# WEAK DIFFUSION LIMITS OF DYNAMIC CONDITIONAL CORRELATION MODELS

CHRISTIAN M. HAFNER
*Université catholique de Louvain*

SEBASTIEN LAURENT
*Aix-Marseille University*

FRANCESCO VIOLANTE
*Aarhus University*

The properties of dynamic conditional correlation (DCC) models, introduced more than a decade ago, are still not entirely known. This paper fills one of the gaps by deriving weak diffusion limits of a modified version of the classical DCC model. The limiting system of stochastic differential equations is characterized by a diffusion matrix of reduced rank. The degeneracy is due to perfect collinearity between the innovations of the volatility and correlation dynamics. For the special case of constant conditional correlations, a nondegenerate diffusion limit can be obtained. Alternative sets of conditions are considered for the rate of convergence of the parameters, obtaining time-varying but deterministic variances and/or correlations. A Monte Carlo experiment confirms that the often used quasi-approximate maximum likelihood (QAML) method to estimate the diffusion parameters is inconsistent for any fixed frequency, but that it may provide reasonable approximations for sufficiently large frequencies and sample sizes.

## 1. INTRODUCTION

Continuous and discrete time volatility models are often considered as two competitive views to modeling financial time series. Thanks to the analytical tractability ensured by Itô calculus, continuous time models have played a central role in theoretical finance. The continuous time setting permits a deeper understanding of the properties of the corresponding discrete time model and to assess probabilistic and statistical properties of discrete time sequences such as stationarity, existence of moments or distributional results which are otherwise intractable in discrete time; see, Nelson (1990), Nelson and Foster (1994), and Nelson (1996).

**691**

From an applied viewpoint, inference on continuous time parameters of stochastic volatility models represents an important issue. The intractable likelihood functions and the unobservable volatility process require sophisticated estimation procedures. Several estimation methods have been proposed, such as the simulation based method of moments, Duffie and Singleton (1993), the quasi-indirect inference of Broze, Scaillet, and Zakoian (1998) or Bayesian Markov chain Monte Carlo methods, Jones (2003). Bollerslev, Engle, and Nelson (1994) and Ghysels, Harvey, and Renault (1996) provide exhaustive surveys on stochastic volatility models. Therefore, discrete time volatility models have been most often preferred by the applied econometrician. Rather than estimating and forecasting with a diffusion model observed at discrete points in time, it is in fact often easier to use a discrete model directly.

The theory of convergence of discrete time Markov sequences towards continuous time diffusion processes, see, e.g., Stroock and Varadhan (1979), Kushner (1984) and Ethier and Kurtz (1986), provides the theoretical foundation to establish mutual complementarities, possible interchangeability and connections between the two approaches. Nelson (1990) provides conditions ensuring the weak convergence of a discrete time Markov chain, defined by a system of stochastic difference equations, towards a diffusion, defined by a system of stochastic differential equations. The proposed approach requires the convergence, as the interval between observations shrinks to zero, of a number of conditional moments to well defined limits at an appropriate rate. In the context of GARCH-type models, Nelson (1990) illustrates the convergence through various GARCH specifications. This approach has been used by Duan (1997) to derive the diffusion limit of the Augmented GARCH model, by Fornari and Mele (1997) to study the continuous time behavior of the class of nonlinear ARCH models proposed by Ding, Granger, and Engle (1993), by Alexander and Lazar (2005) to derive the diffusion limit of a weak GARCH process and in a related setting by Trifi (2006) to illustrate the convergence results for the CEV-ARCH model of Fornari and Mele (2006) and the CMSV model of Jeantheau (2004) and Hobson and Rogers (1998). In the multivariate case, apart from Nelson (1996) in the context of asymptotic filtering theory, to our knowledge, the relationship between discrete and continuous time volatility and correlation models has not been addressed yet.

The potential advantage of the Nelson approximation approach lies essentially in estimation and forecasting. Considering the discrete time model as a diffusion approximation suggests inference on the parameters of the diffusion model via the parameter estimates of a discrete time GARCH-type model. Hence, a natural alternative to the direct estimation of the diffusion parameters consists in inferring the diffusion parameters by means of a tractable likelihood function of an approximating discrete time multivariate GARCH process. Following Fornari and Mele (2006), this approach is called estimation by quasi-approximated maximum likelihood (QAML). Requiring a feasible computational effort, this approach has been advocated, e.g., by Engle and Lee (1996), Lewis (2000), Barone-Adesi, Rasmussen, and Ravanelli (2005) and Stentoft (2011)

among others. Its computational advantage becomes important in the multivariate case, where volatility models within the conditional correlation class can be estimated in a straightforward two-step procedure, estimating first the conditional variances, then conditional correlations. However, the main drawback of estimation by QAML is the difficulty of proving its consistency even if the discrete time approximation is closed under temporal aggregation, see Drost and Nijman (1993) and Drost and Werker (1996). In the univariate GARCH case, Wang (2002) has shown that the statistical experiments resulting from the estimation of the diffusion model and its approximating discrete time model are not equivalent, which would imply inconsistency of the QAML estimator also in the multivariate case.

In this paper, we focus on conditional correlation models with GARCH dynamics for the variances of the marginal processes. We recover the diffusion limit of a modified version of the classical Dynamic Conditional Correlation (DCC) model of Engle (2002), called consistent DCC (cDCC), proposed by Aielli (2006). Unlike DCC, the cDCC model has a martingale difference property of the correlation dynamics and is therefore easier to treat from a theoretical viewpoint. For this specification and the general case of time-varying conditional correlations, we derive the existence of a degenerate weak diffusion limit under suitable convergence conditions for the model parameters. The degeneracy in the general case is due to the particular structure of the discrete time model in which the noise propagation systems of the variances and the one of the correlation driving process are perfectly correlated. This structure is preserved in the diffusion limit which is characterized by a diffusion matrix of reduced rank. More precisely, the diffusion of the variances and of the diagonal elements of the correlation driving process are pairwise governed by the same Brownian motion.

As a special case, we consider the Constant Conditional Correlation model (CCC) of Bollerslev (1990), which can be obtained from the cDCC under suitable parameter restrictions. The CCC-GARCH model is particularly interesting because, unlike the cDCC-GARCH process, it admits a nondegenerate diffusion and, in the bivariate specification, a closed form solution for the diffusion limit.

We then propose and discuss alternative sets of conditions regarding the speed of convergence of parameters of the cDCC-GARCH model. In this way, other types of degenerated diffusions can be obtained which are characterized by a stochastic price process while variances and/or correlations remain time varying but deterministic. In the same spirit of Corradi (2000), we then discuss what kind of processes can be obtained as Euler approximations of the alternative diffusion processes.

The paper is completed by a simulation study to investigate the performance of the QAML estimator of the diffusion parameters in our model framework. For the parameters characterizing the innovation in variances and in correlations, we find a negative bias in all cases, irrespective of the sample size, which only decreases as the time interval shrinks to zero. This confirms the results of Wang (2002) that care needs to be taken in inferring diffusion parameters from a discrete type approximation when there is no statistical equivalence of the likelihood estimators.

For the remaining model parameters, however, no substantial biases are found and the mean square error converges to zero as the sample size increases for a given time interval.

The paper is organized as follows. In Section 2 we study the continuous time behavior of the cDCC and CCC models. We also present the degenerate diffusions induced by a reparameterization of the convergence conditions. In Section 3 we illustrate through a Monte Carlo simulation our convergence results. In Section 4 we conclude and discuss directions for further research. Appendix A presents the general framework of the theorem of weak convergence of discrete time Markov chains. All proofs are provided in Appendix B.

## 2. MAIN THEORETICAL RESULTS

Let $Y_{kh}^{(h)}$ be an $N$-dimensional vector of logarithmic prices indexed by $kh$, $k \in \mathbb{N}$, $h > 0$. The superscript $(h)$ represents the time interval between two consecutive observations, i.e., for given $h$, prices are observed at times $h, 2h, 3h, \ldots$. We let the parameters depend on the sampling frequency. Furthermore, the variance of the innovations is made proportional to $h$. In this paper, we focus on the covariance stationary case, hence usual suitable positivity and stationarity constraints on the parameters of the variances and correlation driving process apply, see Bollerslev (1986), Engle (2002), Aielli (2006), and Aielli (2013).

Throughout the paper, we use the following notation: $||A||$ is the $L_2$ (Frobenius) norm of a matrix $A$, vec() stacks the columns of a matrix into a vector, vech() stacks the lower triangular portion of a square matrix into a vector, vechl() stacks the strictly lower triangular portion of a square matrix into a vector (i.e., excluding the diagonal elements), diag() stacks the diagonal of a square matrix into a vector. For any symmetric matrix $A$, the operation $a = \text{vech}(A)$ can be inverted and we write this as $A = \text{vech}^{-1}(a)$. Similarly, for the diag operator, we can define the inverse operator $\text{diag}^{-1}$ that inserts a vector into a diagonal matrix. Furthermore, $1_N$ is a $(N \times 1)$ vector of ones and $I_N$ is the $(N \times N)$ identity matrix. Let $\Omega$ denote the space of $N \times N$ matrices, and $\Omega' \subset \Omega$ the set of symmetric positive semidefinite $N \times N$ matrices. We also make use of the following elementary matrices: $D_N$ denotes the $(N^2 \times N(N+1)/2)$ duplication matrix, which for any symmetric matrix $A$ transforms vech($A$) into vec($A$), $D_N^+$ its generalized inverse, see, e.g., Lütkepohl (1996) for details, $I^*$ is defined such that $\text{diag}(A) = I^* \text{vech}(A)$ with $I^* = I^{+\prime} D_N$ and $I^+ = (1_N \otimes I_N) \odot [1_N' \otimes \text{vec}(I_N)]$ transforms vec($A$) into diag($A$). Finally, $I^-$ is defined such that $\text{vechl}(A)\text{vechl}(A)' = I^-(\text{vech}(A)\text{vech}(A)')I^{-\prime} = I^- D_N^+(A \otimes A)D_N^{+\prime}I^{-\prime}$.[1]

### 2.1. The cDCC-GARCH process

We consider a system of stochastic difference equations based on the discrete time cDCC-GARCH process of Aielli (2006) for the log price vector of an $N$-dimensional portfolio of assets $Y_{kh}^{(h)}$, $h > 0$ and $k \in \mathbb{N}$. Time is partitioned

more and more finely as in Appendix A, letting the parameters and the covariance matrix of innovations depend on the length $h$ of time intervals. The model is given by

$$Y_{kh}^{(h)} = Y_{(k-1)h}^{(h)} + S_{kh}^{(h)} \eta_{kh}^{(h)}, \tag{1}$$

$$V_{(k+1)h}^{(h)} = c_h + A_h h^{-1} S_{kh}^{(h)2} (\eta_{kh}^{(h)} \odot \eta_{kh}^{(h)}) + B_h V_{kh}^{(h)}, \tag{2}$$

$$Q_{(k+1)h}^{(h)} = \bar{Q}_h + \vartheta_h h^{-1} P_{kh}^{(h)} \eta_{kh}^{(h)} \eta_{kh}^{(h)\prime} P_{kh}^{(h)} + \gamma_h Q_{kh}^{(h)}, \tag{3}$$

$$P_{kh}^{(h)} = \left( Q_{kh}^{(h)} \odot I_N \right)^{1/2}, \tag{4}$$

$$R_{kh}^{(h)} = P_{kh}^{(h)-1} Q_{kh}^{(h)} P_{kh}^{(h)-1}, \tag{5}$$

where $\eta_{kh}^{(h)}$ is an $(N \times 1)$ vector of standardized but potentially conditionally correlated innovations such that $R_{kh}^{(h)-1/2} \eta_{kh}^{(h)} \sim$ i.i.d. N$(0, h I_N)$. Further, $S_{kh}^{(h)}$ is an $(N \times N)$ diagonal matrix of conditional standard deviations with the $(N \times 1)$ vector of conditional variances denoted by $V_{kh}^{(h)} = \text{diag}(S_{kh}^{(h)2})$. For the correlation driving process $Q_{kh}^{(h)}$ we have, under the restrictions $\bar{Q}_h \in \Omega'$, $\vartheta_h, \gamma_h \geq 0$, that $Q_{kh}^{(h)} \in \Omega'$. We will stack the nonredundant elements of $Q_{kh}^{(h)}$ into the vector $q_{kh}^{(h)} = \text{vech}(Q_{kh}^{(h)})$. Note that, for a given $h$, the vector $X_{kh}^{(h)} := \left( Y_{kh}^{(h)}, V_{(k+1)h}^{(h)}, q_{(k+1)h}^{(h)} \right)'$ is a discrete time Markov process with filtration $\mathcal{F}_{kh} := \sigma \left( X_{sh}^{(h)}, s < k \right)$, so that the theory of Appendix A applies.

Note also that the discrete time cDCC-GARCH model of Aielli (2006) can be obtained by setting $h = 1$. The standard DCC model of Engle (2002) is very similar but instead of (3) specifies $Q_{(k+1)h}^{(h)} = \bar{Q}_h + \vartheta_h h^{-1} \eta_{kh}^{(h)} \eta_{kh}^{(h)\prime} + \gamma_h Q_{kh}^{(h)}$. The advantage of the cDCC model is that the recursion for $Q_{kh}^{(h)}$ preserves the martingale difference property, i.e., $h^{-1} \text{E}\left[ \left( P_{kh}^{(h)} \eta_{kh}^{(h)} \eta_{kh}^{(h)\prime} P_{kh}^{(h)} \right) - Q_{kh}^{(h)} | \mathcal{F}_{kh} \right] = 0$, so that, for a given $h$, the process $\left\{ h^{-1/2} P_{kh}^{(h)} \eta_{kh}^{(h)}, q_{kh}^{(h)} \right\}$ is a multivariate semistrong GARCH process in the sense of Drost and Nijman (1993) and Hafner (2008).

Without loss of generality, we reparameterize the drift in the recursion $Q_{kh}^{(h)}$ as a combination of a frequency invariant component and frequency dependent parameters. The drift $\bar{Q}_h$ can be expressed as $\bar{Q}_h = (1 - \vartheta_h - \gamma_h) \bar{Q}$.[2] As shown by Aielli (2013), $\bar{Q}$ is only identified up to scale so that we restrict the diagonal elements of $\bar{Q}$ to one. The reparameterization will be particularly useful when deriving the diffusion limit of the constant conditional correlation (CCC) model of Bollerslev (1990). In fact, under the parameter restriction $\vartheta_h = \gamma_h = 0$, $Q_{kh}^{(h)} = \bar{Q}_h = \bar{Q}$, and therefore, $R_{kh}^{(h)} = \bar{Q}$ for all $h$. Thus, in the CCC model, $\bar{Q}$ is the frequency-invariant unconditional correlation matrix of $\left( Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)} \right)$. We denote by $\bar{q} = \text{vech}(\bar{Q})$ the nonredundant elements of $\bar{Q}$.

Deriving the diffusion limit of the cDCC-GARCH process requires to assume convergence rates of the discrete time parameters such that the first two

conditional moments converge with increasing sampling frequency, as required by Assumption A.1. For the discrete time cDCC-GARCH process $X_{kh}^{(h)} = \left(Y_{kh}^{(h)}, V_{(k+1)h}^{(h)}, q_{(k+1)h}^{(h)}\right)'$ given in (1)–(3), our main result in Theorem 1 will assume the following convergence rates:

$$c_h = h\, c + o(h),\tag{6}$$

$$I_N - A_h - B_h = h\, \Lambda + o(h),\tag{7}$$

$$A_h = \sqrt{h}\, A + o(\sqrt{h}),\tag{8}$$

$$\bar{Q}_h = \bar{Q}\phi h + o(h),\tag{9}$$

$$1 - \vartheta_h - \gamma_h = h\, \phi + o(h),\tag{10}$$

$$\vartheta_h = \sqrt{h}\, \vartheta + o(\sqrt{h}),\tag{11}$$

where $c$ is an $(N \times 1)$ vector, $A$ and $\Lambda$ are diagonal $N \times N$ matrices with positive diagonal elements, $\bar{Q}$ is a positive definite $N \times N$ matrix, and $\phi, \vartheta$ are positive scalars.

As shown in the proof of 1, these convergence rates ensure that the first and the second conditional moments per unit of time converge, as $h \to 0$, to well-behaved limits and that the first difference of the process $\left[Y_{kh}^{(h)\prime}, V_{(k+1)h}^{(h)\prime}, q_{(k+1)h}^{(h)\prime}\right]'$ satisfies Assumption A.1.

Note that $c > 0$ (elementwise) ensures positivity of the variance process, $A > 0$ and $\vartheta > 0$ ensure that the rescaled second conditional moment does not vanish as $h \to 0$, while $\Lambda > 0$ and $\phi > 0$ ensure covariance stationarity of the return process.[3]

Under our assumptions, we have the following result for the diffusion limit of the cDCC-GARCH process $X_{kh}^{(h)} = \left(Y_{kh}^{(h)}, V_{(k+1)h}^{(h)}, q_{(k+1)h}^{(h)}\right)'$.

THEOREM 1 (Diffusion limit of the cDCC-GARCH model). *Assume that the initial value $X_{0h}^{(h)}$ converges in distribution to a random vector $X_0$ as $h \to 0$. Under (6) to (11), the discrete time cDCC-GARCH process $X_{kh}^{(h)}$ given in (1)–(3) weakly converges to the diffusion process $X_t = \left[Y_t', V_t', q_t'\right]'$ which is the solution to the system of stochastic differential equations*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,\tag{12}$$

*where $W_t$ is an $N(N+5)/2$-dimensional vector of mutually independent standard Brownian motions, independent from $X_0$. The drift, $b(X_t)$, is given by*

$$b(X_t) = \begin{bmatrix} 0_N \\ c - \Lambda V_t \\ \phi(\bar{q} - q_t) \end{bmatrix},\tag{13}$$

*while the scale, $\sigma(X_t)$, is a continuous mapping such that, for all $X_t \in \mathbb{R}^{N(N+5)/2}$ and $t \geq 0$, $a(X_t) = \sigma(X_t)\sigma(X_t)'$, where $a(X_t)$ is given by*

$$a(X_t) = \begin{bmatrix} a_{YY} & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N \times N)} & a_{VV} & a_{Vq} \\ 0_{(N^* \times N)} & a'_{Vq} & a_{qq} \end{bmatrix}, \tag{14}$$

*with $N^* := N(N+1)/2$ and*

$$
\begin{aligned}
a_{YY} &= S_t R_t S_t, \\
a_{VV} &= 2AS_t^2(R_t \odot R_t)S_t^2 A, \\
a_{Vq} &= \vartheta \left[ I^* K_t D_N^+(P_t \otimes P_t)D_N^{+\prime} - 1_N q_t' \right]' S_t^2 A, \\
a_{qq} &= \vartheta^2 \left[ D_N^+(P_t \otimes P_t)D_N^{+\prime} K_t D_N^+(P_t \otimes P_t)D_N^{+\prime} - q_t q_t' \right],
\end{aligned}
$$

*where $K_t = 2D_N^+(R_t \otimes R_t)D_N^{+\prime} + vech(R_t)vech(R_t)'$, $Q_t = vech^{-1}(q_t)$, $P_t = (I_N \odot Q_t)^{1/2}$, $R_t = P_t^{-1} Q_t P_t^{-1}$, and $S_t = diag^{-1}(V_t^{1/2})$. The matrix $a(X_t)$ is singular and its rank is equal to $N(N+3)/2 < \dim(X_t) = N(N+5)/2$.*

Note first that the drift term $b(X_t)$ is linear in $X_t$, which is due to the fact that the cDCC-GARCH process satisfies a semistrong GARCH structure, meaning that increments to the state variables have a conditional mean that is linear in the state. In particular, as shown in the proof, we can use that in the cDCC model $E\left[ \Delta q_{(k+1)h}^{(h)} | \mathcal{F}_{kh} \right] = (1 - \vartheta_h - \gamma_h)(\bar{q} - q_{kh}^{(h)})$. This is, however, not the case in the DCC model, where this expectation would be a function of the conditional correlation matrix $R_{kh}^{(h)}$, which is a nonlinear function of the state variable $q_{kh}^{(h)}$. This is the reason why it is tedious to obtain analytical results for the diffusion limit of the DCC model, and in fact more generally why there are few results on the properties of this model such as stationarity or moment conditions.

Note further that one of the examples of Nelson (1990), p.15, of a univariate GARCH(1,1) model setting $c = 0$ in his equation (2.20), can be recovered as a special case of Theorem 1. His drift and diffusion terms (2.37) and (2.38) are the reduced versions of our (13) and (14) terms in the univariate case. Note however the difference in terms of the parameter scaling: Nelson (1990) lets $\alpha_h = \alpha(h/2)^{1/2}$, so that the scaling factor $\sqrt{2}$ should be included in his results to compare them with ours.

The singularity of $a(X_t)$ is due to the particular structure of the model in which the noise propagation of the variance processes and the one of the diagonal elements of the correlation driving processes are pairwise perfectly correlated. This is because, although (possibly) different in terms of level and dynamics, (2) and (3) are driven by the same innovations. For example, in the special case where $(I_N - A_h - B_h)^{-1} c_h = diag(\bar{Q})$, $A_h = \vartheta_h I_N$, $B_h = \gamma_h I_N$, the model reduces to a scalar VEC model with $N$ redundant equations.

To investigate the implications of singularity of the diffusion matrix $a(X_t)$, let us rearrange the order of the elements of the diffusion process $X_t$ as $\left[Y_t', V_t', q_t^{(d)\prime}, q_t^{(l)\prime}\right]'$, where $q_t^{(d)} = \mathrm{diag}(Q_t)$ and $q_t^{(l)} = \mathrm{vechl}(Q_t)$. The two partial diffusion processes $\left[Y_t', V_t', q_t^{(l)\prime}\right]'$ and $\left[Y_t', q_t^{(d)\prime}, q_t^{(l)\prime}\right]'$ share the same correlation structure, while $\mathrm{Corr}(dV_{t,i}, dQ_{t,ii}) = 1$, $i = 1, \ldots, N$, implies that the two partial diffusions are driven by the same vector of Brownian innovations. Thus, the relevant part in terms of the noise propagation system of the diffusion limit of the cDCC-GARCH process consists of a system of $N(N+3)/2$ stochastic differential equations, either $\left[Y_t', V_t', q_t^{(l)\prime}\right]$ or $\left[Y_t', q_t^{(d)\prime}, q_t^{(l)\prime}\right]$, while the remaining $N$ diffusion processes, $q_t^{(d)}$ or $V_t$, respectively, are characterized by a specific deterministic part (drift) but a common, although appropriately rescaled, stochastic component. To illustrate this point, let us consider the following partition of the diffusion matrix in (14), whose elements have been appropriately reordered,

$$
a(X_t) = \begin{bmatrix}
a_{YY} & 0_{(N\times N)} & 0_{(N\times N)} & 0_{(N\times N^+)} \\
0_{(N\times N)} & a_{VV} & a_{Vq^{(d)}} & a_{Vq^{(l)}} \\
0_{(N\times N)} & a'_{Vq^{(d)}} & a_{q^{(d)}q^{(d)}} & a_{q^{(d)}q^{(l)}} \\
0_{(N^+\times N)} & a'_{Vq^{(l)}} & a'_{q^{(d)}q^{(l)}} & a_{q^{(l)}q^{(l)}}
\end{bmatrix},
\tag{15}
$$

where $N^+ := N(N-1)/2$ and

$$
a_{Vq^{(d)}} = 2\vartheta A S_t^2 (R_t \odot R_r) P_t^2 = \vartheta a_{VV} \left(S_t^2 A\right)^{-1} P_t^2,
$$

$$
a_{q^{(d)}q^{(d)}} = 2\vartheta^2 P_t^2 (R_t \odot R_r) P_t^2 = \vartheta^2 P_t^2 \left(A S_t^2\right)^{-1} a_{VV} \left(S_t^2 A\right)^{-1} P_t^2,
$$

$$
a_{Vq^{(l)}} = \vartheta A S^2 \left[I^* K_t D_N^+ (P_t \otimes P_t) D_N^{+\prime} I^{-\prime} - 1_N q_t' I^{-\prime}\right],
$$

$$
a_{q^{(d)}q^{(l)}} = \vartheta^2 P_t^2 \left[I^* K_t D_N^+ (P_t \otimes P_t) D_N^{+\prime} I^{-\prime} - 1_N q_t' I^{-\prime}\right] = \vartheta P_t^2 \left(A S_t^2\right)^{-1} a_{Vq^{(l)}},
$$

$$
a_{q^{(l)}q^{(l)}} = \vartheta^2 I^- \left[D_N^+ (P_t \otimes P_t) D_N^{+\prime} K_t D_N^+ (P_t \otimes P_t) D_N^{+\prime} - q_t q_t'\right] I^{-\prime}.
$$

Let us also define $C_t = \vartheta P_t^2 (A S_t^2)^{-1}$. We can rewrite (15) as

$$
\begin{bmatrix}
a_{YY} & 0_{(N\times N)} & 0_{(N\times N)} & 0_{(N\times N^+)} \\
0_{(N\times N)} & a_{VV} & a_{VV} C_t' & a_{Vq^{(l)}} \\
0_{(N\times N)} & C_t a_{VV} & C_t a_{VV} C_t' & C_t a_{Vq^{(l)}} \\
0_{(N^+\times N)} & a'_{Vq^{(l)}} & a'_{Vq^{(l)}} C_t' & a_{q^{(l)}q^{(l)}}
\end{bmatrix}.
\tag{16}
$$

Therefore, the partial diffusion processes $\left[Y_t', V_t', q_t^{(l)\prime}\right]'$ and $\left[Y_t', q_t^{(d)\prime}, q_t^{(l)\prime}\right]'$ are characterized by the diffusion matrices $a_1(\cdot)$ and $a_2(\cdot)$, respectively, given by

$$
a_1\left(\left[Y_t', V_t', q_t^{(l)\prime}\right]'\right) = \begin{bmatrix}
a_{YY} & 0_{(N\times N)} & 0_{(N\times N^+)} \\
0_{(N\times N)} & a_{VV} & a_{Vq^{(l)}} \\
0_{(N^+\times N)} & a'_{Vq^{(l)}} & a_{q^{(l)}q^{(l)}}
\end{bmatrix},
\tag{17}
$$

$$a_2([Y_t', q_t^{(d)\prime}, q_t^{(l)\prime}]') = \begin{bmatrix} a_{YY} & 0_{(N \times N)} & 0_{(N \times N^+)} \\ 0_{(N \times N)} & C_t a_{VV} C_t' & C_t a_{Vq^{(l)}} \\ 0_{(N^+ \times N)} & a_{Vq^{(l)}}' C_t' & a_{q^{(l)}q^{(l)}} \end{bmatrix}. \tag{18}$$

The structure of (17) and (18) shows that the two partial processes $[V_t', q_t^{(l)\prime}]'$ and $[q_t^{(d)\prime}, q_t^{(l)\prime}]'$, both uncorrelated with $Y_t$, share the same correlation structure. Furthermore, from Theorem 1, it immediately follows that $[V_t', q_t^{(l)\prime}]'$ and $[q_t^{(d)\prime}, q_t^{(l)\prime}]'$ are elementwise perfectly correlated. In fact, more generally, $\mathrm{Corr}(dV_{t,i}, dq_{t,j}^{(d)}) = (R_t \odot R_t)_{ij}$, $i, j = 1, \ldots, N$. The perfect correlation between the variances and the diagonal elements of $Q$ is intuitively obvious also from the discrete time model in (1)–(5), as the driving innovations in (2) and the diagonal elements of (3) are both given by $\eta_{kh}^{(h)} \odot \eta_{kh}^{(h)}$. Although either partial diffusion process $[Y_t', V_t', q_t^{(l)\prime}]'$ or $[Y_t', q_t^{(d)\prime}, q_t^{(l)\prime}]'$ is sufficient alone to fully characterize the noise propagation system of the cDCC diffusion limit, they are both necessary to obtain the distribution of $Y_t$ which depends on both $V_t$ and $q_t = [q_t^{(d)\prime}, q_t^{(l)\prime}]'$ through the correlation process $R_t$.

The degeneracy of the diffusion matrix may have consequences for the estimation method of the parameters of the diffusion process. For example, if maximum likelihood is used as, e.g., in Li (2013) which requires a positive definite diffusion matrix, estimation of the full system would be infeasible in our case.

### 2.2. A Special Case: The CCC-GARCH Process

As a special case, consider the Constant Conditional Correlation (CCC) model of Bollerslev (1990). The cDCC process nests the CCC process under the following parameter restrictions

$$\vartheta_h = \gamma_h = 0 \quad \text{for all } h > 0.$$

The innovation vector $\eta_{kh}^{(h)}$ in the CCC-GARCH process is an $(N \times 1)$ vector of standardized but potentially correlated innovations, such that $\eta_{kh}^{(h)} \sim \mathrm{N}(0, hR)$, where $R$ represents the (frequency invariant) constant conditional correlation matrix. This model, although rather restrictive in practice, is particularly interesting because, unlike the cDCC-GARCH process, it allows for a nondegenerate diffusion and, in the bivariate specification, a closed form solution for the diffusion limit. As a corollary to Theorem 1, Assumption A.1 holds under the following convergence rates for the parameters of the discrete time CCC-GARCH process

$$c_h = h\,c + o(h), \tag{19}$$

$$I_N - A_h - B_h = h\,\Lambda + o(h), \tag{20}$$

$$A_h = \sqrt{h}\,A + o(\sqrt{h}), \tag{21}$$

for some $(N \times 1)$ vector $c$ and $(N \times N)$ diagonal matrices $A$ and $\Lambda$ with positive and finite elements.

The following corollary to Theorem 1 states the result of a diffusion limit of the CCC-GARCH process $X_{kh}^{(h)} = \big(Y_{kh}^{(h)}, V_{(k+1)h}^{(h)}\big)'$.

COROLLARY 1 (Diffusion limit of the CCC-GARCH model). *Assume that the initial value* $X_{0h}^{(h)}$ *converges in distribution to a random vector* $X_0$ *as* $h \to 0$. *Under the convergence conditions in (19)–(21), the CCC-GARCH process weakly converges to the nondegenerate diffusion process* $X_t = [Y_t', \ V_t']'$ *which is a solution to a system of stochastic differential equations of the form (12), with drift*

$$b(X_t) = \begin{bmatrix} 0_N \\ c - \Lambda V_t \end{bmatrix} \tag{22}$$

*and diffusion matrix*

$$a(X_t) = \begin{bmatrix} S_t R S_t & 0_{(N \times N)} \\ 0_{(N \times N)} & 2A S_t^2 (R \odot R) S_t^2 A \end{bmatrix} \tag{23}$$

*and driven by a vector* $W_t$ *of* $2N$ *mutually independent Brownian motions, independent of the initial value* $X_0 = [Y_0, \ V_0]'$.

The diffusion limit of the CCC model is clearly nondegenerate because it is driven by as many Brownian motions as the number of variables in the system and whose covariance matrix is nonsingular.

It is clear that the diffusion limit of the cDCC-GARCH process (as well as that of the CCC-GARCH process) is a continuous time stochastic volatility model (i.e., stochastic variances and correlations). We discuss next the case when rates of convergence other than the ones introduced in Theorem 1, but still satisfying Assumption A.1, are used.

## 2.3. Alternative Convergence Conditions

In this section we reconsider the continuous time approximation of the cDCC-GARCH process (1)–(3). The convergence rate $h^{1/2}$, suggested in Theorem 1, represents the slowest rate of convergence for the parameters $A_h$ and $\vartheta_h$ satisfying Assumption A.1. More generally, the rate $h^{1/2}$ represents the only rate ensuring that the second conditional moments $\mathrm{Var}\big(V_{(k+1)h}^{(h)} - V_{kh}^{(h)}|\mathcal{F}_{kh}\big)$, $\mathrm{Var}\big(q_{(k+1)h}^{(h)} - q_{kh}^{(h)}|\mathcal{F}_{kh}\big)$ and $\mathrm{Cov}\big[(V_{(k+1)h}^{(h)} - V_{kh}^{(h)}), (q_{(k+1)h}^{(h)} - q_{kh}^{(h)})|\mathcal{F}_{kh}\big]$ scaled by $h^{-1}$, do not vanish as $h \to 0$. As shown in Theorem 1, the resulting diffusion limit is characterized by stochastic variances of the marginal processes and a stochastic correlation driving process.

However, there are other admissible convergence rates for $A_h$ and $\vartheta_h$ which also satisfy Assumption A.1. Thus, depending on the continuous time approximation we consider, we can recover different types of diffusion for the process

(1)–(3). Consider the following convergence rates for the parameters $A_h$ and $\vartheta_h$:

$$\lim_{h \to 0} h^{-\left(\frac{1}{2}+\delta_1\right)} A_h = \tilde{A} < \infty \tag{24}$$

and

$$\lim_{h \to 0} h^{-\left(\frac{1}{2}+\delta_2\right)} \vartheta_h = \tilde{\vartheta} < \infty, \tag{25}$$

for some $(N \times N)$ diagonal matrix $\tilde{A} > 0$ (elementwise), $\tilde{\vartheta} > 0$, $\delta_1 \geq 0$ and $\delta_2 \geq 0$. Note that under (24) and (25), $A_h$ and $\vartheta_h$ are of order $h^{1/2+\delta_1}$ and $h^{1/2+\delta_2}$, respectively. Clearly, the special case $\delta_1 = \delta_2 = 0$ is covered by Theorem 1.

As shown in the proof of Theorem 2, Assumption A.1 holds under (6), (7), (9), (10), (24), and (25). The implications are straightforward. If either $\delta_1 > 0$ or $\delta_2 > 0$, then the terms depending on $\eta_{kh}^{(h)}$ on the right hand side of (2) and/or (3) are of order $o(h^{1/2})$. Consequently, the conditional second moments scaled by $h^{-1}$ converge to zero as $h \to 0$. The resulting diffusion limits are degenerate and are characterized by time varying but deterministic variances of the marginal processes and/or a deterministic correlation driving process. We have the following results.

THEOREM 2 (Alternative convergence conditions). *Assume that the initial value $X_{0h}^{(h)}$ converges in distribution to a random vector $X_0$ as $h \to 0$. Replacing in the assumptions of Theorem 1 either (8) by (24), or (11) by (25), or both, the discrete time cDCC-GARCH process (1)–(3) admits a degenerate diffusion limit. The diffusion process $X_t = [Y_t', V_t', q_t']'$ is the solution to a system of stochastic differential equations of the form (12), with drift given by (13) and diffusion matrix given, respectively, by*

  i) *(deterministic variances but stochastic correlation) under (6), (7), (10), (11), and (24)*

$$a(X_t) = \begin{bmatrix} S_t R_t S_t & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N \times N)} & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N^* \times N)} & 0_{(N^* \times N)} & \vartheta^2 [D_N^+(P_t \otimes P_t) D_N^{+'} K_t \\ & & D_N^+(P_t \otimes P_t) D_N^{+'} - q_t q_t'] \end{bmatrix}. \tag{26}$$

  *The diffusion process defined by (12), (13), and (26) is driven by $N(N+3)/2$ independent standard Brownian motions;*

  ii) *(stochastic variance but deterministic correlation) under (6), (7), (8), (10), and (25)*

$$a(X_t) = \begin{bmatrix} S_t R_t S_t & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N \times N)} & 2 A S_t^2 (R_t \odot R_r) S_t^2 A & 0_{(N \times N^*)} \\ 0_{(N^* \times N)} & 0_{(N \times N)} & 0_{(N^* \times N^*)} \end{bmatrix}. \tag{27}$$

  *The diffusion process defined by (12), (13), and (27) is driven by $2N$ independent standard Brownian motions;*

iii) *(deterministic variances and correlation) under (6), (7), (10), and both (24), and (25)*

$$a(X_t) = \begin{bmatrix} S_t R_t S_t & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N \times N)} & 0_{(N \times N)} & 0_{(N \times N^*)} \\ 0_{(N^* \times N)} & 0_{(N^* \times N)} & 0_{(N^* \times N^*)} \end{bmatrix}. \tag{28}$$

*The diffusion process defined by (12), (13), and (28) is driven by $N$ independent standard Brownian motions.*

It is possible to characterize the types of processes that can be obtained as Euler approximation of the different diffusions in Theorem 2. These approximations are not unique. For example, in the univariate GARCH case, Corradi (2000) has shown that an Euler approximation of a degenerate diffusion process is GARCH, while that of a nondegenerate diffusion is stochastic volatility. In the same spirit, and using stochastic calculus results of Steele (2001) p.123, we can show that the following type of processes are Euler approximations of the three diffusions defined in Theorem 2: $i$) a process with stochastic correlation and GARCH variances, $ii$) a process with stochastic volatility and cDCC correlation, and $iii$) a cDCC-GARCH process as in (1)–(3), respectively.

Note further that the results in i) and iii) of Theorem 2 are generalizations of Proposition 2.1 (i) of Corradi (2000) to the multivariate case. Her Proposition 2.1 (ii) corresponds to the Nelson result where the variance process is stochastic. One generalization to the multivariate case was presented in Theorem 1, but Theorem 2, ii) gives an alternative generalization using parameter convergence conditions as in Nelson (1990) for the variances, and as in Corradi (2000) for the correlations.

## 3. MONTE CARLO EVIDENCE ON ESTIMATION BY APPROXIMATION

In this section we investigate the performance of the quasi-approximate maximum likelihood (QAML) procedure of Fornari and Mele (2006), discussed in the introduction, in our model framework using a Monte Carlo simulation study. We infer the diffusion parameters from the estimation of an approximating discrete time cDCC model using the same three-step estimator as in Aielli (2013), Definition 3.4.

Estimation by QAML essentially involves two types of biases: First, the *finite sample bias* due to the availability of a sample of only a finite number of observations. The second, called *approximation bias*, arises from the approximation of an exact, but unknown, discrete time representation of the underlying diffusion process. The approximating model is not even an Euler discretization of the diffusion model, but its sample paths converge weakly (in distribution) to the diffusion process. Fornari and Mele (2006) consider bias correction methods based on the indirect inference principle developed by Broze, Scaillet, and Zakoian (1998) and compare them with the not bias-corrected QAML estimator.

For the drift parameter of a Cox–Ingersoll–Ross type process, Phillips and Yu (2009) have shown that the approximation bias for alternative approximation schemes is typically negligible compared to the finite sample bias. This motivates the use of QAML for GARCH-type processes, where the exact discrete time model is unknown. Rather than comparing with alternative estimation strategies, e.g, simulated MLE and indirect inference as in Kleppe, Yu, and Skaug (2014), we focus on the properties of the simple QAML procedure and, in particular, the relative importance of approximation and estimation biases.

For univariate GARCH models, Wang (2002) has shown the nonequivalence of the statistical experiments resulting from the estimation of the discrete time model and its weak diffusion limit. Nevertheless, many studies have used QAML; see, e.g., Engle and Lee (1996), Broze, Scaillet, and Zakoian (1998), Lewis (2000), Barone-Adesi, Rasmussen, and Ravanelli (2005) and Stentoft (2011), arguing that the approximation bias tends to disappear as the frequency increases. For the related case of estimating temporally aggregated multivariate GARCH models, the bias of QAML has been shown to be negligible, see Hafner and Rombouts (2007). Therefore, it is of interest to see whether this finding extends to the estimation of some or all parameters of the cDCC-GARCH diffusion limit.

We estimate the parameters of a sequence, indexed by $h$, of discrete time cDCC-GARCH models with i.i.d. innovations. Then, for each $h$, we use the relationships given in Theorem 1 to obtain the diffusion parameters and we investigate the behavior of the latter as $h \to 0$. To keep the computational burden feasible, we focus on the bivariate case, $N = 2$, but our results should generalize in an obvious way to higher dimensions. Using the representations of Section 2.1, the cDCC-GARCH diffusion limit can be written as

$$\begin{bmatrix} dY_{1t} \\ dY_{2t} \end{bmatrix} = \begin{bmatrix} \sqrt{V_{1t}} & 0 \\ 0 & \sqrt{V_{2t}} \end{bmatrix} \Upsilon^{(1)}(\rho_t)^{\frac{1}{2}} dW_t^{(1)}, \tag{29}$$

$$\begin{bmatrix} dV_{1t} \\ dV_{2t} \\ dQ_{12t} \end{bmatrix} = \begin{bmatrix} c_1 - \Lambda_{11} V_{1t} \\ c_2 - \Lambda_{22} V_{2t} \\ \phi(\bar{Q}_{12} - Q_{12t}) \end{bmatrix} dt$$

$$+ \sqrt{2} \begin{bmatrix} A_{11} V_{1t} & 0 & 0 \\ 0 & A_{22} V_{2t} & 0 \\ 0 & 0 & \vartheta Q_{12t} \sqrt{\frac{1+\rho_t^2}{2\rho_t^2}} \end{bmatrix} \Upsilon^{(2)}(\rho_t)^{\frac{1}{2}} dW_t^{(2)}, \tag{30}$$

$$\begin{bmatrix} dQ_{11t} \\ dQ_{22t} \\ dQ_{12t} \end{bmatrix} = \begin{bmatrix} \phi(\bar{Q}_{11} - Q_{11t}) \\ \phi(\bar{Q}_{22} - Q_{22t}) \\ \phi(\bar{Q}_{12} - Q_{12t}) \end{bmatrix} dt$$

$$+ \sqrt{2}\vartheta \begin{bmatrix} Q_{11t} & 0 & 0 \\ 0 & Q_{22t} & 0 \\ 0 & 0 & Q_{12t} \sqrt{\frac{1+\rho_t^2}{2\rho_t^2}} \end{bmatrix} \Upsilon^{(2)}(\rho_t)^{\frac{1}{2}} dW_t^{(2)}, \tag{31}$$

where

$$\Upsilon^{(1)}(\rho_t) = \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix}, \qquad \Upsilon^{(2)}(\rho_t) = \begin{bmatrix} 1 & \rho_t^2 & \sqrt{\dfrac{2\rho_t^2}{1+\rho_t^2}} \\[2mm] \rho_t^2 & 1 & \sqrt{\dfrac{2\rho_t^2}{1+\rho_t^2}} \\[2mm] \sqrt{\dfrac{2\rho_t^2}{1+\rho_t^2}} & \sqrt{\dfrac{2\rho_t^2}{1+\rho_t^2}} & 1 \end{bmatrix}$$

and

$$\rho_t = \frac{Q_{12t}}{\sqrt{Q_{11t}Q_{22t}}}.$$

We use an Euler discretization scheme of (29)–(31) and simulate 5000 sample paths using a discretization interval $\Delta t = 1/8192$ and length $k$ ranging from 250 to 2000 periods. The data is generated using the following parameterization: $c = [0.1, 0.15]'$, $A_{11} = 0.07$, $A_{22} = 0.10$, $\Lambda_{11} = 0.13$, $\Lambda_{22} = 0.10$, $\vartheta = 0.08$, and $\phi = 0.04$. The diagonal elements of $\bar{Q}$ are fixed to one for identification as in Aielli (2013), while the off-diagonal element, $\bar{Q}_{12}$, is set to 0.5. For sake of comparison with the drift parameter of the variance equations, we report results in terms of the off-diagonal element of the drift matrix $\bar{Q}_h$, i.e., $\bar{Q}_{12,h}$. The square roots of the correlation matrices of the diffusion, $\Upsilon^{(1)}(\rho_t)$ and $\Upsilon^{(2)}(\rho_t)$, are computed by spectral decomposition.

For each sample path we estimate the model (1)–(5) with $c_h = ch$, $A_h = A\sqrt{h}$, $B_h = I_N - A\sqrt{h} - \Lambda h$, $\bar{Q}_h = \bar{Q}\phi h$, $\vartheta_h = \vartheta\sqrt{h}$, and $\gamma_h = 1 - \vartheta\sqrt{h} - \phi h$. The model is estimated by Gaussian QAML using data sampled at nine frequencies spanning from $h = 1/4$ to $h = 1/512$. The bias and variance of parameter estimates are reported in Tables 1 and 2, respectively. To economize on space, only the results for $1/h \in \{8, 32, 128, 512\}$ are reported, the complete set of results is available in the working paper version of this paper, Hafner, Laurent, and Violante (2016).

Note first that the bias is negative for the innovation parameters and positive for the drift and persistence parameters, which confirms the results of Aielli (2013) for the analogous discrete time parameterization. As the sampling frequency increases, the bias vanishes and the MSE decreases at an appropriate rate for all parameters. For a given frequency, however, there are remarkable differences. For the parameters in the drift terms, bias and MSE decrease as the sample size $k$ increases, suggesting that the finite sample bias dominates the approximation bias, which confirms the results of Phillips and Yu (2009) for the Cox–Ingersoll–Ross diffusion process. However, this is not the case for the parameters $A$ and $\vartheta$ linked to the innovation terms in $V_t$ and $Q_t$, respectively, for which the approximation bias dominates the finite sample bias. Clearly, QAML is inconsistent when only the sample size is increased but not the frequency, which confirms the results of

**TABLE 1.** QAML biases of cDCC diffusion approximation at varying frequencies $1/h$ and number of time periods $k$

| $1/h$ | $\hat{c}_{1,h}$ | $\hat{A}_{11,h}$ | $\hat{\Lambda}_{11,h}$ | $\hat{c}_{2,h}$ | $\hat{A}_{22,h}$ | $\hat{\Lambda}_{22,h}$ | $\hat{\bar{Q}}_{12,h}$ | $\hat{\vartheta}_h$ | $\hat{\phi}_h$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $k=250$ | | | | | |
| 8 | 1.02e-2 | −1.00e-2 | 1.03e-2 | 9.45e-3 | −1.08e-2 | 6.64e-3 | 1.97e-3 | −5.65e-3 | 3.90e-3 |
| 32 | 1.42e-3 | −3.05e-3 | 1.43e-3 | 1.41e-3 | −3.09e-3 | 1.02e-3 | 3.30e-4 | −1.55e-3 | 6.43e-4 |
| 128 | 2.22e-4 | −7.71e-4 | 2.21e-4 | 2.62e-4 | −8.39e-4 | 1.90e-4 | 6.98e-5 | −3.77e-4 | 1.29e-4 |
| 512 | 4.40e-5 | −1.61e-4 | 4.31e-5 | 5.55e-5 | −2.22e-4 | 4.00e-5 | 1.55e-5 | −8.24e-5 | 2.66e-5 |
| | | | | $k=500$ | | | | | |
| 8 | 5.95e-3 | −1.02e-2 | 5.99e-3 | 4.95e-3 | −1.11e-2 | 3.50e-3 | 8.94e-4 | −5.61e-3 | 1.75e-3 |
| 32 | 7.09e-4 | −3.02e-3 | 7.07e-4 | 7.83e-4 | −3.19e-3 | 5.67e-4 | 1.76e-4 | −1.45e-3 | 3.36e-4 |
| 128 | 1.15e-4 | −7.55e-4 | 1.13e-4 | 1.49e-4 | −8.80e-4 | 1.10e-4 | 3.90e-5 | −3.54e-4 | 7.08e-5 |
| 512 | 2.24e-5 | −1.59e-4 | 2.12e-5 | 3.00e-5 | −2.51e-4 | 2.19e-5 | 9.21e-6 | −7.87e-5 | 1.51e-5 |
| | | | | $k=1000$ | | | | | |
| 8 | 2.70e-3 | −1.03e-2 | 2.68e-3 | 2.61e-3 | −1.11e-2 | 1.81e-3 | 4.64e-4 | −5.50e-3 | 8.92e-4 |
| 32 | 3.92e-4 | −2.99e-3 | 3.82e-4 | 4.57e-4 | −3.19e-3 | 3.28e-4 | 9.72e-5 | −1.38e-3 | 1.85e-4 |
| 128 | 6.39e-5 | −7.49e-4 | 6.01e-5 | 8.41e-5 | −9.00e-4 | 6.19e-5 | 2.21e-5 | −3.32e-4 | 3.92e-5 |
| 512 | 1.24e-5 | −1.58e-4 | 1.11e-5 | 1.61e-5 | −2.65e-4 | 1.19e-5 | 5.36e-6 | −7.33e-5 | 8.67e-6 |
| | | | | $k=2000$ | | | | | |
| 8 | 1.56e-3 | −1.03e-2 | 1.53e-3 | 1.61e-3 | −1.13e-2 | 1.13e-3 | 2.79e-4 | −5.41e-3 | 5.18e-4 |
| 32 | 2.45e-4 | −2.99e-3 | 2.35e-4 | 3.26e-4 | −3.26e-3 | 2.38e-4 | 5.98e-5 | −1.39e-3 | 1.09e-4 |
| 128 | 4.62e-5 | −7.45e-4 | 4.28e-5 | 6.34e-5 | −9.22e-4 | 4.79e-5 | 1.33e-5 | −3.42e-4 | 2.32e-5 |
| 512 | 8.83e-6 | −1.55e-4 | 7.71e-6 | 1.22e-5 | −2.76e-4 | 9.33e-6 | 3.19e-6 | −7.46e-5 | 5.00e-6 |

Wang (2002) for the univariate case. With sufficiently high frequency and sample size, however, the bias may be considered negligible in most practical situations.

Finally, to obtain an idea of the importance of the approximation bias relative to the finite sample bias, we simulate a discrete time cDCC model with the parameters of the diffusion approximated to first order by Theorem 1, using $h = 1/512$ and sample sizes $k = 250, 500, 1000, 2000$. The estimation by MLE of this model only involves finite sample bias and variance, but no approximation bias, so that the comparison with the diffusion approximation allows us to draw conclusions about the approximation bias.[4] The results are summarized in Table 3. We see that for the persistence and drift parameters, the relative bias is about 80% for the smallest sample size, meaning that about 20% of the overall bias is explained by the approximation. For the parameters linked to the innovation terms, the relative bias is much smaller and close to zero, since the overall bias is dominated by the approximation bias even for small sample sizes. All relative biases tend to

**TABLE 2.** QAML variances of cDCC diffusion approximation at varying frequencies $1/h$ and number of time periods $k$

| $1/h$ | $\hat{c}_{1,h}$ | $\hat{A}_{11,h}$ | $\hat{\Lambda}_{11,h}$ | $\hat{c}_{2,h}$ | $\hat{A}_{22,h}$ | $\hat{\Lambda}_{22,h}$ | $\hat{\bar{Q}}_{12,h}$ | $\hat{\vartheta}_h$ | $\hat{\phi}_h$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $k=250$ | | | | |
| 8 | 4.41e-4 | 4.87e-5 | 4.48e-4 | 2.90e-4 | 5.55e-5 | 1.43e-4 | 1.49e-5 | 3.65e-5 | 5.66e-5 |
| 32 | 7.58e-6 | 4.33e-6 | 7.87e-6 | 5.42e-6 | 5.05e-6 | 2.82e-6 | 2.94e-7 | 3.61e-6 | 1.10e-6 |
| 128 | 1.64e-7 | 4.23e-7 | 1.72e-7 | 1.84e-7 | 5.53e-7 | 1.00e-7 | 1.28e-8 | 4.06e-7 | 4.84e-8 |
| 512 | 6.42e-9 | 4.78e-8 | 6.89e-9 | 8.72e-9 | 6.36e-8 | 4.87e-9 | 6.7e-10 | 4.78e-8 | 2.54e-9 |
| | | | | | $k=500$ | | | | |
| 8 | 2.09e-4 | 2.53e-5 | 2.14e-4 | 9.92e-5 | 2.64e-5 | 4.91e-5 | 3.13e-6 | 1.77e-5 | 1.16e-5 |
| 32 | 2.28e-6 | 2.06e-6 | 2.36e-6 | 2.06e-6 | 2.38e-6 | 1.06e-6 | 1.02e-7 | 1.81e-6 | 3.67e-7 |
| 128 | 6.06e-8 | 2.06e-7 | 6.42e-8 | 7.75e-8 | 2.61e-7 | 4.23e-8 | 4.75e-9 | 2.05e-7 | 1.72e-8 |
| 512 | 2.55e-9 | 2.42e-8 | 2.74e-9 | 3.75e-9 | 3.17e-8 | 2.12e-9 | 2.6e-10 | 2.38e-8 | 9.6e-10 |
| | | | | | $k=1000$ | | | | |
| 8 | 6.71e-5 | 1.12e-5 | 6.85e-5 | 3.58e-5 | 1.23e-5 | 1.72e-5 | 1.00e-6 | 8.28e-6 | 3.67e-6 |
| 32 | 9.22e-7 | 1.01e-6 | 9.45e-7 | 8.87e-7 | 1.20e-6 | 4.50e-7 | 3.84e-8 | 8.82e-7 | 1.37e-7 |
| 128 | 2.74e-8 | 1.02e-7 | 2.89e-8 | 3.42e-8 | 1.32e-7 | 1.85e-8 | 1.89e-9 | 1.01e-7 | 6.72e-9 |
| 512 | 1.20e-9 | 1.21e-8 | 1.29e-9 | 1.72e-9 | 1.53e-8 | 9.7e-10 | 1.0e-10 | 1.18e-8 | 3.8e-10 |
| | | | | | $k=2000$ | | | | |
| 8 | 2.56e-5 | 5.41e-6 | 2.58e-5 | 1.50e-5 | 6.01e-6 | 7.19e-6 | 4.34e-7 | 3.93e-6 | 1.54e-6 |
| 32 | 4.19e-7 | 4.99e-7 | 4.29e-7 | 4.24e-7 | 5.76e-7 | 2.15e-7 | 1.76e-8 | 4.18e-7 | 6.03e-8 |
| 128 | 1.33e-8 | 5.18e-8 | 1.40e-8 | 1.68e-8 | 6.42e-8 | 9.08e-9 | 8.6e-10 | 4.86e-8 | 2.94e-9 |
| 512 | 6.0e-10 | 5.89e-9 | 6.4e-10 | 8.7e-10 | 7.69e-9 | 4.8e-10 | 4.8e-11 | 6.03e-9 | 1.6e-10 |

zero as $k$ increases because, unlike the approximation bias, the finite sample bias disappears. The relative variances in the bottom half of the table give the relative efficiency of MLE with respect to QAML. They are for all parameters in the range of 70 to 86 percent at the highest sample size.

## 4. CONCLUSIONS

This paper considered weak diffusion limits of two conditional correlation GARCH specifications, namely the cDCC model of Aielli (2006) and the CCC model of Bollerslev (1990). For the cDCC-GARCH model, the diffusion limit is degenerate in the sense that the diffusion of the variances and that of the diagonal elements of the correlation driving process are pairwise governed by the same Brownian motion. We show that this result is due to the particular structure

**TABLE 3.** MLE bias and variance of discrete time cDCC models simulated at time interval $h = 1/512$, divided by corresponding QAML bias and variances of continuous time approximation, in percent

| $k$ | $\hat{c}_{1,h}$ | $\hat{A}_{11,h}$ | $\hat{\Lambda}_{11,h}$ | $\hat{c}_{2,h}$ | $\hat{A}_{22,h}$ | $\hat{\Lambda}_{22,h}$ | $\hat{\bar{Q}}_{12,h}$ | $\hat{\vartheta}_h$ | $\hat{\phi}_h$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Relative bias | | | | | |
| 250 | 82.7 | 2.4 | 86.4 | 79.4 | 2.3 | 79.8 | 74.3 | 11.6 | 86.7 |
| 500 | 77.3 | 0.6 | 84.0 | 74.7 | 0.6 | 74.5 | 71.3 | 6.1 | 84.1 |
| 1000 | 65.5 | 0.5 | 73.9 | 59.7 | 0.3 | 59.8 | 63.3 | 2.5 | 78.6 |
| 2000 | 39.9 | 0.5 | 45.7 | 39.6 | 0.2 | 37.7 | 45.9 | 0.3 | 57.5 |
| | | | | Relative variance | | | | | |
| 250 | 99.0 | 95.1 | 98.1 | 91.5 | 88.5 | 89.5 | 83.0 | 79.2 | 80.1 |
| 500 | 100.9 | 68.8 | 102.2 | 103.7 | 83.8 | 100.3 | 87.5 | 66.7 | 83.8 |
| 1000 | 96.9 | 93.2 | 96.4 | 90.3 | 90.1 | 88.9 | 106.4 | 68.1 | 103.9 |
| 2000 | 86.3 | 83.6 | 84.8 | 75.5 | 81.1 | 73.0 | 81.6 | 72.5 | 84.7 |

of the noise propagation system of the variances and of the correlation driving process. The CCC model, which can be obtained from cDCC under suitable parameter restrictions, admits a nondegenerate diffusion. Under an alternative set of conditions regarding the convergence rates of the parameters, we obtain diffusion limits characterized by a stochastic price process where either the variances, the correlations, or both, are time-varying but deterministic. Our Monte Carlo study confirms that estimation of the diffusion parameters by QAML is inconsistent for any fixed frequency, but may provide good approximations if the frequency and sample size are sufficiently large.

There are several possible extensions of this work. First of all, the assumption of Gaussian innovations may be relaxed. One may also extend the results to allow for volatility spillover. Furthermore, similar to Nelson (1990) it may be possible to derive the stationary distribution of the continuous time limit of returns, variances and correlations. Also, different GARCH specifications for the dynamics of the variances of the marginal processes could be used, which would potentially solve the redundancy problem in the sense of Theorem 1. Finally, it would be useful to extend the results of this paper to jump-diffusion processes, based on the results of Ethier and Kurtz (1986).

*NOTES*

1. Examples for $N = 3$: $I^+ = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix}'$ and $I^- = \begin{bmatrix} 0\,1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0 \end{bmatrix}$.

2. The same transformation can be carried out for the intercept of the $V^{(h)}_{(k+1)h}$ process, i.e., $c_h = (I_N - A_h - B_h)\bar{c}$. The vector $\bar{c}$ is frequency invariant and contains the (rescaled) unconditional

variances of the return process $\left(Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)}\right)$, i.e., $\bar{c} = \mathrm{E}\left[\left(Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)}\right) \odot \left(Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)}\right)\right]/h = \mathrm{E}\left[V_{(k+1)h}^{(h)}\right]$, for all $h > 0$.

3. In the univariate setting, the cases $\Lambda = 0$ (integrated variance) and $\Lambda < 0$ (strictly stationary but not covariance stationary GARCH process) are also discussed in Nelson (1990). In this paper we restrict the analysis to the covariance stationary case.

4. We thank the editor and an anonymous referee for making this suggestion.

## REFERENCES

Aielli, G.P. (2006) Consistent estimation of large scale dynamic conditional correlations. Department of Statistics, University of Florence.

Aielli, G.P. (2013) Dynamic conditional correlations: On properties and estimation. *Journal of Business and Economic Statistics* 31, 282–299.

Alexander, C. & E. Lazar (2005) On the continuous limit of GARCH. Discussion Papers in Finance 2005–13. ICMA Centre Discussion Paper No. DP2005-13. University of Reading, UK.

Barone-Adesi, G., H. Rasmussen, & C. Ravanelli (2005) An option pricing formula for the GARCH diffusion model. *Computational Statistics and Data Analysis* 49(2), 287–310.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T. (1990) Modeling the coherence on short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498–505.

Bollerslev, T., R. Engle, & D. Nelson (1994) Arch Models. *Handbook of Econometrics*, vol. IV. North-Holland.

Broze, L., O. Scaillet, & J. Zakoian (1998) Quasi indirect inference for diffusion processes. *Econometric Theory* 14, 161–186.

Corradi, V. (2000) Reconsidering the continuous time limit of the GARCH(1,1) process, *Journal of Econometrics* 96, 145–153.

Ding, Z., C. Granger, & R. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Drost, C. & T. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.

Drost, F. & J. Werker (1996) Closing the gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.

Duan, J. (1997) Augmented *p,q* process and its diffusion limit. *Journal of Econometrics* 79, 97–127.

Duffie, D. & K. Singleton (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica* 61, 929–952.

Engle, R. (2002) Dynamic conditional correlation - A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.

Engle, R. & G. Lee (1996) Estimating diffusion models of stochastic volatility. In P. Rossi (ed.), *Modeling Stock Market Volatility: Bridging the Gap to Continuous Time*, pp. 333–384. Academic Press.

Ethier, S. & T. Kurtz (1986) *Markov Processes: Characterization and Convergence*. Wiley.

Fornari, F. & A. Mele (1997) Weak convergence and distributional assumptions for a general class of nonlinear ARCH models. *Econometric Reviews* 16, 205–229.

Fornari, F. & A. Mele (2006) Approximating volatility diffusions with CEV-ARCH models. *Journal of Economic Dynamics & Control* 30(6), 931–966.

Ghysels, E., A. Harvey, & E. Renault (1996) Stochastic volatility. In Maddala, G. (ed.), *Handbook of Statistics*, vol. 14, pp. 119–191. North-Holland.

Hafner, C. (2003) Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics* 1, 26–54.

Hafner, C. (2008) Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics* 142, 467–483.

Hafner, C. S. Laurent, & F. Violante (2016) Weak diffusion limits of dynamic conditional correlation models, CORE discussion paper 2016/09, Université catholique de Louvain, Belgium.

Hafner, C. & J. Rombouts (2007) Estimation of temporally aggregated multivariate GARCH models. *Journal of Statistical Computation and Simulation* 77, 629–650.

Hobson, D., & L. Rogers (1998) Complete models with stochastic volatility. *Mathematical Finance* 8, 27–48.

Jeantheau, T. (2004) A link between complete models with stochastic volatility and ARCH models. *Finance and Stochastics* 8, 111–131.

Jones, C. (2003) The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics* 116, 181–224.

Kleppe, T., J. Yu, & H. Skaug (2014) Maximum likelihood estimation of partially observed diffusion models. *Journal of Econometrics* 180, 73–80.

Kushner, H. (1984) *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*. The MIT Press.

Lewis, A. (2000) *Option Valuation under Stochastic Volatility*. Finance Press.

Li, C. (2013) Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Annals of Statistics* 41, 1350–1380.

Lütkepohl, H. (1996) *Handbook of Matrices*. Wiley.

Magnus, J. (1988) *Linear Structures*. Griffin.

Nelson, D. (1990) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.

Nelson, D. (1996) Asymptotic filtering theory for multivariate ARCH models. *Journal of Econometrics* 71, 1–47.

Nelson, D. & D. Foster (1994) Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62, 1–41.

Phillips, P. & J. Yu (2009) Maximum likelihood and Gaussian estimation of continuous time models in finance. In T.G. Anderson et al. (eds.), *Handbook of Financial Time Series*, pp. 497–530. Springer.

Steele, J. (2001) *Stochastic Calculus and Financial Applications*. Springer.

Stentoft, L. (2011) American option pricing with discrete and continuous time models: An empirical comparison. *Journal of Empirical Finance* 18, 880–902.

Stroock, D. & S. Varadhan (1979) *Multidimensional Diffusion Processes*. Springer-Verlag.

Trifi, A. (2006) Issues of aggregation over time of conditional heteroscedastic volatility models: What kind of diffusion do we recover? *Studies in Nonlinear Dynamics & Econometrics* 10(4), 1314–1323.

Wang, Y. (2002) Asymptotic nonequivalence of GARCH models and diffusions. *Annals of Statistics* 30, 754–783.

# APPENDIX A: Weak convergence of stochastic systems

Based on the work of Stroock and Varadhan (1979), Kushner (1984), Ethier and Kurtz (1986) and Nelson (1990), we introduce a set of conditions for the weak convergence of a system of discrete time stochastic difference equations towards a system of stochastic differential equations.

Let $Pr_h$ be a fixed probability measure for each $h > 0$. Let $\mathcal{F}_{kh}$ be the $\sigma$-field generated by $(kh, X_0^{(h)}, X_h^{(h)}, X_{2h}^{(h)}, \ldots, X_{(k-1)h}^{(h)})$, where $X_{kh}^{(h)}$ is an $N$-dimensional discrete time Markov chain indexed by $h > 0$, $k \in \mathbb{N}$, with $\nu_h$ a probability measure on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, where $\mathcal{B}(\mathbb{R}^N)$ are the Borel sets on $\mathbb{R}^N$, such that $Pr_h[X_0^{(h)} \in \Gamma] = \nu_h(\Gamma)$ for any $\Gamma \in \mathcal{B}(\mathbb{R}^N)$ defines the distribution of the starting point $X_0^{(h)}$, and with transition probabilities

$Pr_h\big[X_{kh}^{(h)} \in \Gamma \big| \mathcal{F}_{kh}\big] = \Pi_{h,kh}\big(X_{(k-1)h}^{(h)}, \Gamma\big), \forall\, k \in \mathbb{N}, \Gamma \in \mathcal{B}\big(\mathbb{R}^N\big)$. Let us now define $X_t^{(h)}$ a continuous time process, formed from the discrete time process $X_{kh}^{(h)}$ as a càdlàg step function with jumps at $h, 2h, 3h, \ldots$, such that $Pr_h\big[X_t^{(h)} = X_{kh}^{(h)}, kh < t < (k+1)h\big] = 1$. Finally, let $X_t$ be a continuous time process obtained from $X_t^{(h)}$ by shrinking $h$ towards zero. $X_t$ represents the limiting diffusion process to which, under Assumptions A.1 to A.4 given below, the discrete time process $X_t^{(h)}$ weakly converges as $h \to 0$.

For the convergence results, we need the following assumptions.

**Assumption A.1.** There exist a continuous, measurable function $a(x,t) : \mathbb{R}^N \times [0,\infty) \to \Omega'$ and a continuous, measurable function $b(x,t) : \mathbb{R}^N \times [0,\infty) \to \mathbb{R}^N$ such that for all $r > 0$ and $(k-1)h < t < kh$

$a)\quad \lim_{h\to 0}\sup_{\|x\|\leqslant r}\left\| h^{-1}\mathrm{E}\Big[ X_{(k+1)h}^{(h)} - X_{kh}^{(h)} \Big| X_{kh}^{(h)} = x \Big] - b(x,t)\right\| = 0, \qquad (A.1)$

$b)\quad \lim_{h\to 0}\sup_{\|x\|\leqslant r}\left\| h^{-1}\mathrm{E}\Big[ \big(X_{(k+1)h}^{(h)} - X_{kh}^{(h)}\big)\big(X_{(k+1)h}^{(h)} - X_{kh}^{(h)}\big)' \Big| X_{kh}^{(h)} = x \Big] - a(x,t)\right\| = 0,$
$$(A.2)$$

$c)\quad \exists\, \delta > 0 : \lim_{h\to 0}\sup_{\|x\|\leqslant r}\left\| h^{-1}\mathrm{E}\Big[ \big|(X_{(k+1)h}^{(h)} - X_{kh}^{(h)})_i\big|^{2+\delta} \Big| X_{kh}^{(h)} = x \Big]\right\| = 0,$ where $(.)_i$ is

the $i^{th}$ element of the vector $\big(X_{(k+1)h}^{(h)} - X_{kh}^{(h)}\big)$.

**Assumption A.2.** There exists a continuous function $\sigma(x,t) : \mathbb{R}^N \times [0,\infty) \to \Omega$ such that for all $x \in \mathbb{R}^N$ and $t \geq 0$, $a(x,t) = \sigma(x,t)\sigma(x,t)'$.

**Assumption A.3.** $X_0^{(h)}$ converges in distribution, as $h \to 0$, to a random variable $X_0$ with probability measure $\nu_0$ on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.

**Assumption A.4.** $\nu_0, b(x,t), a(x,t)$ uniquely specify the distribution of a diffusion process $X_t$ with initial distribution $\nu_0$, drift vector $b(x,t)$ and diffusion matrix $a(x,t)$.

We can now state the following theorem for the weak convergence of discrete time stochastic sequences.

**Theorem of weak convergence (Nelson, 1990).** *Under Assumptions A.1 to A.4, the sequence of discrete time process $X_{kh}^{(h)}$ indexed by $h > 0$, $k \in \mathbb{N}$, converges in distribution, as $h \to 0$, to the diffusion process $X_t$, i.e., the solution of the system of stochastic differential equations*

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t, \qquad (A.3)$$

*where $W_t$ is an N-dimensional vector of mutually independent standard Brownian motions, independent from $X_0$, and with initial distribution $\nu_0$. The process $X_t$ exists, it is finite in finite time intervals almost surely, it is distributionally unique and its distribution does not depend on the choice of $\sigma(x,t)$.*

For the proof, we refer to Nelson (1990). Conditions under which $\nu_0$, $b(x,t)$, and $a(x,t)$ ensure finiteness of the process in finite time intervals and uniqueness of the limiting diffusion are extensively discussed in Stroock and Varadhan (1979), Ethier and Kurtz (1986),

and Nelson (1990). To ensure weak existence, uniqueness, and nonexplosion of the diffusion process $X_t$ on compact sets we rely on 'Condition A' of Nelson (1990), i.e.,

**Condition A.1** (weak existence and uniqueness). *Let $a(x,t)$ and $b(x,t)$ be continuous in both $x$ and $t$ with two partial derivatives with respect to $x$.*

Following Theorem 10.2.2 of Stroock and Varadhan (1979), we impose the following conditions of nonexplosiveness of $X_t$.

**Condition A.2** (nonexplosiveness). *For each $T > 0$, there is a $C_T < \infty$ such that*

$$\sup_{0 \leq t \leq T} \|a(x,t)\| \leq C_T \left(1 + |x|^2\right), \quad x \in \mathbb{R}^N$$

*and*

$$\sup_{0 \leq t \leq T} \langle x, b(x,t) \rangle \leq C_T \left(1 + |x|^2\right), \quad x \in \mathbb{R}^N.$$

These are not the weakest possible conditions, but they are easy to check and will suffice in our model framework.

# APPENDIX B: Proofs

**Proof of Theorem 1.** The process (1)–(5) is Markovian with drift and second moment per unit of time given by (B.1)–(B.3) (drift) and (B.4), (B.6), and (B.8)–(B.11) (second moments), respectively. The theorem of weak convergence applies if Assumptions A.1 to A.4 hold. Assumption 3 has been assumed directly, so that it remains to verify Assumptions A.1, A.2, and A.4.

We first show that the parameter convergence conditions (B.12)–(B.17) satisfy the requirements of Assumption A.1. The first step is to compute the increments of the process (1)–(3), that is

$$Y_{kh}^{(h)} - Y_{(k-1)h}^{(h)} = S_{kh}^{(h)} \eta_{kh}^{(h)},$$

$$V_{(k+1)h}^{(h)} - V_{kh}^{(h)} = c_h + A_h S_{kh}^{(h)2} h^{-1} \left(\eta_{kh}^{(h)} \odot \eta_{kh}^{(h)}\right) + (B_h - I_N) V_{kh}^{(h)},$$

$$q_{(k+1)h}^{(h)} - q_{kh}^{(h)} = (1 - \vartheta_h - \gamma_h)\bar{q} + \vartheta_h h^{-1} \text{vech}\left(P_{kh}^{(h)} \eta_{kh}^{(h)} \eta_{kh}^{(h)\prime} P_{kh}^{(h)}\right) + (\gamma_h - 1)q_{kh}^{(h)},$$

where we have used that $\bar{q}_h = \bar{q}(1 - \vartheta_h - \gamma_h)$.

Second, we compute the conditional moments to define suitable convergence conditions as required by Assumption A.1. To simplify the notation, let us define the difference operator over an interval of size $h$ as $\Delta : \Delta X_{kh}^{(h)} = X_{kh}^{(h)} - X_{(k-1)h}^{(h)}$. The first conditional moment per unit of time of the increments of (1)–(3) is given by

$$h^{-1}\text{E}\left[\Delta Y_{kh}^{(h)}|\mathcal{F}_{kh}\right] = S_{kh}^{(h)} \text{E}\left[\eta_{kh}^{(h)}|\mathcal{F}_{kh}\right] = 0, \tag{B.1}$$

$$h^{-1}\text{E}\left[\Delta V_{(k+1)h}^{(h)}|\mathcal{F}_{kh}\right] = h^{-1}c_h + A_h h^{-2} S_{kh}^{(h)2} \text{E}\left[\eta_{kh}^{(h)} \odot \eta_{kh}^{(h)}|\mathcal{F}_{kh}\right] + h^{-1}(B_h - I_N)V_{kh}^{(h)}$$

$$= h^{-1}c_h + h^{-1}(A_h + B_h - I_N)V_{kh}^{(h)}, \tag{B.2}$$

$$h^{-1}\mathrm{E}\left[\Delta q_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right] = h^{-1}(1-\vartheta_h-\gamma_h)\bar{q} + h^{-2}\vartheta_h\,\mathrm{vech}\left(\mathrm{E}\left[P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\big|\mathcal{F}_{kh}\right]\right)$$
$$+ h^{-1}(\gamma_h-1)q_{kh}^{(h)}$$
$$= h^{-1}(1-\vartheta_h-\gamma_h)\bar{q} + h^{-1}(\vartheta_h+\gamma_h-1)q_{kh}^{(h)}, \tag{B.3}$$

where $\quad \mathrm{E}\left[\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\big|\mathcal{F}_{kh}\right] = h\,1_N \quad$ and $\quad \mathrm{vech}\left(P_{kh}^{(h)}\mathrm{E}[\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}|\mathcal{F}_{kh}]P_{kh}^{(h)}\right) =$
$h\,\mathrm{vech}\left(P_{kh}^{(h)}R_{kh}^{(h)}P_{kh}^{(h)}\right) = hq_{kh}^{(h)}.$

To compute the second moments per unit of time, consider the following partition

$$\mathrm{vech}\left(\mathrm{Var}\left(\left[\Delta Y_{kh}^{(h)\prime}, \Delta V_{(k+1)h}^{(h)\prime}, \Delta q_{(k+1)h}^{(h)\prime}\right]'\big|\mathcal{F}_{kh}\right)\right) = \begin{bmatrix} \mathrm{Var}\left(\Delta Y_{kh}^{(h)}\big|\mathcal{F}_{kh}\right) \\ \mathrm{Cov}\left(\Delta Y_{kh}^{(h)}, \Delta V_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right)' \\ \mathrm{Cov}\left(\Delta Y_{kh}^{(h)}, \Delta q_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right)' \\ \mathrm{Var}\left(\Delta V_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right) \\ \mathrm{Cov}\left(\Delta V_{(k+1)h}^{(h)}, \Delta q_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right)' \\ \mathrm{Var}\left(\Delta q_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right) \end{bmatrix}.$$

The conditional variance of $\Delta Y_{kh}^{(h)}$ standardized by $h$ is given by

$$h^{-1}\mathrm{Var}\left[\Delta Y_{kh}^{(h)}\big|\mathcal{F}_{kh}\right] = h^{-1}S_{kh}^{(h)}\mathrm{E}\left(\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}\big|\mathcal{F}_{kh}\right)S_{kh}^{(h)} = S_{kh}^{(h)}R_{kh}^{(h)}S_{kh}^{(h)}. \tag{B.4}$$

Similarly the conditional variance of $\Delta V_{(k+1)h}^{(h)}$ is given by

$$h^{-1}\mathrm{Var}\left[\Delta V_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right] = A_h S_{kh}^{(h)2}h^{-3}\left[\mathrm{E}\left[\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)'\big|\mathcal{F}_{kh}\right]\right.$$
$$\left. -\mathrm{E}\left[\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)\big|\mathcal{F}_{kh}\right]\mathrm{E}\left[\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)\big|\mathcal{F}_{kh}\right]'\right]S_{kh}^{(h)2}A_h'. \tag{B.5}$$
$$= 2h^{-1}A_h S_{kh}^{(h)2}\left(R_{kh}^{(h)}\odot R_{kh}^{(h)}\right)S_{kh}^{(h)2}A_h', \tag{B.6}$$

where the second equality uses that, under the conditional normality assumption, $\mathrm{E}[\eta_{kh,i}^{(h)}\eta_{kh,j}^{(h)}|\mathcal{F}_{kh}] = h^2(1+2R_{kh,ij}^{(h)2})$, $i,j \in \{1,\ldots,N\}$. Moreover, $\mathrm{E}[(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)})|\mathcal{F}_{kh}] = h1_N$.

The variance of $\Delta q_{(k+1)h}^{(h)}$ is given by

$$h^{-1}\mathrm{Var}\left[\Delta q_{(k+1)h}^{(h)}\big|\mathcal{F}_{kh}\right]$$
$$= \vartheta_h^2 h^{-3}\mathrm{E}\left[\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)'\big|\mathcal{F}_{kh}\right]$$
$$- \vartheta_h^2 h^{-3}\mathrm{E}\left[\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)\big|\mathcal{F}_{kh}\right]\mathrm{E}\left[\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)\big|\mathcal{F}_{kh}\right]' \tag{B.7}$$
$$= h^{-1}\vartheta_h^2\left[D_N^+\left(P_{kh}^{(h)}\otimes P_{kh}^{(h)}\right)D_N^{+\prime}K_N^{(h)}D_N^+\left(P_{kh}^{(h)}\otimes P_{kh}^{(h)}\right)D_N^{+\prime} - q_{kh}^{(h)}q_{kh}^{(h)\prime}\right], \tag{B.8}$$

where the second equality uses that $\mathrm{E}\big[\mathrm{vech}\big(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\big)|\mathcal{F}_{kh}\big]=hq_{kh}^{(h)}$, and where $K_{kh}^{(h)}=h^{-2}\mathrm{E}[\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}\otimes\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}|\mathcal{F}_{kh}]$ is the $(N(N+1)/2\times N(N+1)/2)$ matrix of conditional fourth moments of $\eta_{kh}^{(h)}$ which, given the normality assumption of the innovations, is given by

$$K_{kh}^{(h)}=2D_N^+\left(R_{kh}^{(h)}\otimes R_{kh}^{(h)}\right)D_N^{+\prime}+\mathrm{vech}\left(R_{kh}^{(h)}\right)\mathrm{vech}\left(R_{kh}^{(h)}\right)',$$

which is a consequence of Theorem 10.2 of Magnus (1988), see the proof of Theorem 1 of Hafner (2003).

Finally, the conditional covariances are

$$h^{-1}\mathrm{Cov}\left[\Delta Y_{kh}^{(h)},\Delta V_{(k+1)h}^{(h)}\right)|\mathcal{F}_{kh}\right]=h^{-1}\mathrm{E}\left[\left(S_{kh}^{(h)}\eta_{kh}^{(h)}\right)\left(A_hS_{kh}^{(h)2}h^{-1}\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)\right)'|\mathcal{F}_{kh}\right]$$

$$=h^{-2}S_{kh}^{(h)}\mathrm{E}\left[\eta_{kh}^{(h)}\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)'|\mathcal{F}_{kh}\right]S_{kh}^{(h)2}A_h=0,\ \textbf{(B.9)}$$

because all conditional third moments of $\eta_{kh}^{(h)}$ are equal to zero given the normality assumption. Furthermore, we have

$$h^{-1}\mathrm{Cov}\left[\Delta Y_{kh}^{(h)},\Delta q_{(k+1)h}^{(h)}|\mathcal{F}_{kh}\right]$$

$$=h^{-2}\mathrm{E}\left[\left(S_{kh}^{(h)}\eta_{kh}^{(h)}\right)\left(\vartheta_h\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)\right)'|\mathcal{F}_{kh}\right]$$

$$=h^{-2}\vartheta_hS_{kh}^{(h)}\mathrm{E}\left[\eta_{kh}^{(h)}\mathrm{vech}\left(\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}\right)'|\mathcal{F}_{kh}\right]\left(D_N^+\left(P_{kh}^{(h)}\otimes P_{kh}^{(h)}\right)D_N^{+\prime}\right)=0\quad\textbf{(B.10)}$$

and

$$h^{-1}\mathrm{Cov}\left[\Delta V_{(k+1)h}^{(h)},\Delta q_{(k+1)h}^{(h)}|\mathcal{F}_{kh}\right]=$$

$$=h^{-3}\mathrm{E}\left[\left(A_hS_{kh}^{(h)2}\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)\right)\left(\vartheta_h\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)\right)'|\mathcal{F}_{kh}\right]$$

$$-h^{-3}\mathrm{E}\left[A_hS_{kh}^{(h)2}\left(\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}\right)|\mathcal{F}_{kh}\right]\mathrm{E}\left[\vartheta_h\mathrm{vech}\left(P_{kh}^{(h)}\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}P_{kh}^{(h)}\right)|\mathcal{F}_{kh}\right]'$$

$$=h^{-1}\vartheta_hA_hS_{kh}^{(h)2}\left[I^*K_{kh}^{(h)}D_N^+\left(P_{kh}^{(h)}\otimes P_{kh}^{(h)}\right)D_N^{+\prime}-1_Nq_{kh}^{(h)\prime}\right],\qquad\textbf{(B.11)}$$

where the second equality uses $\eta_{kh}^{(h)}\odot\eta_{kh}^{(h)}=\mathrm{diag}\big(\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}\big)=I^*\mathrm{vech}\big(\eta_{kh}^{(h)}\eta_{kh}^{(h)\prime}\big)$.

For the conditional moments (B.1)–(B.3) (drift) and (B.4), (B.6), and (B.8)–(B.11) (second moments) to converge to well behaved functions as $h\to0$, as required by Assumption 1 $a$) and $b$), the following limits must exist and be finite

$$\lim_{h\to0}h^{-1}c_h=c,\tag{B.12}$$

$$\lim_{h\to0}h^{-1}(I_N-A_h-B_h)=\Lambda,\tag{B.13}$$

$$\lim_{h\to0}h^{-1/2}A_h=A,\tag{B.14}$$

$$\lim_{h\to0}h^{-1}\bar{Q}_h=\bar{Q}\phi,\tag{B.15}$$

$$\lim_{h \to 0} h^{-1}(1 - \vartheta_h - \gamma_h) = \phi, \tag{B.16}$$

$$\lim_{h \to 0} h^{-1/2}\vartheta_h = \vartheta, \tag{B.17}$$

where $c$ is a $(N \times 1)$ vector, $A$ and $\Lambda$ are $(N \times N)$ diagonal matrices and $\phi$ and $\vartheta$ are scalars with all elements positive and finite, such that $c > 0$ (elementwise) ensures positivity of the variance process, $A > 0$ and $\vartheta > 0$ ensure that the rescaled second conditional moment of $V_{kh}^{(h)}$ and $q_{kh}^{(h)}$ does not vanish as $h \to 0$, while $\Lambda > 0$ and $\phi > 0$ ensure covariance stationarity of the return process. Finally, by straightforward computation as in Nelson (1990), under (B.12)–(B.17), Assumption A.1 $c$) holds for $\delta = 2$, i.e.,

$$h^{-1}\lim_{h \to 0} E\left[\left|\left(\Delta Y_{kh}^{(h)}\right)_i\right|^4 \bigg| \mathcal{F}_{kh}\right] = 0, \ i,i = 1,\dots,N,$$

$$h^{-1}\lim_{h \to 0} E\left[\left|\left(\Delta V_{(k+1)h}^{(h)}\right)_i\right|^4 \bigg| \mathcal{F}_{kh}\right] = 0, \ i,i = 1,\dots,N,$$

$$h^{-1}\lim_{h \to 0} E\left[\left|\left(\Delta q_{(k+1)h}^{(h)}\right)_i\right|^4 \bigg| \mathcal{F}_{kh}\right] = 0, \ i,i = 1,\dots,N(N+1)/2,$$

which shows that Assumption A.1 holds under our convergence conditions (6) to (11), and the drift and diffusion matrix for the system of stochastic differential equations are defined, which completes the first part of the proof.

Substituting (B.12)–(B.17) into (B.1)–(B.3) (first moments) and (B.4), (B.6), and (B.8)–(B.11) (second moments), we obtain

$$h^{-1}\mathrm{E}\left[\Delta Y_{kh}^{(h)} | \mathcal{F}_{kh}\right] = 0$$

$$h^{-1}\mathrm{E}\left[\Delta V_{(k+1)h}^{(h)} | \mathcal{F}_{kh}\right] = c - \Lambda V_{kh}^{(h)} + o(1)$$

$$h^{-1}\mathrm{E}\left[\Delta q_{(k+1)h}^{(h)} | \mathcal{F}_{kh}\right] = \phi\left(\bar{q} + q_{kh}^{(h)}\right) + o(1)$$

for the drift, while for the moments of second order we have

$$h^{-1}\mathrm{Var}\left[\Delta Y_{kh}^{(h)} | \mathcal{F}_{kh}\right] = S_{kh}^{(h)} R_{kh}^{(h)} S_{kh}^{(h)},$$

$$h^{-1}\mathrm{Var}\left[\Delta V_{(k+1)h}^{(h)} | \mathcal{F}_{kh}\right] = 2A_h S_{kh}^{(h)2}\left(R_{kh}^{(h)} \odot R_{kh}^{(h)}\right)S_{kh}^{(h)2} A_h + o(1),$$

$$h^{-1}\mathrm{Var}\left[\Delta q_{(k+1)h}^{(h)} | \mathcal{F}_{kh}\right]$$
$$= \vartheta^2\left[\left(D_N^+\left(P_{kh}^{(h)} \otimes P_{kh}^{(h)}\right)D_N^{+\prime}\right)K_{kh}^{(h)}\left(D_N^+\left(P_{kh}^{(h)} \otimes P_{kh}^{(h)}\right)D_N^{+\prime}\right) - q_{kh}^{(h)}q_{kh}^{(h)\prime}\right] + o(1),$$

$$h^{-1}\mathrm{Cov}\left[\Delta Y_{kh}^{(h)}, \Delta V_{(k+1)h}^{(h)}\right] | \mathcal{F}_{kh}\right) = 0,$$

$$h^{-1}\mathrm{Cov}\left[\Delta Y_{kh}^{(h)}, \Delta q_{(k+1)h}^{(h)}\right] | \mathcal{F}_{kh}\right) = 0,$$

$$h^{-1}\mathrm{Cov}\left[\Delta V_{(k+1)h}^{(h)}, \Delta q_{(k+1)h}^{(h)}\right] | \mathcal{F}_{kh}\right)$$
$$= \vartheta A S_{kh}^{(h)2}\left[I^* K_{kh}^{(h)} D_N^+\left(P_{kh}^{(h)} \otimes P_{kh}^{(h)}\right)D_N^{+\prime} - 1_N q_{kh}^{(h)\prime}\right] + o(1).$$

Hence, as $h \to 0$, the functions (13) and (14) are solutions of (A.1) and (A.2) and represent the drift and the diffusion matrix of the diffusion process $X_t = [Y_t', V_t', q_t']'$.

From the representation (16), we have that the columns $N+1, \ldots, 2N$ and $2N+1, \ldots, 3N$ are collinear. Thus, the diffusion matrix is singular with $\mathrm{rank}(a([Y_t', V_t', q_t']')) = N(N+3)/2 < \dim(X_t)) = N(N+5)/2$.

The matrix $a(X_t)$ is positive semidefinite so that its matrix square root $\sigma(X_t)$ can be obtained via Cholesky or spectral decomposition of (14), which shows that Assumption A.2 holds. Condition A.1 is satisfied: The drift term $b(x)$ is linear in $x$ such that its second order derivatives are zero. The diffusion term $a(x)$ consists of components $a_{YY}, a_{VV}, a_{Vq}$, and $a_{qq}$, all of which are matrix products, Kronecker of Hadamard products of $S_t$, $R_t$, $P_t$, and $q_t$. By the chain rule and the continuous differentiability of Kronecker and Hadamard products, the diffusion term $a(x)$ is also twice continuously differentiable.

Condition A.2 holds since the diffusion matrix and the inner product of the drift and the state variable $X$ are at most of order two in $X$: $\|a(x)\| = \mathrm{Tr}^{1/2}\big(a_{YY}^2 + a_{VV}^2 + a_{qq}^2 + 2a_{Vq}'a_{Vq}\big) \leq \|a_{YY}\| + \|a_{VV}\| + \|a_{qq}\| + 2\|a_{Vq}\|$, where $\|a_{YY}\| = \|S_t R_t S_t\| \leq \|S_t^2\|\|R_t\| \leq N\|V\| \leq N(1+\|V\|^2)$, $\|a_{VV}\| = 2\|AS_t^2(R_t \odot R_t)S_t^2 A\| \leq 2N\|A\|^2\|S_t^2\|^2 = 2N\|A\|^2\|V_t\|^2$, $\|a_{qq}\| \leq \vartheta^2(\|D_N^+(P_t \otimes P_t)D_N^{+'}K_t D_N^+(P_t \otimes P_t)D_N^{+'}\| + \|q_t q_t'\|)$. The first term can be bounded by $\vartheta^2\|D_N^+\|^4\|K_t\|\|P_t \otimes P_t\|^2$, where $\|K_t\| \leq 2\|D_N^+\|^2\|R_t \otimes R_t\| + \|\mathrm{vech}(R_t)\mathrm{vech}(R_t)'\| \leq 2N^2\|D_N^+\|^2 + N(N+1)/2 =: C_N$ and $\|P_t \otimes P_t\|^2 = \mathrm{Tr}(P_t^2 \otimes P_t^2) = \mathrm{Tr}(P_t^2)^2 = \mathrm{Tr}(Q_t)^2$. By the $c_r$-inequality, there is a constant $c_r$ such that $\mathrm{Tr}(Q_t)^2 \leq c_r \mathrm{Tr}(Q_t^2)$, and $\mathrm{Tr}(Q_t^2) \leq 2\|q_t\|^2$. Hence, $\|a_{qq}\| \leq \vartheta^2(2\|D_N^+\|^4 c_r C_N + 1)\|q_t\|^2$. By the Cauchy–Schwarz inequality, $\|a_{Vq}\| \leq \sqrt{\|a_{VV}\|\|a_{qq}\|} \leq \|a_{VV}\| \vee \|a_{qq}\|$. Finally, $\langle x, b(x)\rangle = cV_t - \Lambda V_t^2 + \phi\bar{q}q_t - \phi q_t^2$ and each of the terms can be bounded by $C_T(1+|x|^2)$. This proves that Condition A.2 holds and together with Condition A.1 ensures weak existence, uniqueness and nonexplosion of the diffusion process $X_t$ on compact sets. Thus Assumption A.4 holds, which completes the proof. ∎

**Proof of Theorem 2.** We first show that under the new conditions, Assumption A.1 holds. Assumption 1a) holds trivially. To show Assumption A.1b), consider the limit, as $h \to 0$, of the moments of interests (B.6), (B.8), and (B.11). The case $\delta_1 = \delta_2 = 0$ is covered by Theorem 1. If $\delta_1 > 0$, then $\lim_{h \to 0} h^{-1}\mathrm{Var}[\Delta V_{(k+1)h}^{(h)}|\mathcal{F}_{kh}] = 0$ provided that

$$\lim_{h \to 0} h^{-1/2}A_h = 0 \tag{B.18}$$

that is, $A_h$ is of order $h^{1/2+\delta_1}$, $\delta_1 > 0$.

Similarly, if $\delta_2 > 0$, then $\lim_{h \to 0} h^{-1}\mathrm{Var}[\Delta q_{(k+1)h}^{(h)}|\mathcal{F}_{kh}] = 0$ provided that

$$\lim_{h \to 0} h^{-1/2}\vartheta_h = 0 \tag{B.19}$$

that is, $\vartheta_h$ is of order $h^{1/2+\delta_2}$, $\delta_2 > 0$.

Either $\delta_1 > 0$ or $\delta_2 > 0$, or both, also ensure that

$$\lim_{h \to 0} h^{-1}\mathrm{Cov}\left[\Delta V_{(k+1)h}^{(h)}, \Delta q_{(k+1)h}^{(h)}|\mathcal{F}_{kh}\right] = 0.$$

Hence, under (B.18) and (B.19) Assumption A.1b) holds. Finally, Assumption A.1c) can be shown similar to the case $\delta_1 = \delta_2 = 0$ (Theorem 1).

The theorem of weak convergence applies by analogy to Theorem 1. Furthermore, depending on the combination of convergence conditions for $A_h$ and $\vartheta_h$, we either obtain a diffusion with deterministic variances and stochastic correlations (i.e., $\delta_1 > 0$ and $\delta_2 = 0$), or stochastic variances and deterministic correlations (i.e., $\delta_1 = 0$ and $\delta_2 > 0$), or deterministic variances and correlations (i.e., $\delta_1 > 0$ and $\delta_2 > 0$). The drift and diffusion matrices can be derived from those of Theorem 1. ∎