the Cauchy-Schwarz inequality, (20), Lemma 3.4, and (G1), we get

$$\begin{aligned}
D_{4j1}^E &\le \mathrm{E} \left\{ \max_{0\le k\le n-1} \|\sigma_j(\tilde{y}^E, t_k) - \sigma_j(z^E, t_k)\|^p \cdot \max_{0\le k\le n-1} \max_{1\le i\le m_k} |\Delta_{i,0}^k w_j|^p \right\} \\
&\le K \left(\mathrm{E} \max_{0\le k\le n-1} \|\tilde{y}^E(t_k) - z^E(t_k)\|^{2p} \cdot \mathrm{E} \max_{0\le k\le n-1} \max_{1\le i\le m_k} |\Delta_{i,0}^k w_j|^{2p} \right)^{1/2} \\
&\le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p \left(\mathrm{E} \max_{0\le k\le n-1} \sup_{t_k\le t\le t_{k+1}} |w_j(t) - w_j(t_k)|^{2p} \right)^{1/2} \\
&\le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p h^{p/2} (1+\ln n)^{p/2} \\
&\le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}} \right)^p ,
\end{aligned} \tag{24}$$

since $h(1+\ln n) = h(1+\ln \frac{T-t_0}{h}) \le K$ for $h \in (0, T-t_0]$. This estimate was done so roughly since, for the final result, here a better estimate than in (20) does not pay. This consideration applies also to the following estimate: By (V1), (G3), Lemma 3.5b), and (G1) it follows that

$$\begin{aligned} D^E_{4j2} &\le K \cdot \mathrm{E}\left\{ \max_{0\le k\le n-1} \max_{1\le i\le m_k} \left(\tfrac{h}{m_k}\right)^{p/2} \left| \sqrt{\tfrac{m_k}{h}} \Delta^k_{i,0} w_j - \sum_{s=1}^{i} \xi^k_{js} \right|^p \right\} \\ &\le K \left(\tfrac{h}{m(h)}\right)^{p/2} (1 + \ln n + \ln m(h))^p \le K \left(\tfrac{1+\ln n + \ln m(h)}{\sqrt{n \cdot m(h)}}\right)^p \qquad (25) \\ &\le K \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}}\right)^p . \end{aligned}$$

The last step is implied by

$$\begin{aligned} \tfrac{1+\ln\gamma+\ln\delta}{\sqrt{\gamma\delta}} &\le \tfrac{1+\ln\gamma+\ln\delta+\ln\gamma\ln\delta}{\sqrt{\gamma\delta}} = \left(\tfrac{1+\ln\gamma}{\sqrt{\gamma}}\right)\left(\tfrac{1+\ln\delta}{\sqrt{\delta}}\right) \\ &\le \tfrac{2}{\sqrt{e}} \left(\tfrac{1+\ln\delta}{\sqrt{\delta}}\right) \qquad \text{for all real} \quad \gamma, \delta \ge 1. \end{aligned} \qquad (26)$$

Now it follows from (21) – (25) that

$$\begin{aligned} &\mathrm{E} \max_{0\le k\le n-1} \max_{0\le i\le m_k} \|\tilde{y}^E(u^k_i) - z^E(u^k_i)\|^p \\ &\le K \left\{ \mathrm{E} \max_{0\le k\le n-1} \|\tilde{y}^E(t_k) - z^E(t_k)\|^p + \left(\tfrac{1+\ln m(h)}{\sqrt{m(h)}}\right)^p \right\}, \end{aligned}$$

and by (20) we get the assertion. ∎

Acknowledgement. We would like to thank Werner Römisch for helpful discussions and advice while working on this subject.

References

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. *J. Econometrics* **52**, 5–59.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1008.

Gelbrich, M. (1989). *L^p-Wasserstein-Metriken und Approximationen stochastischer Differentialgleichungen.* Dissertation A, Humboldt-Universität zu Berlin, Sektion Mathematik.

Gelbrich, M. (1990). On a Formula for the L^2 Wasserstein Metric between Measures on Euclidean and Hilbert Spaces. *Math. Nachr.* **147**, 185–203.

Gelbrich, M. (1995). Simultaneous time and chance discretization for stochastic differential equations. *J. Comput. Appl. Math.*, to appear.

GIKHMAN, I. I., and SKOROKHOD, A. W. (1977). *Introduction to the Theory of Stochastic Processes* (russian). Nauka, Moscow.

GIVENS, C. R., and SHORTT, R. M. (1984). A Class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31**, 231–240.

IKEDA, N., and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

JANSSEN, R. (1984). Discretization of the Wiener-process in difference-methods for stochastic differential equations. *Stochastic Proc. Appl.* **18**, 361–369.

KANAGAWA, S. (1986). The rate of convergence for approximate solutions of stochastic differential equations. *Tokyo J. Math.* **12**, 33–48.

KARIYA, T. (1993). *Quantitative Methods for Portfolio Analysis: MTV Model Approach.* Kluwer, Dordrecht.

KLOEDEN, P. E., and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin.

KOMLÓS, J., MAJOR, P., and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's, and the sample DF, I. *Z. Wahrsch. verw. Gebiete* **32**, 111–131.

KOMLÓS, J., MAJOR, P., and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF, II. *Z. Wahrsch. verw. Gebiete* **34**, 33–58.

MARUYAMA, G. (1955). Continuous Markov processes and stochastic equations. *Rend. Circ. Math. Palermo* **4**, 48–90.

MITTNIK, S., and RACHEV, S. T. (1993). Modeling asset returns with alternative stable distributions. *Econom. Rev.* **12**, (no. 3), 261–330.

PLATEN, E. (1981). An approximation method for a class of Itô processes. *Liet. Math. Rink.* **XXI** (vol. 1), 121–133.

RACHEV, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley, Chichester.

SHIRYAEV, A. N. (1984). *Probability Theory.* Springer, Berlin.

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR MATHEMATIK
UNTER DEN LINDEN 6
10099 BERLIN, GERMANY

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILILTY
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106
rachev@bernoulli.ucsb.edu