Continuous‐time threshold autoregressions with jumps: properties, estimation, and application to electricity markets

Daniel Lingohr, Gernot Müller

Angaben zur Veröffentlichung / Publication details:

Lingohr, Daniel, and Gernot Müller. 2023. "Continuous‐time threshold autoregressions with jumps: properties, estimation, and application to electricity markets." *Scandinavian Journal of Statistics* 50 (2): 638–64. https://doi.org/10.1111/sjos.12597.

Received: 16 April 2021 | Revised: 30 March 2022 | Accepted: 15 April 2022

DOI: 10.1111/sjos.12597

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Continuous-time threshold autoregressions with jumps: Properties, estimation, and application to electricity markets

Daniel Lingohr | Gernot Müller

Institute of Mathematics, University of Augsburg, Augsburg, Germany

Correspondence
Gernot Müller, Institute of Mathematics, University of Augsburg, Universitätsstraße 14, 86159 Augsburg, Germany.
Email: gernot.mueller@uni-a.de

Abstract

Continuous-time autoregressive processes have been applied successfully in many fields and are particularly advantageous in the modeling of irregularly spaced or high-frequency time series data. A convenient nonlinear extension of this model are continuous-time threshold autoregressions (CTAR). CTAR allow for greater flexibility in model parameters and can represent a regime switching behavior. However, so far only Gaussian CTAR processes have been defined, so that this model class could not be used for data with jumps, as frequently observed in financial applications. Hence, as a novelty, we construct CTAR processes with jumps in this paper. Existence of a unique weak solution and weak consistency of an Euler approximation scheme is proven. As a closed form expression of the likelihood is not available, we use kernel-based particle filtering for estimation. We fit our model to the Physical Electricity Index and show that it describes the data better than other comparable approaches.

KEYWORDS

consistency, discontinuous drift, Euler method, Girsanov formula, particle filtering, stochastic differential equations

1 INTRODUCTION

Continuous-time stochastic models have become increasingly important in recent decades. In particular, their successful application in derivatives pricing has brought them into the field of mathematical finance. Due to the outstanding importance of autoregressive processes in classical time series analysis, continuous-time processes with similar properties have been defined. The class of continuous-time autoregressive moving average (CARMA) processes (see Brockwell, 2014 for an overview) plays an important role here. CARMA processes are a continuous-time equivalent to classical ARMA time series. They allow the representation of a functional relationship between successive observations. As with ARMA processes, this relationship is limited to linear behavior. To overcome this, Tong (1983) introduced the class of threshold autoregressions (TAR) in discrete-time. Here, the parameters of an AR process are piecewise constant functions depending on the value of a stochastic quantity. This construction allowed Tong (1983) to successfully map nonlinear relationships. Brockwell and Hyndman (1992) extended the class of CARMA processes in the same way by defining continuous-time threshold autoregressions (CTAR). In contrast to CARMA processes, however, there is no direct link to the discrete-time analogue, due to the fact that for the continuous-time process any number of parameter changes between two observations are possible. In addition, since a parameter change always corresponds to a discrete transition, results for CARMA processes can not be transferred to CTAR. Therefore it is necessary besides a proof of existence to develop own estimation methods for CTAR. The existence and uniqueness of CTAR have been proved in Stramer et al. (1996). Moreover, in Brockwell and Hyndman (1992) a consistent simulation scheme was specified, which forms the basis for a heuristic procedure to estimate model parameters. Due to the discontinuity of the drift coefficient of the stochastic differential equation that defines CTAR, these results do not follow from the standard literature. However, it can be shown that the drift coefficient meets a linear growth condition. Such a property has already been shown to be useful to prove the convergence of approximation schemes for stochastic differential equation (SDE) (Gyöngy & Rasonyi, 2011; Yan, 2002). Also the papers by Chan and Stramer (1998) and Stramer (1999) deal with approximations for models with discontinuous coefficients. In particular, Stramer (1999) considers a local linearization scheme and this way extends the work in Chan and Stramer (1998) where the Euler scheme was investigated. Those papers also assess the CTAR(1) and CTAR(2) model (however, without a jump component, which is a main contribution of this paper, see below).

In addition to the occurrence of nonlinear dependencies, it has been found that data often has characteristics that do not interfere with normal distribution assumptions. Especially in financial markets extreme events can often be observed, which suggest the influence of a jump component. While corresponding extensions have been defined for CARMA processes (Brockwell, 2001b; Garcia et al., 2011), to our knowledge this is not the case for CTAR. This gap should be filled in this work. In doing so, the first two parts of the work represent a generalization of the result in Brockwell and Hyndman (1992). For the estimation, however, we present a new approach based on kernel-based approximation of the likelihood found by particle filtering. This approach is based on the idea of Rossi and Vila (2006). Kernel-based approximation in combination with particle filters can also be found in Crisan and Miguez (2014).

The paper is structured as follows. Section 2 proves the existence of a CTAR driven by a jump diffusion process using a generalized version of the Girsanov theorem. In Section 3 we show that an Euler approximation of the stochastic differential equation defining the CTAR converges weakly. This allows realizations to be generated from the unknown transition distribution. This is used in Section 4 to construct a maximum likelihood estimator using particle filtering methods. We test its quality in a simulation study. Finally, we apply our new model to the Physical Electricity Index and show that its use can provide a significant advantage over comparable models. For the readers convenience we shifted the long proofs of Sections 2 and 3 to an Appendix A.

2 | CTAR WITH JUMPS

Definition 1. A CTAR (p) with jumps $\{X(t)\}_{t\in[0,T]}$, $T>0$, of order $p\in\mathbb{N}$ is defined as the first component of the p-dimensional process $\{\mathbf{X}(t)\}_{t\in[0,T]}$ with initial $x^0\in\mathbb{R}^p$ satisfying

$$d\mathbf{X}(t) = \left[A(X(t))\mathbf{X}(t) - \mathbf{1}_p\beta(X(t))\right]dt + \mathbf{1}_p\left[\sigma dW(t) + dJ(t)\right], \tag{1}$$

$$\mathbf{X}(0) = x^0 \ a.s., \tag{2}$$

with $\mathbf{1}_p^T = (0 \ \cdots \ 0\ 1)$, $A(x) = \begin{pmatrix} 0 & I_{p-1} \\ & a(x)^T \end{pmatrix}$, and parameter functions

$$a(x)^T = (-a_{pi} \ \cdots \ -a_{1i}) \in \mathbb{R}^p, \ \ \beta(x) = \beta_i \in \mathbb{R}, \ \text{for } x \in R_i := [r_{i-1}, r_i). \tag{3}$$

The threshold values $-\infty = r_0 < r_1 < \cdots < r_l = \infty$ partition the real line. SDE (1) is driven by a Lévy jump-diffusion $L(t) := \sigma W(t) + J(t)$, where $\sigma > 0$, $\{W(t)\}_{t\in[0,T]}$ is a standard Brownian motion and $J(t) = \sum_{i=1}^{N_t}\gamma_i$ is a compound Poisson process with constant intensity λ and jumps of size $\gamma_i \overset{\text{i.i.d}}{\sim} F_\gamma$ independent of N. For $p=1$, $A(x)$ reduces to $-a_1(x)$.

By (3) we clearly see that the drift coefficient of the stochastic differential equation (1) is a discontinuous function. Therefore construction of a (strong) solution by common theorems fails, see for example, Protter (1990, V,3.). Fortunately, by use of Girsanov's theorem a unique weak solution still can be found.

Note, that in order to setup a CTAR model driven by a general Lévy process, we would have to allow the driving noise to be of pure jump type, that is, to account also for the case $\sigma = 0$. However, without Brownian motion, existence of a weak solution cannot be proven using any Girsanov-type theorem. Hence, for this a completely different approach would have to be established.

Remark 1. A weak solution to a SDE driven by a Lévy jump-diffusion $\sigma W + J$ exists, if there is a solution X on the driving system $(\Omega, \mathcal{F}, \mathbb{F}, P; Z)$, where $Z = \sigma\tilde{W} + \tilde{J}$ is a Lévy jump-diffusion with driving terms $\tilde{W} \overset{\text{d}}{=} W$, $\tilde{J} \overset{\text{d}}{=} J$, see Jacod and Shiryaev (2002, III,§2c).

For application of the Girsanov formula it is essential that for a suitable measurable function $H : [0,T] \times D([0,T],\mathbb{R}) \to \mathbb{R}$, the Doleans–Dade exponential,

$$Z(t) := Z(t,X) = \exp\left(\int_0^t H(s,X)dW_s - \frac{1}{2}\int_0^t H(s,X)^2 ds\right), \ 0 \le t \le T, \tag{4}$$

is a true martingale. For $X(t) \equiv W(t)$, Beneš approach, see Karatzas and Shreve (1998, 3.5.16), shows that this is true under a growth condition on H. Klebaner and Lipster (2011) extended this result for more general forms of the stochastic exponential. In the following theorem we show that under a second order condition on $L(t)$ this extension can be used to prove that there is a weak solution to (1) and (2) in sense of Remark 1. Furthermore this solution in unique and nonexplosive.

Furthermore, we always use the canoncial σ-algebra on $D([0,T),\mathbb{R})$, that is, the Borel σ-algebra associated with the Skorokhod topology.

Theorem 1. *For each Lévy jump-diffusion $\{L(t)\}_{t\in[0,T]}$ as in Definition 1 with $E(L(t)^2)<\infty$ and $x^0\in\mathbb{R}^p$, there is a unique (in law), nonexplosive weak solution of (1)-(2).*

Proof. See Appendix A.1. ■

Remark 2. The proof that $\hat{P}^i$ appearing in Appendix A.1 (proof of Theorem 1) is a probability measure can be carried out in the same way as for $\tilde{P}$. The main argument here is that $\hat{Z}$ is a martingale by the same reasons as $Z(t)$.

At first glance, our proof of uniqueness looks simpler as the one in Stramer et al. (1996, theorem 2.1) for the case of no jumps, which is based on a result occurring, for example, in Karatzas and Shreve (1998, proposition 5.3.10). However, this is because they have to prove first, that a proper probability measure exists, which we have already done in the main part of the proof of Theorem 1.

3 | APPROXIMATION BY A DISCRETE-TIME PROCESS

In practical application it is often necessary to simulate trajectories of the considered stochastic processes. This is because analytic expressions for functional relationships are hard to derive or even do not exist. Therefore Monte Carlo methods are used instead. As for CTAR we do not know how the explicit solution looks like so that, for simulation we have to rely on the defining SDE (1) instead. The most common numerical solution to a SDE is given by the Euler method, a first-order approximation. In the following an Euler representation for the CTAR with jumps is given and we show that this approximation is consistent. This is not trivial as for proving consistency of the Euler approximation one usually requires smoothness of the associated coefficient functions, see, for example, Jacod et al. (2005).

Let $(\Omega,\mathcal{A},P)$ be a probability space on which a solution $\{\mathbf{X}(t)\}_{t\in[0,T]}$ to the SDE (1) with $\mathbf{X}(0)=x^0$ exists. Then an approximation $\{\mathbf{X}^n(\tau_k)\}_{0=\tau_0<\cdots<\tau_n=T}$, $\tau_{k+1}-\tau_k=:\delta\equiv T/n$, to $\{\mathbf{X}(t)\}_{t\in[0,T]}$ is given by

$$\mathbf{X}^n(\tau_{k+1})=\mathbf{X}^n(\tau_k)+\left[A(X^n(\tau_k))\mathbf{X}^n(\tau_k)-\mathbf{1}_p\beta(X^n(\tau_k))\right]\delta$$
$$+\mathbf{1}_p\left[\sigma v_{k+1}\sqrt{\delta}+\gamma_{k+1}q_{k+1}\right],\ k=0,\ \dots,n-1, \tag{5}$$

$$\mathbf{X}^n(\tau_0)=x^0, \tag{6}$$

where $\{v_k\}\overset{\text{iid}}{\sim}\mathcal{N}(0,1)$ approximates the increments of a standard Brownian motion, $\{\gamma_k\}\overset{\text{iid}}{\sim}F_\gamma$ is a sequence of stochastic jump amplitudes with zero mean independent of the jump times $\{q_k\}\overset{\text{iid}}{\sim}\text{ber}(\lambda\delta)$. Thereby $\text{ber}(p)$ denotes the Bernoulli distribution with parameter p. This type of approximation of the process jump dynamic is also used in the Bernoulli Diffusion model, a discretized version of the Merton model, see Honoré (1998). If we extend the discrete time process $\mathbf{X}^n$ to the unit interval by defining $\mathbf{X}^n(t):=\mathbf{X}^n(\lfloor t/\delta\rfloor\delta)$, then $\{\mathbf{X}^n(t)\}_{t\in[0,T]}$ is called the discretized Euler scheme.

In the following we prove that (5) is a valid approximation of the SDE (1) as the distribution of $\mathbf{X}^n$ converges to the distribution of $\mathbf{X}$. The construction of this proof is based on the ideas of Yan (2002) who derived weak consistency of an Euler method with discontinuous coefficients on $C[0,T]$ using the occupation time formula. In the multidimensional case this requires the existence of at least one projection of the set of discontinuities onto a coordinate axis with nondegenerate diffusion coefficient such that the projected set has Lebesgue measure 0. Unfortunately, this is not possible in our case. Therefore, beside extending the proof of Yan (2002) to $D[0,T]$, we adopt an idea of Brockwell and Williams (1997) to show that the local time technique can still be used.

For this note that $b(\mathbf{X}) := A(X_1)\mathbf{X}$ satisfies a linear growth condition similar to (A6),

$$\|b(\mathbf{X}^n(t))\|^2 \leq K[1 + \|\mathbf{X}^n(t)\|^2], \ 0 \leq t \leq 1, \ n \in \mathbb{N},$$

for some $K > 0$, where $\|\cdot\|$ stands for the Euclidean norm in the appropriate space. In the following w.l.o.g. we assume that $T = 1$, $\beta = 0$, $x^0 \equiv 0$.

Lemma 1. *If $m_{\gamma,k} := E_{F_\gamma}(\gamma_1^k) < \infty$ for $k = 4$, then $E(\|\mathbf{X}^n(t)\|^4) < \infty$ for all $n \in \mathbb{N}$ and $t \in [0,1]$.*

Proof. As $\mathbf{X}^n(t) = \mathbf{X}^n(\tau_k)$, $\tau_k \leq t < \tau_{k+1}$, we only have to prove $E(\|\mathbf{X}^n(\tau_k)\|^4) < \infty$ for $k = 0, 1, \dots, n$. Let $n \in \mathbb{N}$ be arbitrary but fixed. By

$$\mathbf{X}^n(\tau_{k+1}) = \mathbf{X}^n(\tau_k) + \frac{1}{n} b(\mathbf{X}^n(\tau_k)) + \frac{\sigma \nu_{k+1}}{\sqrt{n}} \mathbf{1}_p + \gamma_{k+1} q_{k+1} \mathbf{1}_p,$$

we find

$$E(\|\mathbf{X}^n(\tau_{k+1})\|^4) \leq 4^3 \left[E(\|\mathbf{X}^n(\tau_k)\|^4) + \frac{2K^2}{n^4} \left(1 + E(\|\mathbf{X}^n(\tau_k)\|^4)\right) + \frac{3\sigma^4}{n^2} + \frac{\lambda}{n} m_{\gamma,4} \right],$$

where we used $(\sum_{i=1}^m a_i)^4 \leq m^3 (\sum_{i=1}^m a_i^4)$ for any real numbers a_i by Hölder's inequality. The statement follows now by induction on k. ■

Lemma 2. *If the condition of Lemma 1 is satisfied, then $\sup_{n \geq 1} E(\|\mathbf{X}^n(t)\|^4) < \infty$ for all $t \in [0,1]$.*

Proof. See Appendix A.2. ■

Proposition 1. *If the condition of Lemma 1 is satisfied, then the Euler scheme $\{\mathbf{X}^n : n \geq 1\}$ is tight in $D[0,1]$.*

Proof. See Appendix A.3. ■

Since $\{\mathbf{X}^n : n \geq 1\}$ is tight in $D[0,1]$, which is a separable and complete space under a metric d^0, topologically equivalent to the Skorokhod metric d, by Prohorov's theorem, see Billingsley (1999, theorem 5.1), $\{\mathbf{X}^n : n \geq 1\}$ is relatively compact in $D[0,1]$. Therefore each sequence $\{\mathbf{X}^{n_i} : i \geq 1\}$ contains some subsequence $\{\mathbf{X}^{n_{i(m)}} : m \geq 1\}$ converging weakly to some $\mathbf{X}$. By Skorokhod's representation theorem, see Billingsley (1999, theorem 6.7), there exist random elements $\mathbf{Y}^m$ and $\mathbf{Y}$ taking values in $D[0,1]$, defined on a common probability space $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{P})$, such that $\mathcal{L}(\mathbf{Y}^m) = \mathcal{L}(\mathbf{X}^{n_{i(m)}})$, $\forall m$, $\mathcal{L}(\mathbf{Y}) = \mathcal{L}(\mathbf{X})$ and $\mathbf{Y}^m \underset{m\to\infty}{\to} \mathbf{Y}$ almost surely in $D[0,1]$. In addition, by Van der Vaart and Wellner (1996, addendum 1.10.5), $\mathbf{Y}^m$ and $\mathbf{Y}$ can be chosen according to

$$\mathbf{Y}^m(\cdot,\overline{\omega}) = \mathbf{X}^{n_{i(m)}}(\cdot,\phi^m(\overline{\omega})),\ \ \mathbf{Y}(\cdot,\overline{\omega}) = \mathbf{X}(\cdot,\phi(\overline{\omega})), \tag{7}$$

with measurable maps $\phi^m : \overline{\Omega} \to \Omega$ and $P = \overline{P}\circ\phi^m$, for $m = 1, 2, \dots$. If we define $\overline{\nu}_k^m := \nu_k\circ\phi^m$, $\overline{\gamma}_k^m := \gamma_k\circ\phi^m$ and $\overline{q}_k^m := q_k^m\circ\phi^m$, $m = 1, 2, \dots$, the distribution of the random variables is not changed under $\overline{P}$. Therefore for every $m \geq 1$, $\mathbf{Y}^m$ satisfies

$$\mathbf{Y}^m(t) = \mathbf{Y}^m(\tau_k) + \frac{1}{n_{i(m)}}A(Y_1^m(\tau_k))\mathbf{Y}^m(\tau_k) + \mathbf{1}_p\left[\frac{\sigma\overline{\nu}_{k+1}^m}{\sqrt{n_{i(m)}}} + \overline{\gamma}_{k+1}^m\overline{q}_{k+1}^m\right],\ \ \tau_k \leq t < \tau_{k+1}. \tag{8}$$

Using the same representation of $\mathbf{Y}^m$ for $\overline{\gamma}^m \equiv 0$, that is, having no jump component in the driving process, Yan (2002) proved, that $\mathbf{Y}$ is a weak solution of the approximated SDE. Hence, for this case $\{\mathbf{X}^{n_{i(m)}} : m \geq 1\}$ converges weakly to the unique weak solution, which implies weak convergence of the Euler scheme. Extending the local time technique of Yan (2002), we now show that the occupation time of Y_1 in the set of discontinuities $D_a := \{x \in \mathbb{R} : x = r_i,\ i = 1, \dots, l\}$ of the drift function b is for $\overline{\gamma}^m \neq 0$ still of Lebesgue measure 0 almost surely. For this, let $C_b^2(\mathbb{R})$ denote the space of continuous functions f on $\mathbb{R}$ with $|f|$, $|f'|$, $|f''|$ bounded by a constant $b > 0$.

Lemma 3. *If the condition of Lemma 1 is satisfied, then for* $[Y_p](t) := [Y_p, Y_p](t)$, $t \in [0, 1]$,

$$E[Y_p](t) = t(\sigma^2 + m_{J,2}\lambda).$$

Furthermore, for any $f \in C_b^2(\mathbb{R})$,

$$E[f_b(Y_i), Y_p](t) = 0,\ \ i \neq p.$$

Proof. See Appendix A.4. ■

Lemma 4. *If the condition of Lemma 1 is satisfied, then*

$$E\left(\int_0^1 \mathbf{1}(Y_1(s) \in D_a)ds\right) = 0.$$

Proof. See Appendix A.5. ■

Lemma 5. *If the conditions of Lemma 1 are satisfied, then*

$$\frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} A(Y_1^m(\tau_{k-1}))\mathbf{Y}^m(\tau_{k-1}) \xrightarrow[m\to\infty]{L^1(\Omega)} \int_0^t A(Y_1(s))\mathbf{Y}(s)ds.$$

Proof. See Appendix A.6. ■

Theorem 2. *If* $E_{F_\gamma}(\gamma_1^4) < \infty$, *then the Euler scheme defined in (5) weakly converges to the unique weak solution of SDE (1) as* $n \to \infty$.

Proof. See Appendix A.7. ■

4 | STATISTICAL INFERENCE

In this section the problem of fitting a CTAR with jumps to a finite set of possibly irregularly spaced data is considered. For (linear) CAR the explicit solution to the SDE (1) allows to calculate the Gaussian likelihood of the observations with help of the discrete-time Kalman recursions, see Jones (1981). For CTAR an explicit solution is not found and so this approach cannot be used. If the data additionally is uniformly spaced, an alternative procedure for estimation of CAR processes is given by their discrete-time representation. A sampled CAR process satisfies a standard ARMA equation. By this fact fitting of a CAR process can be traced back to fitting of an ARMA process (see Brockwell, 2014, and references therein). Indeed, such a relationship does not hold for CTAR as noted by Hyndman (1992). Nonetheless, Gaussian CTAR models have been fitted to a variety of datasets. If the data are observed frequently one can use the stochastic exponential (4) to get estimators of the autoregressive parameters, see Brockwell et al. (2007). This approach is also possible for the CTAR with jumps but since high-frequency data is rarely available in practice, we do not want to go into detail here. Instead, we want to introduce an approach that can be used in a very general setup.

Let $\mathbf{Y}_n := \{y(t_1), y(t_2), \dots, y(t_n)\}$, $t_1 < t_2 < \cdots < t_n$, be a set of observations, where w.l.o.g we assume that $t_i - t_{i-1} = \delta\ \forall i$ as the introduced procedures extend in an obvious way for irregularly spaced data. For fitting of a CTAR process we are interested in evaluation of the likelihood $L(\theta; \mathbf{Y}_n)$, being given by

$$L(\theta; \mathbf{Y}_n) = f_\theta(y(t_1)) \prod_{i=2}^{n} f_\theta(y(t_i)|\mathbf{Y}_{i-1}), \tag{9}$$

where θ denotes the vector of all model parameters; f_θ will be used from now on generically for any density, where the exact meaning can be easily seen from its arguments.

A problem arising in representation (9) is that the densities $f_\theta(y(t_i)|\mathbf{Y}_{i-1})$ for the CTAR process are unknown. This makes a direct implementation of a maximum likelihood approach impossible. A possible way out is to estimate the parameters based on an estimator of the likelihood itself. A method to find a suitable estimate $\widehat{L}(\theta; \mathbf{Y}_n)$ is given via *particle filters*, also known as *sequential Monte Carlo* methods.

Particle filters were introduced by Gordon et al. (1993), Kitagawa (1993, 1996) in the framework of nonlinear and non-Gaussian state space models. Their use for likelihood evaluation was investigated, for example, in Pitt (2002). Subsequently, a large number of publications appeared, all addressing the application of particle filters in lots of different setups. As for continuous time models, Fearnhead et al. (2008) and Johannes et al. (2009) used them for diffusion-driven models, and Creal (2008) analyzed particle filters in different settings of Lévy-driven stochastic volatility models using Monte Carlo experiments. In the same framework such sequential Monte Carlo methods were also used in Jasra et al. (2011). Of course, also lots of publications on adapted versions of particle filters (e.g., random-weight particle filtering, Fearnhead et al., 2010) in the context of continuous time processes can be found in the literature.

Particle filters are Monte Carlo-type algorithms that represent the posterior density of a stochastic process by sampling a set of particles. They are designed for hidden Markov models, where the observations $\{y_t|\mathbf{X}_t\}$, conditional on a preliminary assumed to be unknown state $\mathbf{X}_t$, being independent and $\{\mathbf{X}_t\}$ is assumed to be Markovian. Obviously for our observations $y(t_i)$ and the state vector $\mathbf{X}(t_i)$ of (1) these assumptions are satisfied. Pitt (2002) used the representation

$$f_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \int f_\theta(y(t_i)|\mathbf{X}(t_i))f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})d\mathbf{X}(t_i), \tag{10}$$

to estimate the transition densities with help of Monte Carlo integration. The first density f_θ in the integrand is often called observation density. As for the CTAR process

$$f_\theta(y(t)|\mathbf{X}(t)) = \delta_{X_1(t)}(y(t)), \tag{11}$$

direct application of (10) is infeasible. Instead we restrict to the unknown state $\overline{\mathbf{X}}(t) := (X_2, \dots, X_p)(t)$, and use

$$f_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \int f_\theta(y(t_i)|\overline{\mathbf{X}}(t_{i-1}), y(t_{i-1}))f_\theta(\overline{\mathbf{X}}(t_{i-1})|\mathbf{Y}_{i-1})d\overline{\mathbf{X}}(t_{i-1}). \tag{12}$$

instead. Now the integrand is not longer known explicitly, but can be approximated by

$$\begin{aligned} f_\theta(y(t_i)|\overline{\mathbf{X}}^k(t_{i-1}), \mathbf{Y}_{i-1}) &= \int f_\theta(y(t_i), \overline{\mathbf{X}}(t_i)|\overline{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))d\overline{\mathbf{X}}(t_i) \\ &\approx \int \frac{1}{L}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i))K_h(\overline{\mathbf{X}}(t_i) - \overline{\mathbf{X}}^{kj}(t_i))d\overline{\mathbf{X}}(t_i) \\ &= \frac{1}{L}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i)), \end{aligned} \tag{13}$$

where K_h is a kernel with bandwidth h and $\{(y, \overline{\mathbf{X}})^{kj}(t_i)\}_{1\le j\le L}$ are random variables from $f_\theta(\mathbf{X}(t_i)|\overline{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$, conditional on $\{\overline{\mathbf{X}}^k(t_{i-1})\}_{1\le k\le N}, \overline{\mathbf{X}}^k(t_{i-1}) \sim f_\theta(\overline{\mathbf{X}}(t_{i-1})|\mathbf{Y}_{i-1})$; the latter density f_θ is often called filtering density. Now, by Monte Carlo integration of (12), we get

$$\hat{f}_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \frac{1}{NL}\sum_{k=1}^{N}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i)). \tag{14}$$

By the properties of kernel density estimators, convergence of $\hat{f}_\theta(y(t_i)|\mathbf{Y}_{i-1})$ to $f_\theta(y(t_i)|\mathbf{Y}_{i-1})$ is ensured. But this approximation requires $N \times L$ random number generations. A more efficient approximation scheme is given by the *convolution particle filter* of Rossi and Vila (2006). Rossi and Vila (2006) propose to sample from the joint density $f(y(t_i), \overline{\mathbf{X}}(t_i)|\mathbf{Y}_{i-1}))$ by first simulate $\mathbf{X}^k(t_i) \sim f_\theta(\mathbf{X}(t_i)|\overline{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$ followed by generation of an observation according to $g_\theta(y(t_i)|\mathbf{X}^k(t_i))$. Then the density of $(y(t_i), \mathbf{X}(t_i))$ is approximated by $(y^k(t_i), \mathbf{X}^k(t_i))$ with help of kernel density estimation. For CTAR by (11) we only have to simulate from $f_\theta(\mathbf{X}(t_i)|\overline{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$ which can be done directly with the Euler scheme presented in Section 3. Then

$$\begin{aligned} f_\theta(\overline{\mathbf{X}}(t_i)|\mathbf{Y}_i) &= \frac{f_\theta(\overline{\mathbf{X}}(t_i), y(t_i)|\mathbf{Y}_{i-1})}{f_\theta(y(t_i)|\mathbf{Y}_{i-1})} \\ &\approx \frac{\frac{1}{N}\sum_{k=1}^{N} K_h(y(t_i) - y^k(t_i))K_h(\overline{\mathbf{X}}(t_i) - \overline{\mathbf{X}}^k(t_i))}{\frac{1}{N}\sum_{k=1}^{N} K_h(y(t_i) - y^k(t_i))}. \end{aligned} \tag{15}$$

This approximation allows to generate samples according to $f_\theta(y(t_i)|\mathbf{Y}_{i-1})$. For that, starting with a guess of the initial states, we recursively resample from the set $\{\overline{\mathbf{X}}^k(t_i)\}_k$ with weights proportional to (15) followed by an Euler step. Thus, an estimator for the likelihood of the CTAR with jumps is

$$\widehat{L}(\theta, \mathbf{Y}_n) = \hat{f}_\theta(y(t_1)) \prod_{i=2}^{n} \hat{f}_\theta(y(t_i)|\mathbf{Y}_{i-1}), \tag{16}$$

where $L = 1$ in (14).

Remark 3. Brockwell and Hyndman (1992) replaced the transition densities in (12) by Gaussian densities with fitted first- and second-order moments and used Riemann sums to approximate the integrals. Beside the lack of theoretical backing, for the highly nonlinear CTAR with jumps model this method can be very inaccurate. Moreover, for large p and n, this approach suffers from the complicated approximation of higher-dimensional integrals.

We now try to find the maximum likelihood estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta, \mathbf{Y}_n),$$

by replacing the true likelihood $L(\theta, \mathbf{Y}_n)$ by its approximation $\widehat{L}(\theta, \mathbf{Y}_n)$.

Sampling with respect to the the true transition density, Vila (2012) proved that (15) and (16) converge almost surely to the true filtering density resp. likelihood by convergence properties of kernel density estimators. A similar result is still true if we use the Euler approximation at the evolving step. Therefore we assume K to be a bounded, positive, symmetrical application from $\mathbb{R}^p \to \mathbb{R}$, such that $\int K d\lambda = 1$, where λ is the Lebesgue measure. If we further assume $\lim_{||x|| \to \infty} ||x|| K(x) = 0$, K_h is called *Parzen–Rosenblatt kernel.*

Lemma 6. *Let $K_{h_N}(x) := K(x/h_N)/h_N^p$, where K is a Parzen–Rosenblatt kernel and h_N the bandwidth parameter with $0 < h_N \searrow 0$ and $Nh_N^p \to \infty$ as $N \to \infty$. Furthermore we assume the transition density f_θ is a positive function satisfying $f_\theta \in C_b(\mathbb{R}^p)$. Then*

$$\lim_{N\to\infty} \lim_{\delta_{\text{sim}}\to 0} E(\widehat{L}(\theta, \mathbf{Y}_n)) = L(\theta, \mathbf{Y}_n),$$

where the expectation is build with respect to all simulated variables and δ_{sim} is the step size of the Euler method in Section 3.

Proof. Given $\mathbf{X}^k(t_{i-1}) \sim f_\theta(\mathbf{X}(t_{i-1})|\mathbf{Y}_{i-1})$, $1 \le k \le N$, and samples $\mathbf{X}^k_{\delta_{\text{sim}}}(t_i)$ generated by the Euler method of Section 3 with step size δ_{sim} conditional on $\mathbf{X}^k(t_{i-1})$, using a kernel density estimator of $\hat{f}_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})$, we get

$$E_f(\hat{f}_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})) = E_f\left(\frac{1}{N}\sum_{k=1}^{N} K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k_{\delta_{\text{sim}}}(t_i))\right) \xrightarrow[\delta_{\text{sim}}\to 0]{} \frac{1}{N}\sum_{k=1}^{N} E_f(K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k(t_i))) \xrightarrow[N\to\infty]{} f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1}),$$

where the expectation is taken with respect to $f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})$. The first limit is true because $K_h \in C_b(\mathbb{R}^p)$ and the weak convergence of $\mathbf{X}^{kj}_{\delta_{\text{sim}}}(t_i)$, whereas the last equation holds by the asymptotic unbiasedness of kernel density estimators in continuity points of f_θ. By this it is easy to see that the approximated resampling weights (15) converge to the true resampling weights. Obviously this is also true for the likelihood approximation (14). Now the statement follows by induction using an iterated expectations argument as the particle filter estimator of the likelihood function is an unbiased estimator regardless of the number of particles N (see Pitt et al., 2012 for a direct proof). ∎

Theorem 3. *If the conditions in Lemma 6 are satisfied and $\int K^2 d\lambda < \infty$, then*

$$\hat{L}(\theta, \mathbf{Y}_n) \xrightarrow[N\to\infty, \delta_{\text{sim}}\to 0]{p} L(\theta, \mathbf{Y}_n).$$

Proof. Using Markov's inequality

$$P(|\hat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)| \geq \epsilon) \leq \frac{E\left(\hat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)\right)^2}{\epsilon^2}.$$

Hence, in view of Lemma 6 and by the Cauchy-Schwarz inequality it is sufficient to show

$$E\left(\hat{L}(\theta, \mathbf{Y}_n) - E(\hat{L}(\theta, \mathbf{Y}_n))\right)^2 \to 0, \ as \ \ N \to \infty, \ \ \delta_{\text{sim}} \to 0.$$

However, since $\forall t_i$

$$\text{Var}\left(\frac{1}{N}\sum_{k=1}^{N} K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k_{\delta_{\text{sim}}}(t_i))\right) \xrightarrow[\delta\to 0]{} \frac{1}{Nh_N^p}\int K^2(y) f_\theta(x + yh|\mathbf{Y}_{i-1})dy \xrightarrow[N\to\infty]{} 0,$$

the statement follows by an iterated expectations argument as in the proof of Lemma 6. ∎

Remark 4. The direct application of Lemma 6 and Theorem 3 to CTAR is generally not possible because we do not expect that the associated transition density is a continuous function. This assumption is corroborated by the fact that for the Gaussian CTAR(1) the density of the stationary distribution is discontinuous (Brockwell, 2001a). Since the discontinuities form almost surely a (Lebesgue) null set (see Lemma 4), there is hope that the asymptotic of likelihood is still valid. This, however, requires further investigation.

Chopin (2004), cf. also Malik and Pitt (2011), shows that for the particle filter-based estimator of the likelihood function

$$\sqrt{N}(\hat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, \sigma^2_{pf}), \tag{17}$$

where σ^2_{pf} is the particle filter variance and conditions under which $\sigma^2_{pf} < \infty$ can be found. This central limit theorem gives an idea about the error of likelihood approximation which can be interesting for model comparison as we will show in Section 5.

Remark 5. In numerical application one usually uses the logarithmic likelihood, that is, $\log(L)$, instead of L. Then the estimator $\log(\hat{L})$ is no longer asymptotically unbiased. Nevertheless by Chopin (2004), a result similar to (17) is still valid for $\log(\hat{L})$, where a bias of magnitude $-\sigma^2_{pf}/(2N)$ has to be considered.

Theorem 3 serves as a basis for the assumption that $\hat{\theta} = \text{argmax}_\theta \hat{L}(\theta, \mathbf{Y}_n)$ is a meaningful estimator for our process. This is also supported by the results of Gourieroux and Monfort (1996), which could show that given explicit densities

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\theta_0)),$$

where $I(\theta_0)$ is the expected information matrix at the true parameter value θ_0 as long as $n, N \to \infty$ and $\sqrt{n}/N \to 0$. However, even if $\hat{L}(\theta, \mathbf{Y}_n)$ is a consistent estimator of a smooth likelihood function and $\hat{\theta}$ is consistent, for a finite number of particles, it is hard to optimize (16) by usual numerical procedures, as approximation of the likelihood by particle filtering leads to a non-smooth behavior in θ. This is because resampling particles, in fact, is sampling from a discrete distribution. Thus, by each change of θ, resampling weights will change, and so possibly some particles are exchanged. As the replaced particles in general are not alike, $\hat{f}_\theta(y(t_i)|\mathbf{Y}_{i-1})$ will shift excessively. Maximizing the resulting rough surface can be extremely problematic. Note that this also is not overcome by using common random numbers. Therefore, Campillo and Rossi (2009) include the unknown parameters θ as an additional state. Although they do not assume any additional noise of the parameter state, parameter estimates in extended state space models suffer from a strong dependence of close observations. In order to obtain an estimator that takes all observations equally into account, we prefer a classic maximum likelihood approach. Lee (2008) proposed a tree-based resampling scheme to smooth a likelihood obtained by particle filtering with resampling by inducing significant correlation among the selected particles of consecutive runs. Figure 1 shows the effect of algorithm 6 of Lee (2008) for linear CAR(2) in the first dimension of the parameter space.

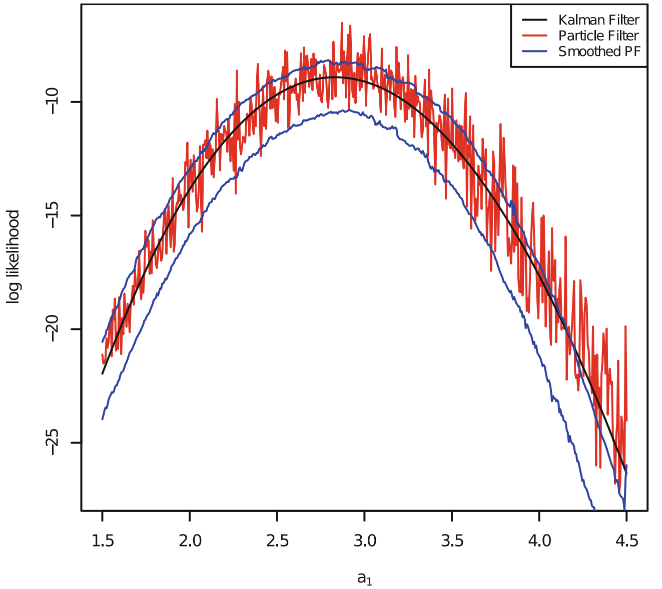

FIGURE 1 Estimated log-likelihood for a CAR(2) model as a function of a_1 using the smoothing procedure of Lee (2008) (blue) and the vanilla particle filter (red). The true log-likelihood is shown in black. For the smoothed log-likelihood we show the estimate for three sets of common random numbers.

To test the general quality of the likelihood approximation for continuous-time autoregressive models, beside the smoothed estimator $\log \hat{L}(\theta, \mathbf{Y}_n)$, we also drew the true log-likelihood $\log L(\theta, \mathbf{Y}_n)$ of the CAR(2) calculated by the Kalman filter. We conclude that the smoothed particle filter results in a likelihood estimate similar to the true density of the process. This gives hope to find reasonable estimates of the parameters based on an optimization of $\log \hat{L}(\theta, \mathbf{Y}_n)$.

In selecting an optimization procedure to determine $\hat{\theta}$ two features of the utility function $\hat{L}(\theta, \mathbf{Y}_n)$ should be considered. First, calculation of $\hat{L}(\theta, \mathbf{Y}_n)$ can be computationally intensive, which is a general drawback of many particle filter methods. The bottleneck of our method is given by smooth resampling procedure of Lee (2008) which requires construction of a binary tree. Construction of such a tree can be done in $\mathcal{O}(NL \log NL)$, see Lee (2008). Note that (15) is a mixture distribution which is easy to sample from as long as K is selected appropriately. Therefore, efficient ways of optimizing in the multidimensional search space are preferred. Secondly, the smoothing algorithm of Lee (2008) only provides correlated likelihood estimates, but does not guarantee continuity. A technique that can handle both is *simultaneous perturbation stochastic approximation* (SPSA), see Spall (2003, chapter 7). SPSA is a steepest ascent algorithm,

$$\hat{\theta}_{k+1} = \hat{\theta}_k + a_k \hat{\nabla}_{\hat{\theta}_k} L(\hat{\theta}_k, \mathbf{Y}_n), \; k = 1, \ldots, I_{\max},$$

where $\hat{\nabla}_{\hat{\theta}_k} L(\hat{\theta}_k, \mathbf{Y}_n)$ is a finite difference estimate of the gradient calculated by randomly perturbing all elements of $\hat{\theta}_k$ together to obtain two measurements of $L(\cdot, \mathbf{Y}_n)$. In context of particle filtering with intractable observation density $g_\theta(y(t)|\mathbf{X}(t))$, SPSA was already used successfully by Ehrlich et al. (2015). Table 1 reports the results of a simulation study in estimating the parameters of a jump diffusion CTAR(2) process by SPSA with the smoothed particle filter likelihood as utility function. The overall results are very satisfactory. Already with 500 observations meaningful estimation results are obtained. An increase in the number of observations will also increase the precision of the estimate in most cases. Since jumps are rare events, reliable estimates for the parameters of the associated distribution can only be made with an even greater number of observations.

TABLE 1 Estimated coefficients based on 40 replicates on $[0, T]$ of the CTAR(2) with jumps. The jumps are $\pm$Unif(a_γ, b_γ) distributed. For simulating the processes we used a bandwith of $\delta_{sim} = 0.01$ whereas observations are taken according to the stepsize $\delta_{obs} = 1$. For estimating the likelihood we used $N = 2048$ particles

		a_{11}	a_{12}	a_{21}	a_{22}	σ	λ	a_γ	b_γ	r_1
	True	**1.5**	**0.5**	**3**	**1**	**1**	**0.2**	**0.7**	**2.1**	**0.2**
$T = 500$	Mean	1.27	0.40	2.97	0.87	0.94	0.42	0.57	1.77	0.28
	Bias	−0.23	−0.10	−0.03	−0.13	−0.06	0.22	−0.13	−0.33	0.08
	SD	0.31	0.10	0.61	0.22	0.21	0.45	0.32	0.95	0.18
	MSE	0.15	0.02	0.37	0.07	0.05	0.25	0.12	1.01	0.04
$T = 1500$	Mean	1.22	0.41	2.93	0.83	0.96	0.30	0.64	1.77	0.24
	Bias	−0.28	−0.09	−0.07	−0.17	−0.04	0.10	−0.06	−0.33	0.04
	SD	0.20	0.05	0.39	0.22	0.11	0.18	0.38	0.56	0.10
	MSE	0.12	0.01	0.16	0.08	0.01	0.04	0.15	0.42	0.01

Looking at the MSE, we see clearly an improvement—with the exception of a_{22} and a_γ. Such phenomena can occur if the likelihood is quite flat around the maximum in the direction of those two parameters. In this case the precise estimation for these specific parameters is more difficult than for the other parameters, but at the same time the model fit is not deteriorating significantly, for example, in view of the AIC criterion, since the likelihood changes only very slowly along those parameters. Hence, also for those parameters precise estimates can only be achieved with an even greater number of observations.

5 | APPLICATION

Since the liberalization of the European energy market electricity, prices have shown a behavior rarely seen in other commodity markets. A pronounced seasonal pattern driven by the seasonality of demand is overlaid by a high volatility. Not only sudden outages of power plants but especially the integration of renewable energy sources has led to an electricity production that is difficult to predict. The fact that conventional power plants, in particular, are often unable to change their production flexibly is used by market participants to impose extreme prices in times of high or low demand. As failed supply is usually replaced fast by other flexible producers and extreme demand arises only for a few hours even extreme price levels return back to the seasonal mean in short time. The temporary persistence of deviations from the average level can be modeled by using a stochastic process with mean reversion. To incorporate spikes jump-diffusions are routinely used as driving process. However, as extreme prices are only temporary stable the true mean-reversion will not be fast enough. In estimation, this leads to an erroneous specification of the degree of mean-reversion, which will usually be upward biased. The common approach to overcome this is using multi-factor models, see Bierbrauer et al. (2007). These have the drawback that separating the effect of the different components can be difficult. Instead of combining different processes Borovkova and Permana (2006) allowed the mean-reversion parameter to be a continuous function depending on the value the price. While this approach seems intuitively appealing, there is no direct approach to transfer this to a multivariate model including higher autoregressive orders. As CTAR allows for a linear interpolation of arbitrary accuracy of the mean-reversion function, the CTAR with jumps can be seen as a multivariate extension to the model of Borovkova and Permana (2006).

In Figure 2 the *Phelix* (Physical Electricity Index) of the years 2014–2015 is shown.

The Phelix is a stock market index on the European Power Exchange for trading electricity in Germany and Austria. It represents the daily average of the day-ahead auction results. Here, we consider the daily base load index, which is calculated as average over all 24 h every day, for all 7 days of the week. As the most important underlying on the derivatives market, accurate modeling is of great importance. The CTAR with jumps is applied to the deseasonalized time-series, that is,

$$Y(t) = S(t) - \Lambda(t), \ \ t \geq 0,$$

where $S(t)$ is the price process and $\Lambda(t)$ is a deterministic seasonality function. Motivated by the seasonality function used in Benth et al. (2014), we take the seasonality function as a periodic function

$$\Lambda(t) = m_0 + \sum_{k=1}^{q} a_k \cos\left(\frac{2\pi t}{s_k}\right) + b_k \sin\left(\frac{2\pi t}{s_k}\right),$$

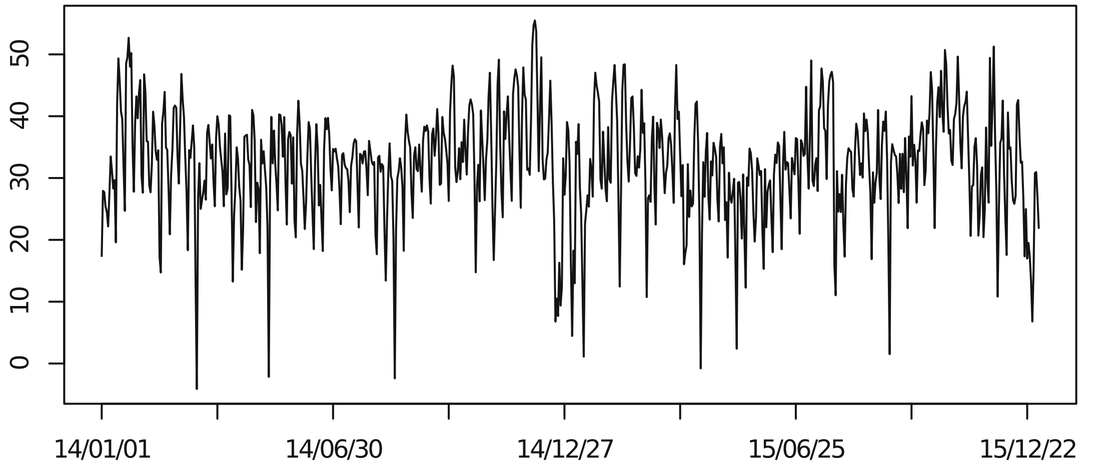

FIGURE 2 Physical electricity index January 1, 2014 to December 31, 2015

TABLE 2 Estimated parameters of the seasonality function $\Lambda(\cdot)$

m_0	a_1	b_1	a_2	b_2	a_3	b_3
32.23	5.04	3.21	−1.58	1.22	−2.68	−2.30

where we take $q = 3$ and the periods s_k are the dominant periods found by spectral analysis. The first three significant periods are $s_1 = 7, s_2 = 365$ and $s_3 = 3.5$. To estimate the parameters we used least-squares (Table 2).

After subtracting $\Lambda(t)$ from the price process, we want to test if there are really nonlinear effects in $Y(t)$, which would reinforce the use of a threshold model. Therefore we use an idea of Borovkova and Permana (2006). For the Gaussian CTAR(1) as noted in Brockwell (2001a), the stationary distribution has the density

$$\pi(x) = \frac{k}{\sigma^2} \exp\left(\frac{a(x)x^2}{\sigma^2}\right), \tag{18}$$

where k is a normalization constant. Thus, if we replace π by a estimate of the observations marginal density (e.g., a kernel estimator), then

$$-\log(\hat{\pi}(x)) = \frac{a(x)}{\sigma^2}x^2 + \log(\sigma^2) - \log(k).$$

By this,

$$-\log(\hat{\pi}(x))' \propto a(x)x,$$

which should be a linear function if there is no nonlinear effect, that is, $a(x) = a_1$. As (18) is only valid for the Gaussian process, we first filter the spikes in the data by considering those price movements as jumps which are outside of a $\pm 2\hat{\sigma}$ interval, where $\hat{\sigma}$ denotes the empirical standard deviation (note that this is just one possibility for filtering spikes, and also other procedures might be reasonable). The derivative of the negative logarithmic density estimator is shown in Figure 3 (for estimation we used the standard kernel function in R, which by default employs a

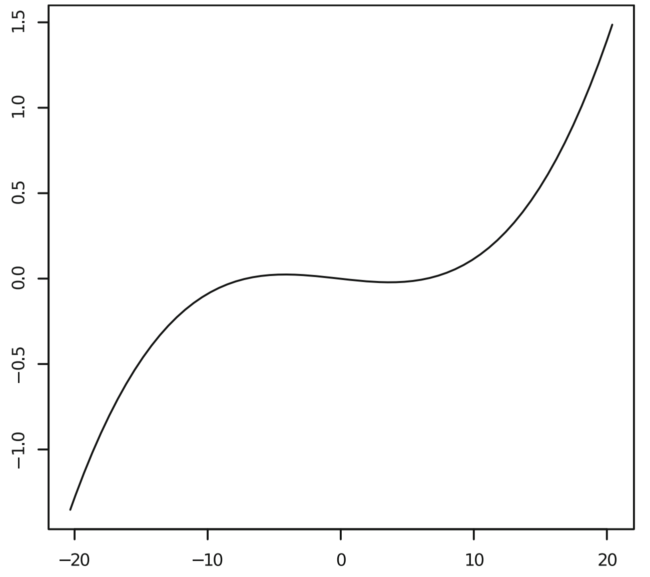

FIGURE 3 Estimated mean-reversion function (drift function)

Gaussian kernel; as bandwidth the function uses Silverman's rule of thumb, cf. Silverman, 1986, p. 48, equation (3.31)).

Figure 3 indicates that the mean-reversion is a nonlinear function with a high value for extreme prices and a low value for normal price levels. This is exactly what we expect for electricity prices and indicates the use of a nonlinear model.

To test if there is really a benefit in using a nonlinear autoregressive model we finally fit a Gaussian CTAR(2), a linear CAR(2) driven by a jump-diffusion and a CTAR(2) with jumps to $Y(t)$, using the observations in 2015. As jump distribution we use $\pm\text{Unif}(a_\gamma, b_\gamma)$, as in the simulation study (cf. Table 1). By Figure 3 the use of three regimes seems reasonable. As the upper regime would include only very less observations, to get clear estimates, we restrict to one threshold instead. This is also consistent with the fact that the negative jumps are dominant for our observations. For the autoregressive order we assumed the same value as in Benth et al. (2014). Table 3 contains all estimated model parameters, whereas Table 4 compares the fitted models by means of the deviation of empirical moments, their likelihood values and the AIC and BIC.

First we see that a linear CAR is not able to model the skewness in the data. This is obvious as this process is symmetric by definition as long as we use a driving process with unskewed distribution. Furthermore for the Gaussian CTAR there is a large bias in the mean. This can be explained by the dominant negative jumps in the data (see Figure 2), which cannot be adequately represented by a continuous process. In order to generate such extreme values, however, the parameters were estimated in such a way that the process is more frequently in the negative region. This leads to a downward biased mean. For the CTAR with jumps the empirical moments suit best to our data. This is further supported by its likelihood value which outperforms the values of the other model.

A more sophisticated measure to compare different models is given by the AIC. We see that the CTAR with jumps minimizes this criterion, from which we conclude that the use could offer an advantage. In Table 4 we report $\log(L)$ calculated for $N = 8192$ and $\delta = 1/50$. The corresponding standard deviation of the estimate was approximated numerically based on $M = 200$ simulations and can be found in Table 5.

TABLE 3 Fitted models for Phelix (2015)

Parameter	CAR(2) w.j.	CTAR(2)	CTAR(2) w.j.
a_{11} (a_1 resp.)	4.65	2.74	3.06
a_{12}	–	2.13	2.07
a_{21} (a_2 resp.)	3.11	6.38	4.86
a_{22}	—	2.83	2.21
σ	37.05	38.73	28.89
λ	0.09	—	0.08
a_γ	19.38	—	49.30
b_γ	55.48	—	70.10
r_1	-	−4.65	−4.93

TABLE 4 Linear CAR versus CTAR for Phelix (2015)

	Phelix	CAR(2) w.j.	CTAR(2)	CTAR(2) w.j.
Mean	−0.01	0.01	−0.66	−0.03
SD	7.08	7.21	7.97	7.86
Skewness	−0.54	0.00	−0.76	−0.33
Kurtosis	3.84	3.02	4.12	3.91
log(L)	\	−1145.54	−1139.94	−1125.20
AIC	\	2303.08	2291.88	2268.40
BIC	\	2345.15	2333.95	2331.50

TABLE 5 Empirical standard deviation of $\log(\hat{L})$ based on $M = 200$ simulations

	CTAR(2)	CAR(2) w.j.	CTAR(2) w.j.
$\hat{\sigma}_{pf}$	4.29	2.70	2.61

By these standard deviations and (17), the approximated log likelihood $\log(\hat{L})$ seems to be sufficiently close to the true log likelihood $\log(\hat{L})$ for such a conclusion (the bias introduced by using the logarithmic likelihood is negligible by Remark 5). Also the BIC is minimal for the CTAR(2) with jumps, although a conclusion in favor of this model is not as compelling as with the AIC.

Another way of comparing, for example, the CTAR(2) w.j. and the CAR(2) w.j. model is to use the quantity $\exp((\text{AIC}_{\text{CTAR(2)w.j.}} - \text{AIC}_{\text{CAR(2)w.j.}})/2)$ which is the (generalized) *relative likelihood* of CAR(2) w.j. w.r.t. to CTAR(2) w.j. It is closely related to the likelihood ratio in the likelihood-ratio test and can be interpreted as being proportional to the probability of minimizing the (estimated) information loss (cf. Burnham and Anderson (2002, sections 2.8, 2.9.1, and 6.4.5). In our application this leads to $\exp((2268.40 - 2303.08)/2) \approx 2.9 \cdot 10^{-8}$, which means that CTAR(2) w.j. minimizes the information loss with an extremely larger probability (factor $3.4 \cdot 10^7$) than CAR(2) w.j. To develop a test for these model classes which compares, for example, the parameters a_{ji} directly is, of course, an interesting task. A thorough investigation of this topic is, however, beyond the scope of this publication and is therefore left to a subsequent paper.

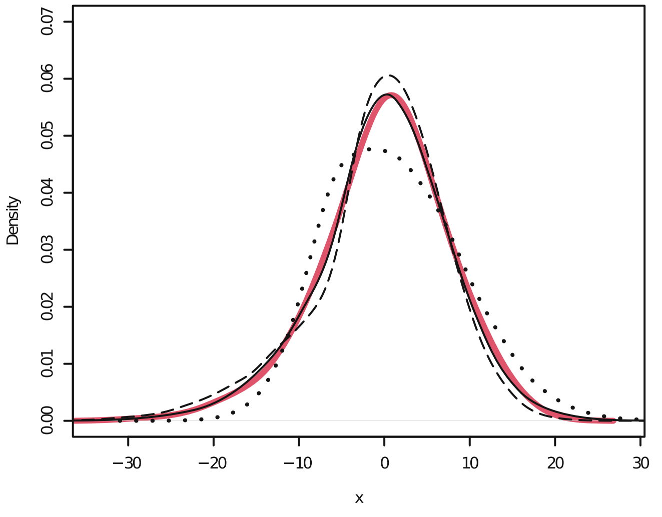

FIGURE 4 Kernel density estimates: Daily spot price data (red thick curve), CTAR(2) w.j. (solid black curve), CAR(2) w.j. (dotted curve), CTAR(2) (dashed curve)

Finally, as we use stationary processes, we compare the empirical distributions of the different models (based on 10^5 simulations each) together with the one of the deseasonalized spot price using kernel density estimates. The result is shown in Figure 4.

One can clearly see that the density curve of CTAR(2) w.j. is closer to the density curve of the observed data than the curves from the two competing models.

6 | CONCLUSION AND OUTLOOK

In this work we have extended CTAR such that the driving process can include a jump component. In addition to the proof of the existence of the newly defined process, we have introduced a consistent Euler method which enables the generation of observations from the correct distribution. While these points represent extensions of the ideas used by Brockwell and Hyndman (1992), we have examined a method for estimation that has not yet been considered for continuous autoregressive processes. The particle filter utilized has been formulated in such a way that it can be used for non-analytically representable densities and observation quantities which depend deterministically on the state vector. In order to numerically optimize the estimated objective function, a suitable procedure was selected and its quality was tested by means of a simulation study. Finally, we have shown that CTAR with jumps can offer a significant advantage over comparable models in describing electricity spot prices.

Deriving the stationary distribution of a CTAR model with jumps would be an interesting task. However, this follows not directly from the concepts used so far. Moreover, note that even for the Gaussian CTAR model the stationary distribution has been found only for $p = 1$, compare with Equation (18). Of course, an approximation approach based on simulations of the process seems feasible. However, since we do not need the stationary distribution for our purposes, we skipped this investigation to another paper.

Another interesting question is how to calculate the forward price (usually written as $E_Q(S_T|\mathcal{F}_t)$) based on our CTAR model and a risk neutral probability measure Q. A general idea to find risk neutral probability measures, also in the field of electricity price models, is using Esscher transforms, since they preserve the structure of the model dynamics, compare Benth and Sgarra (2012). In particular, the new measure is equivalent to the old one and the process with respect to the new measure inherits important features as, for example, the independent increment property. For an overview of how to apply Esscher transforms in the financial context see for example, Gerber and Shiu (1994). For estimating the forward price in our CTAR context, a similar approach as in Benth et al. (2014, sections 3 and 4.6), could probably be used. In that paper, the authors derived pricing formulas for electricity forwards based on a CARMA(2,1) spot price dynamics and also used Esscher transforms.

ACKNOWLEDGMENTS

We thank two referees for their excellent suggestions which led to an improved presentation of our paper.

ORCID

Gernot Müller https://orcid.org/0000-0002-4426-6252

REFERENCES

Benth, F. E., Klüppelberg, C., Müller, G., & Vos, L. (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics*, *44*, 392–406.

Benth, F. E., & Sgarra, C. (2012). The risk premium and the Esscher transform in power markets. *Stochastic Analysis and Applications*, *30*(1), 20–43.

Bierbrauer, M., Menn, C., Rachev, S. T., & Trück, S. (2007). Spot and derivative pricing in the EEX power market. *Journal of Banking and Finance*, *31*, 3462–3485.

Billingsley, P. (1999). *Convergence of probability measures* (2nd ed.). Wiley.

Borovkova, S., & Permana, F. J. (2006). *Modelling electricity prices by the potential jump-diffusion*. In A. Shiryaev, M. Grossinho, P. Oliveira, & M. Esquivel (Eds.), *Stochastic finance*. Springer.

Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference*, *39*, 291–303.

Brockwell, P. J. (2001a). *Continuous-time ARMA processes*. In D. N. Shanbhag & R. C. Rao (Eds.), *Handbook of statistics* (Vol. *19*, pp. 249–276). North-Holland.

Brockwell, P. J. (2001b). Levy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics*, *53*, 113–124.

Brockwell, P. J. (2014). Recent results in the theory and application of CARMA processes. *Annals of the Institute of Statistical Mathematics*, *66*, 647–685.

Brockwell, P. J., Davis, R. A., & Yu, Y. (2007). Continuous-time Gaussian autoregression. *Statistica Sinica*, *17*, 63–80.

Brockwell, P. J., & Hyndman, R. J. (1992). On continuous-time threshold autoregression. *International Journal of Forecasting*, *8*, 157–173.

Brockwell, P. J., & Williams, R. J. (1997). On the existence and application of continuous-time threshold autoregressions of order two. *The Advances in Applied Probability*, *29*, 205–227.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.

Campillo, F., & Rossi, V. (2009). Convolution particle filtering for parameter estimation in general state-space models. *IEEE Transactions on Aerospace and Electronic Systems*, *45*(3), 1063–1072.

Chan, K. S. & Stramer, O. (1998). Weak consistency of the Euler method for numerically solving stochastic differential equations with discontinuous coefficients. *Stochastic Processes and Their Applications*, *76*(1), 33–44.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, *32*(6), 2385–2411.

Cont, R., & Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC Press.

Creal, D. D. (2008). Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models. *Computational Statistics & Data Analysis*, *52*(6), 2863–2876.

Crisan, D., & Miguez, J. (2014). Particle-kernel estimation of the filter density in state-space models. *Bernoulli*, *20*(4), 1879–1929.

Ehrlich, E., Jasra, A., & Kantas, N. (2015). Gradient free parameter estimation for hidden Markov models with intractable likelihoods. *Methodology and Computing in Applied Probability*, *17*(3), 315–349.

Ethier, S. N., & Kurtz, T. G. (1986). *Markov processes: Characterization and convergence*. Wiley.

Fearnhead, P., Papaspiliopoulos, O., & Roberts, G. O. (2008). Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 755–777.

Fearnhead, P., Papaspiliopoulos, O., Roberts, G. O., & Stuart, A. (2010). Random-weight particle filtering of continuous time processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(4), 497–512.

Garcia, I., Klüppelberg, C., & Müller, G. (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Statistical Modelling*, *11*(5), 447–470.

Gerber, H. U., & Shiu, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of Society of Actuaries*, *46*, 99–140.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, *140*(2), 107–113.

Gourieroux, C., & Monfort, A. (1996). *Simulation based econometric methods*. Oxford University Press.

Gyöngy, I., & Rasonyi, M. (2011). A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic Processes and Their Applications*, *121*, 2189–2200.

Honoré, P. (1998). *Pitfalls in estimating jump-diffusion models* (Technical report). University of Aarhus Graduate School of Business.

Hyndman, R. H. (1992). *Continuous time threshold autoregressive models* [Ph. D. thesis]. University of Melbourne.

Jacod, J., Kurtz, T., Méléard, S., & Protter, P. (2005). The approximate Euler method for Lévy driven stochastic differential equations. *Annales de l'IHP Probabilités et statistiques*, *41*(3), 523–558.

Jacod, J., & Shiryaev, A. N. (2002). *Limit theorems for stochastic processes* (2nd ed.). Springer.

Jasra, A., Stephens, D. A., Doucet, A., & Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scandinavian Journal of Statistics*, *38*(1), 1–22.

Johannes, M. S., Polson, N. G., & Stroud, J. R. (2009). Optimal filtering of jump diffusions: Extracting latent states from asset prices. *The Review of Finanical Studies*, *22*(7), 2759–2799.

Jones, R. H. (1981). *Fitting a continuous time autoregression to discrete data*. In D. F. Findley (Ed.), *Applied time series analysis II* (pp. 651–682). Academic Press.

Karatzas, I., & Shreve, S. E. (1998). *Brownian motion and stochastic calculus* (2nd ed.). Springer Science+Business Media.

Kitagawa, G. (1993). *A Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space models*. In *Proceedings of the 2nd U. S.-Japan joint seminar on statistical time series analysis* (Vol. *2*, pp. 110–131).

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, *5*(1), 1–25.

Klebaner, F., & Lipster, R. (2011). When a stochastic exponential is a true martingale. Extension of the Beneš method. *Theory of Probability and Its Applications*, *58*(1), 38–62.

Kurtz, T. G., & Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, *19*(3), 1035–1070.

Lee, A. (2008). *Towards smoother multivariate particle filters* [Master's thesis]. University of British Columbia, Vancouver, BC.

Malik, S., & Pitt, M. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, *165*(2), 190–209.

Pitt, M., Silva, R., Giordani, P., & Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, *171*(2), 134–151.

Pitt, M. K. (2002). *Smooth particle filters for likelihood evaluation and maximisation* (Warwick Economic Research Papers). University of Warwick.

Protter, P. (1990). *Stochastic integration and differential equations*. Springer-Verlag.
Rossi, V., & Vila, J. (2006). Nonlinear filtering in discrete time: A particle filtering approach. *Annales de l'Institut de Statistique de l'Université de Paris*, *3*, 71–102.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC Press.
Spall, J. C. (2003). *Introduction to stochastic search and optimization*. Wiley.
Stramer, O. (1999). The local linearization scheme for nonlinear diffusion models with discontinuous coefficients. *Statistics & Probability Letters*, *42*(3), 249–256.
Stramer, O., Tweedie, R. L., & Brockwell, P. J. (1996). Existence and stability of continuous time threshold ARMA processes. *Statistica Sinica*, *6*(3), 715–732.
Tong, H. (1983). *Threshold models in non-linear time series analysis Springer lecture notes in statistics* (Vol. *21*). Springer-Verlag.
Van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag.
Vila, J. (2012). Enhanced consistency of the resampled convolution particle filter. *Statistics & Probability Letters*, *82*, 786–797.
Yan, L. (2002). The Euler scheme with irregular coefficients. *Annals of Probability*, *30*(3), 1172–1194.

How to cite this article: Lingohr, D., & Müller, G. (2023). Continuous-time threshold autoregressions with jumps: Properties, estimation, and application to electricity markets. *Scandinavian Journal of Statistics*, *50*(2), 638–664. https://doi.org/10.1111/sjos.12597

APPENDIX. PROOFS

A.1 Proof of Theorem 1

First we adopt the idea of Brockwell (1994), attributing a solution of (1) to a solution of a one-dimensional SDE. To this end, fix $x^0 \in \mathbb{R}^p$ and note that $L(t) = \sigma W(t) + J(t)$ has characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$. Writing equation (1) in coordinate form, we get

$$\begin{aligned} dX_1(t) &= X_2(t)dt, \\ dX_2(t) &= X_3(t)dt, \\ &\vdots \\ dX_{p-1}(t) &= X_p(t)dt, \\ dX_p(t) &= [-a_p X_1(t) - \cdots - a_1 X_p(t) - \beta]dt + dL(t), \end{aligned} \tag{A1}$$

where we have abbreviated $a_j(X_1(t))$ and $\beta(X_1(t))$ to a_j and β, respectively. Assuming $\mathbf{X}(0) = x^0$, we can write $\mathbf{X}(t)$ in terms of $\{X_p(s), 0 \le s \le t\}$ using the relation

$$X_j(t) = x_j^0 + \int_0^t \int_0^{s_{p-1-j}} \cdots \int_0^{s_2} X_p(s_1)ds_1 \cdots ds_{p-j},\ j = 1, 2, \ldots, p-1. \tag{A2}$$

The resulting functional relationship will be denoted by

$$\mathbf{X}(t) = \mathbf{F}(t, X_p). \tag{A3}$$

Substituting (A3) into the last equation in (A1), we see that it can be written in the form,

$$dX_p(t) = G(t, X_p)dt + dL(t), \tag{A4}$$

where $G(t, X_p)$, like $\mathbf{F}(t, X_p)$, depends on $\{X_p(s), 0 \le s \le t\}$.

Let $\Omega = D([0,T),\mathbb{R})$ be the Skorokhod space, $\mathcal{F}$ the canonical σ-algebra and $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ the canoncial filtration. Now, $G : [0,T) \times D([0,T),\mathbb{R}) \to \mathbb{R}$ is progressively measurable, see Karatzas and Shreve (1998, 1.1.13). We now want to show by use of the Girsanov formula for semimartingales, that there is a (weak) solution of equation (A4). For this, let P be a probability measure under which the coordinate mapping process $L(t)(\omega) = \omega(t)$, $0 \le t \le T$, $\omega \in \Omega$, is a Lévy process with the given characteristics. For this mapping $\mathcal{F}^L = \mathcal{F}$. Then we define for all $t \ge 0$,

$$Z(t) := Z(t,L) = \exp\left(\int_0^t \frac{G(s,L)}{\sigma} dW(s) - \frac{1}{2}\int_0^t \frac{G^2(s,L)}{\sigma^2} ds\right), \tag{A5}$$

By (A2), the function $G(t,L)$ satisfies the linear growth condition

$$G^2(t,L) \le K\left[1 + \sup_{s\le t} L^2(s)\right], \ t \le T. \tag{A6}$$

Since $E(L(t)) = 0$, by Cont and Tankov (2004, 3.17), $L(t)$ is a martingale and by Karatzas and Shreve (1998, 1.3.7), $L^2(t)$ is a submartingal. Therefore Doob‘s inequality can be used to show that by (A6) and $E(L^2(t)) < \infty$, $G(t,L)$ satisfies the conditions of Klebaner and Lipster (2011). Hence, $Z(t)$ is a martingale and

$$\tilde{P}(t,A) := \mathrm{E}(1_A Z(t)), \ A \in \mathcal{F}(t), \ 0 \le t \le T, \tag{A7}$$

defines a consistent family of probability measures, see Karatzas and Shreve (1998, p. 191). By Girsanov's theorem for semi-martingales (Jacod & Shiryaev, 2002, 3.24, III) the characteristics of $L(t)$ relative to $\tilde{P}(t)$ are

$$\left(\int_0^t \sigma\rho(s)ds, \sigma^2 t, Y \cdot t \cdot \lambda F_\gamma\right), \tag{A8}$$

where Y is a measurable nonnegative function and $\rho(t)$ is a predictable process uniquely determined by the equations

$$Y = M^P_{\mu^L}\left(\frac{Z}{Z_-}|\tilde{\mathcal{P}}\right), \tag{A9}$$

$$[Z^c, L^c] = \int \sigma\rho(s)Z(s-)ds. \tag{A10}$$

In Equation (A9), $M^P_\mu(X|\tilde{\mathcal{P}})$ is the conditional expectation of X on the probability space $(\tilde{\Omega}, \tilde{\mathcal{P}}, M^P_\mu)$ (see Jacod & Shiryaev, 2002,3.15, III, for further details) and $[X,Y] = \{[X,Y](t)\}_{t\in[0,T]}$ is the quadratic covariation of X, Y, see Protter (1990,6, II).

As $Z(t)$ is continuous, $Z^c(t) = Z(t-) = Z(t)$. We conclude $Y \equiv 1$ directly and $\rho(s) = G(s,L(s))/\sigma$ since

$$\begin{aligned}[Z^c, L^c](t) = [Z, \sigma W](t) &= \left[1 + \int Z\frac{G}{\sigma}dW, \sigma W\right](t) \\ &= \sigma\int_0^t Z(s)\frac{G(s,L)}{\sigma}d[W,W](s) = \int_0^t Z(s)G(s,L)ds = \int_0^t \sigma\rho(s)Z(s)ds.\end{aligned}$$

Hence by Jacod and Shiryaev (2002, 2.32, II) the process

$$\tilde{L}(t) := L(t) - \int_0^t G(s, L)ds,\ 0 \le t < T, \tag{A11}$$

has characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$ under $\tilde{P}$. As $\tilde{L}(t)$ is a Lévy process, see Jacod and Shiryaev (2002, 4,19, II), $(x^0 + L(t), \tilde{L}(t))$ is a weak solution of (A4) in sense of Remark 1. Furthermore by Klebaner and Lipster (2011, 5.1) this solution does not explode on any finite time interval $[0, T]$, that is $\sup_{t\in[0,T]} \mathrm{E}(X_p(t)^2) < \infty$.

To prove that the solution is unique in law, assume there are two weak solutions (X^i, L^i), $(\Omega^i, \mathcal{F}^i, P^i)$, $i = 1, 2$, to (A4) with the same initial value x^0. Now, using the above arguments reversed, the process $X^i(t)$ is a Lévy jump-diffusion with characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$ for the probability measure $\hat{P}^i(t)$ on $\mathcal{F}_t^i$, according to the prescription $d\hat{P}^i(t)/dP^i(t) = \hat{Z}(t, X^i)$, where

$$\hat{Z}(t, X^i) = \exp\left(-\int_0^t \frac{G(s, X^i)}{\sigma} dW(s) - \frac{1}{2}\int_0^t \frac{G^2(s, X^i)}{\sigma^2} ds\right).$$

Therefore, for $0 = t_0 < t_1 < \cdots < t_n \le T$ and $\Gamma \in \mathcal{B}(\mathbb{R}^{n+1})$, we have

$$\begin{aligned} P^1\left[(X^1(t_0), \ldots, X^1(t_n)) \in \Gamma\right] &= \int_\Gamma \frac{1}{\hat{Z}(s, X^1)} d\hat{P}^1 = \int_\Gamma \frac{1}{\hat{Z}(s, X^2)} d\hat{P}^2 \\ &= P^2\left[(X^2(t_0), \ldots, X^2(t_n)) \in \Gamma\right], \end{aligned}$$

concluding that the solution is unique in law. ∎

A.2 Proof of Lemma 2

First note that by Equation (5),

$$\mathbf{X}^n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1})) + \sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}} \mathbf{1}_p + \sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i \mathbf{1}_p,$$

and so by Hölder's inequality,

$$\|\mathbf{X}^n(t)\|^4 \le 4^3 \left[\left\|\frac{1}{n}\sum_{i=0}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1}))\right\|^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}}\right)^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i\right)^4\right].$$

By the fact that $\sum_{i=0}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1}))$ is constant for $\lfloor tn \rfloor / n \le t < \lfloor t(n+1) \rfloor / n$, Hölder's inequality, the at most linear growth of $\|b(\mathbf{X})\|$ and the Cauchy–Schwarz inequality,

$$\begin{aligned} \|\mathbf{X}^n(t)\|^4 &\le 4^3 \left[t^3 \int_0^t \|b(\mathbf{X}^n(s))\|^4 ds + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}}\right)^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i\right)^4\right] \\ &\le 4^3 \left[t^3 2K^2 \left(t + \int_0^t \|\mathbf{X}^n(s)\|^4 ds\right) + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}}\right)^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i\right)^4\right]. \end{aligned}$$

Therefore there exist two positive constants c_1 and c_2 independent of n such that

$$E\left(\|\mathbf{X}^n(t)\|^4\right) \leq c_1 + c_2\left[\int_0^t E\left(\|\mathbf{X}^n(s)\|^4\right) ds\right],$$

where the equality holds by Tonelli's theorem and Lemma 1. Let $f_t^n := E(\|\mathbf{X}^n(s)\|^4)$, then $f_t^n \leq c_1 + c_2\int_0^t f_s^n ds$. Now by Gronwall's lemma, $f_t^n \leq c_1\exp(c_2 t) \leq c_1\exp(c_2)$. This concludes the lemma, as c_1 and c_2 are independent of n. ∎

A.3 Proof of Proposition 1

To prove the tightness of the sequence $\{\mathbf{X}^n : n \geq 1\}$, we use Ethier and Kurtz (1986, theorem 9.8.6). By applying Markov's inequality and Lemma 2, we get first that for all t in $[0,1]$,

$$\limsup_{n\geq 1} P(\|\mathbf{X}^n(t)\| \geq a) \leq \limsup_{n\geq 1}\frac{E(\|\mathbf{X}^n(t)\|^4)}{a^4} \leq \frac{C}{a^4} \xrightarrow[a\to\infty]{} 0.$$

This shows, that conditions (a) in Ethier and Kurtz (1986, theorem 7.2) holds. To use Ethier and Kurtz (1986, theorem 9.8.6), by Ethier and Kurtz (1986, theorem 9.8.8), it remains to prove that

$$E\left(\|\mathbf{X}^n(t+h) - \mathbf{X}^n(t)\|\,\|\mathbf{X}^n(t) - \mathbf{X}^n(t-h)\|\right) \leq Ch,\ t\in[0,1],\ 0\leq h\leq t, \tag{A12}$$

for some $C > 0$. Note that for $h \geq 1/n$,

$$\begin{aligned}E(\|\mathbf{X}^n(t+h) - \mathbf{X}^n(t)\|^4) &= E\left(\left\|\sum_{i=\lfloor tn\rfloor+1}^{\lfloor (t+h)n\rfloor}\frac{1}{n}b(\mathbf{X}^n(\tau_{i-1})) + \frac{\sigma v_i}{\sqrt{n}}\mathbf{1}_p + \gamma_i q_i \mathbf{1}_p\right\|^4\right)\\ &\leq 3^3 E\left(\left\|\frac{1}{n}\sum_{i=\lfloor tn\rfloor+1}^{\lfloor (t+h)n\rfloor} b(\mathbf{X}^n(\tau_{i-1}))\right\|^4 + \left(\sum_{i=\lfloor tn\rfloor+1}^{\lfloor (t+h)n\rfloor}\frac{\sigma v_i}{\sqrt{n}}\right)^4 + \left(\sum_{i=\lfloor tn\rfloor+1}^{\lfloor (t+h)n\rfloor}\gamma_i q_i\right)^4\right)\\ &\leq 3^3 h\left(\frac{1}{n}\sum_{i=\lfloor tn\rfloor+1}^{\lfloor (t+h)n\rfloor}E(\|b(\mathbf{X}^n(\tau_{i-1}))\|^4) + \frac{3\sigma^2}{n} + \lambda m_{\gamma,4}\right) \leq Ch,\end{aligned}$$

where $C > 0$ is independent of n and we used the Cauchy–Schwarz inequality, the at most linear growth of $\|b(\mathbf{X})\|$ and Lemma 1. Since

$$\left(\mathbf{X}^n(t+h) - \mathbf{X}^n(t)\right) \wedge \left(\mathbf{X}^n(t) - \mathbf{X}^n(t-h)\right) = 0,\ h < \frac{1}{n},$$

condition (A12) holds. Therefore $\{\mathbf{X}^n : n \geq 1\}$ is tight in $D[0,1]$. ∎

A.4 Proof of Lemma 3

By the almost sure representation (8) it follows that

$$Y_p^m(\cdot) = y_p^0 + \frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor \cdot n_{i(m)}\rfloor} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \sum_{k=1}^{\lfloor \cdot n_{i(m)}\rfloor}\frac{\sigma \overline{v}_k^m}{\sqrt{n_{i(m)}}} + \sum_{k=1}^{\lfloor \cdot n_{i(m)}\rfloor}\overline{\gamma}_k^m\overline{q}_k^m.$$

$Y_p^m(t,\overline{\omega})$ is a cadlag function in t and therefore bounded on $[0,1]$ for each $\overline{\omega} \in \overline{\Omega}$. As $Y_p^m(t,\overline{\omega})$ is constant in t everywhere except of its $n_{i(m)}$ jump points, it is of finite variation almost sure. Therefore $Y_p^m(t)$ is a pure jump semi-martingale with respect to its natural filtration $\overline{\mathcal{F}}_t^m = \sigma(Y_p^m(t) : 0 \le t \le 1)$, thus, by Jacod and Shiryaev (2002, I.4.52),

$$
\begin{aligned}
[Y_p^m](t) &= \sum_{k=0}^{\lfloor tn_{i(m)}\rfloor} (\Delta Y_p^m(\tau_k))^2 = \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \left(\frac{1}{n_{i(m)}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{\sigma \overline{v}_k^m}{\sqrt{n_{i(m)}}} + \overline{\gamma}_k^m \overline{q}_k^m \right)^2 \\
&= \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \Bigg[\frac{1}{n_{i(m)}^2} b(\mathbf{Y}^m(\tau_{k-1}))^2 \mathbf{1}_p + \frac{\sigma^2 (\overline{v}_k^m)^2}{n_{i(m)}} + (\overline{\gamma}_k^m)^2 (\overline{q}_k^m)^2 + \\
&\quad + \frac{2\sigma \overline{v}_k^m}{n_{i(m)}^{3/2}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{2\overline{\gamma}_k^m \overline{q}_k^m}{n_{i(m)}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{2\sigma \overline{v}_k^m}{\sqrt{n_{i(m)}}} \overline{\gamma}_k^m \overline{q}_k^m \Bigg].
\end{aligned}
$$

and so,

$$
E[Y_p^m](t) = \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \left[\frac{1}{n_{i(m)}^2} E\left(b(\mathbf{Y}^m(\tau_{k-1}))^2 \mathbf{1}_p \right) + \frac{\sigma^2}{n_{i(m)}} + \frac{m_{\gamma,2}\lambda}{n_{i(m)}} \right].
$$

Now note that $\|b(\mathbf{X})\|$ has at most linear growth. At the same time, by Kurtz and Protter (1991, theorem 2.2), $[Y_p^m](t) \xrightarrow[m\to\infty]{p} [Y_p](t)$, as $[X] = X^2 - 2\int X_- dX$ and $\mathbf{Y}^m \xrightarrow[m\to\infty]{} \mathbf{Y}$ a.s. From Lemma 1 it can be seen that $E[Y_p^m]^2(t)$ is uniformly bounded for all t and m, and, therefore, $[Y_p^m]$ is uniformly integrable.

For the proof of the second statement w.l.o.g assume $p = 2$. Note that $f(Y_p^m)$ is still a semi-martingale as f is twice continuously differentiable. Hence, by the Lipschitz continuity of f and the Cauchy–Schwartz inequality, we get

$$
\begin{aligned}
E[f(Y_1^m), Y_2^m]^2(t) &= E\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \Delta f(Y_1^m(\tau_k)) \Delta Y_2^m(\tau_k) \right)^2 \\
&\le \lfloor tn_{i(m)}\rfloor E\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} (\Delta f(Y_1^m(\tau_k)))^2 (\Delta Y_2^m(\tau_k))^2 \right) \\
&\le \lfloor tn_{i(m)}\rfloor b^2 E\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} (\Delta(Y_1^m(\tau_k)))^2 (\Delta Y_2^m(\tau_k))^2 \right) \\
&\le \frac{tb^2}{n_{i(m)}} \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \left(E(Y_2^m(\tau_{k-1})^4) \right)^{\frac{1}{2}} \left(E(\Delta Y_2^m(\tau_k))^4 \right)^{\frac{1}{2}} \le \frac{Cb^2}{n_{i(m)}} \xrightarrow[m\to\infty]{} 0,
\end{aligned}
$$

where for the last inequality we used Lemma 1 and $\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \left(E(\Delta Y_2^m(\tau_k))^4 \right)^{\frac{1}{2}} \le C$, for some $C > 0$ independent of $n_{i(m)}$. This can be proved in the same way as in the proof of Proposition 1. By Kurtz and Protter (1991, theorem 2.2), we know in addition that $[f(Y_1^m), Y_2^m](t) \xrightarrow[m\to\infty]{p} [f(Y_1), Y_2](t)$, as $[X,Y] = XY - \int X_- dY - \int Y_- dX$, if $(f(Y_1^m), Y_2^m) \xrightarrow[m\to\infty]{p} (f(Y_1), Y_2))$ in the Skorohod topology on

$D_{\mathbb{R}^2}[0,1]$. But this is obvious, as $\mathbf{Y}^m \xrightarrow[m\to\infty]{\text{a.s.}} \mathbf{Y}$ and the continuous mapping theorem. Therefore, by the same argument as above, $[f(Y_1), Y_2](t) = 0, \forall t$, a.s.. ■

A.5 Proof of Lemma 4

We first consider the case $p = 1$. Apart from the driving noise, Y_1 has then the same properties as the process considered in Yan (2002, section 2). Therefore, we simply can extend the proof found there which is using the occupation time formula for a SDE driven by a Brownian motion. However, there is an extension of this formula for semi-martingales, see Protter (1990, section IV.5, corollary 1), and by Kurtz and Protter (1991, theorem 2.2), Y_1 is a semi-martingale. Let L^x be the local time of Y_1. Then

$$\int_0^1 \mathbf{1}(Y_1(s) \in D_a)d[Y_1](s) = \int_{x\in D_a} L_1^x dx = 0,$$

since $\lambda(D_a) = 0$ and $L^x(t) < \infty$ a.s. as each cadlag function has at most countable many discontinuities (Billingsley, 1999, p. 124). We conclude the lemma by Lemma 3 and $\sigma^2 + m_{J,2}\lambda > 0$.

To prove the Lemma in case of $p \geq 2$ we adopt an argument of Brockwell and Williams (1997), to show that even for a degenerate diffusion coefficient the amount of time that Y_1 spends in a neighborhood of r_1 is small with respect to the Lebesgue measure. W.l.o.g. we only consider $p = 2$, $l = 2$ and $r_1 = 0$. For $K > 0$ define $\nu_K := \inf\{t \in [0,1] : \|\mathbf{Y}(t)\| \geq K\}$ and fix ϵ such that $K > 1 > 2\epsilon > 0$. Then we choose a function $g \in C_b^3(\mathbb{R})$, where g is such that $g(y) = \int_0^y \int_0^w g''(u)dudw$ and g'' is an even function, $g''(y) = 1$ for $0 \leq y \leq \epsilon$, $g''(y) = 0$ for $2\epsilon \leq y \leq K$ and $g''(y)$ is nonincreasing for $\epsilon \leq y \leq 2\epsilon$. In particular for $|y| \leq K$, $|g'(y)| \leq 2\epsilon$ and $|g(y)| \leq 2\epsilon K$. Thus, we may assume g' is defined for $|y| > K$ such that $|g(y)| \leq 4\epsilon K$ and $|g'(y)| \leq 2\epsilon$ for all y. Brockwell and Williams (1997) used a martingale property of $Y_2 g'(Y_1)$ to show $\lambda(\{|Y_1| \leq \epsilon\}) \xrightarrow[\epsilon\to 0]{} 0$. Instead, we utilize a more direct approach based on the integration by parts formula for semi-martingales, see Protter (1990, II.2), to derive

$$E\left(\int_0^{t\wedge\nu_K} g''(Y_1(s))Y_2^2(s)ds\right) = E\left(Y_2(t\wedge\nu_K)g'(Y_1(t\wedge\nu_K))\right)$$
$$- E\left(\int_0^{t\wedge\nu_K} g'(Y_1(s))dY_2(s)\right) - [g'(Y_1), Y_2](t\wedge\nu_K) \leq CKt\epsilon, \text{ a.s.},$$

by Lemma 2, Lemma 3 and the boundedness of g', where $C > 0$. Thus, on letting ϵ tend to zero, $t \to 1$, and then $K \to \infty$,

$$E\left(\int_0^1 Y_2^2(s)1_{\{0\}}(Y_1(s))ds\right) = 0.$$

As $\int_0^1 1_{\{0\}}(Y_2(s))ds = 0$, a.s. (see case $p = 1$), it follows $\int_0^1 1_{\{0\}}(Y_1(s))ds = 0$, a.s.. This completes the proof. ■

A.6 Proof of Lemma 5

As the first $p-1$ components of $A(X_1)\mathbf{X}$ are continuous functions of $\mathbf{X}$ by $\mathbf{Y}^m \xrightarrow[m\to\infty]{\text{a.s.}} \mathbf{Y}$, Lemma 2 and the dominated convergence theorem, the statement can be proved directly for all components unequal to p. For the last component we have to use Lemma 4 in addition.

By Skorokhod's representation theorem,

$$d(\mathbf{Y}^m, \mathbf{Y}) \xrightarrow[m\to\infty]{} 0, \text{ a.s..}$$

As Skorokohod convergence implies that $\mathbf{Y}^m(t) \xrightarrow[m\to\infty]{} \mathbf{Y}(t)$ for all continuity points t of $\mathbf{Y}$ (Billingsley, 1999, p. 124), we have

$$b_p(\mathbf{Y}^m(t)) \xrightarrow[m\to\infty]{} b_p(\mathbf{Y}(t)), \ \forall t \in \overline{D_{\mathbf{Y}}} \cap \overline{D_{a(Y_1)}}, \text{ a.s.,}$$

where $D_{\mathbf{Y}}$ is the set of jump times of $\mathbf{Y}$ and $D_{a(Y_1)} := \{t \in [0,1] : Y_1 \in D_a\}$. As $\mathcal{L}(Y^m) = \mathcal{L}(X^{n_{i(m)}})$, by the at most linear growth of $\|b(\mathbf{X})\|$ and the proof of Lemma 2, $\sup_{m\geq 1} E\|b(\mathbf{Y}^m(t))\|^4$ is dominated by some constant on $[0,1]$. Therefore $\{b_p(\mathbf{Y}^m(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}}\}_{m\geq 1}$ is uniformly integrable. By Vitali's convergence theorem,

$$b_p(\mathbf{Y}^m(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}} \xrightarrow[m\to\infty]{L^1(\Omega)} b_p(\mathbf{Y}(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}}, \ \forall t \in [0,1]. \tag{A13}$$

Hence, we get

$$\begin{aligned}
&E\left|\frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} b_p(\mathbf{Y}^m(\tau_{k-1})) - \int_0^t b_p(\mathbf{Y}(s))ds\right| \\
&\leq E\left|\frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} b_p(\mathbf{Y}^m(\tau_{k-1})) - \int_0^t b_p(\mathbf{Y}^m(s))ds\right| + E\left|\int_0^t b_p(\mathbf{Y}^m(s))ds - \int_0^t b_p(\mathbf{Y}(s))ds\right| \\
&\leq E|b_p(\mathbf{Y}^m_{\tau_{\lfloor tn_{i(m)}\rfloor-1}})||t-\tau_{\lfloor tn_{i(m)}\rfloor-1}| + \int_0^t E|b_p(\mathbf{Y}^m(s)) - b_p(\mathbf{Y}(s))|ds \xrightarrow[m\to\infty]{} 0,
\end{aligned}$$

since $|t-\tau_{\lfloor tn_{i(m)}\rfloor-1}| \leq 1/n_{i(m)}$, $E(\lambda\left(D_{\mathbf{Y}} \cup D_{a(Y_1)}\right)) = 0$ by Lemma 4, Equation (A13) and the dominated convergence theorem. ∎

A.7 Proof of Theorem 2

We define

$$\begin{aligned}
\mathbf{Z}^m(t) &:= \mathbf{Y}^m(t) - \frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} A(Y_1^m(\tau_{k-1}))\mathbf{Y}^m(\tau_{k-1}) \\
&= \mathbf{1}_p\left[\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor}\frac{\sigma\overline{v_k}^m}{\sqrt{n_{i(m)}}} + \sum_{K=1}^{\lfloor tn_{i(m)}\rfloor}\overline{\gamma_k}^m\overline{q_k}^m\right], \\
\mathbf{Z}(t) &:= \mathbf{Y}(t) - \int_0^t A(Y_1^m(s))\mathbf{Y}^m(s)ds.
\end{aligned}$$

As $\mathbf{Y}^m \xrightarrow[m\to\infty]{} \mathbf{Y}$ a.s. and Lemma 5, $\mathbf{Z}^m$ converges to $\mathbf{Z}$ in probability. Therefore, we only have to prove

$$Z_p^m(t) \xrightarrow[m\to\infty]{d} \sigma W(t) + \sum_{i=1}^{N(t)}\gamma_i =: Z^*(t). \tag{A14}$$

Then, $Y_p(t) \stackrel{d}{=} \int_0^t b_p(\mathbf{Y}(s))ds + \sigma W(t) + \sum_{i=1}^{N(t)} \gamma_i$, that is, $\mathbf{Y}$ is the unique weak solution of SDE (1). Since $\mathbf{X}^{n_{i(m)}}$ converges weakly to $\mathbf{Y}$ and $\mathcal{L}(\mathbf{Y}) = \mathcal{L}(\mathbf{X})$, the Euler scheme converges weakly to the unique weak solution of SDE (1).

To prove (A14), we use Billingsley (1999, theorem 13.5). First, for any $\epsilon > 0$,

$$P(|Z^*(1) - Z^*(1-)| > \epsilon) = P(|\gamma_{N(1)}(N(1) - N(1-\delta))| > \epsilon) = 0,$$

by the continuity in probability of $N(t)$. Next, denote by $T_N \subset [0,1]$ the set of discontinuities of $N(t)$. W.l.o.g. let $t_1 < t_2$, $t_1, t_2 \in T_N$. Then

$$Ee^{i(s_1 Z^m(t_1) + s_2 Z^m(t_2))} = Ee^{i(s_1+s2)Z^m(t_1) + is_2(Z^m(t_2) - Z^m(t_1))} = Ee^{i(s_1+s2)\sum_{k=1}^{\lfloor t_1 n_{i(m)}\rfloor} \frac{\sigma \bar{v}_k^m}{\sqrt{n_{i(m)}}} + is_2 \sum_{k=\lfloor t_1 n_{i(m)}\rfloor+1}^{\lfloor t_2 n_{i(m)}\rfloor} \frac{\sigma \bar{v}_k^m}{\sqrt{n_{i(m)}}}}$$
$$\times Ee^{i(s_1+s2)\sum_{k=1}^{\lfloor t_1 n_{i(m)}\rfloor} \gamma_k^m q_k^m + is_2 \sum_{k=\lfloor t_1 n_{i(m)}\rfloor+1}^{\lfloor t_2 n_{i(m)}\rfloor} \gamma_k^m q_k^m}.$$

By Donsker‘s theorem, see Billingsley (1999, theorem 14.1), the first factor converges to $Ee^{i(s_1 W(t_1) + s_2 W(t_2))}$. To prove that the second factor tends to $Ee^{i(s_1 \sum_{i=1}^{N(t_1)} \gamma_i + s_2 \sum_{i=1}^{N(t_2)} \gamma_i)}$, we show $\sum_{i=1}^{\lfloor tn\rfloor} \bar{\gamma}_i^m \bar{q}_i^m \xrightarrow[n\to\infty]{d} \sum_{i=1}^{N(t)} \gamma_i$. This is true, because by the law of total probability

$$P\left(\sum_{i=1}^{\lfloor tn\rfloor} \bar{\gamma}_i^m \bar{q}_i^m \le x\right) = \sum_{k=1}^{\infty} F_{\gamma_1 + \dots + \gamma_k}(x) \mathrm{bin}\left(\lfloor tn\rfloor, \frac{\lambda}{n}\right)(k),$$

where $\mathrm{bin}(\lfloor tn\rfloor, \lambda/n) \xrightarrow[n\to\infty]{} \mathrm{pois}(\lambda t)$, since $\lfloor tn\rfloor \lambda/n \xrightarrow[n\to\infty]{} \lambda t$. Now, to conclude the proof of (A14), it is sufficient that Z^m fulfills a certain tightness condition, compare Billingsley (1999, inequality (13.14)), namely for $s \le u \le t$, $s, u, t \in [0,1]$, and $m \ge 1$,

$$E|Z^m(u) - Z^m(s)|^{2\beta} |Z^m(t) - Z^m(u)|^{2\beta} \le c(t-s)^{2\alpha},$$

where $\beta \ge 0$, $\alpha > 1/2$ and $c > 0$. This is true, as

$$\begin{aligned} E|Z^m(u) - Z^m(s)|^2 |Z^m(t) - Z^m(u)|^2 &= E|Z^m(u) - Z^m(s)|^2 E|Z^m(t) - Z^m(u)|^2 \le \\ &\le \left(E\left| \sum_{k=\lfloor sn_{i(m)}\rfloor+1}^{\lfloor un_{i(m)}\rfloor} \frac{\sigma \bar{v}_k^m}{\sqrt{n}} \right|^2 + E\left| \sum_{k=\lfloor sn_{i(m)}\rfloor+1}^{\lfloor un_{i(m)}\rfloor} \bar{\gamma}^m \bar{p}_k^m \right|^2 \right) \\ &\quad \times \left(E\left| \sum_{k=\lfloor un_{i(m)}\rfloor+1}^{\lfloor tn_{i(m)}\rfloor} \frac{\sigma \bar{v}_k^m}{\sqrt{n}} \right|^2 + E\left| \sum_{k=\lfloor un_{i(m)}\rfloor+1}^{\lfloor tn_{i(m)}\rfloor} \bar{\gamma}^m \bar{p}_k^m \right|^2 \right) \le \\ &\le (\lfloor un_{i(m)}\rfloor - \lfloor sn_{i(m)}\rfloor)(\lfloor tn_{i(m)}\rfloor - \lfloor un_{i(m)}\rfloor) \left(\frac{\sigma^2}{n} + m_{J,2} \frac{\lambda}{n} \left(1 - \frac{\lambda}{n}\right) \right)^2 \le \\ &\le \begin{cases} c(u - s + \frac{1}{n})(t - u + \frac{1}{n}) \le 4c(t-s)^2, & t - s \ge \frac{1}{n}, s < u < t, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

■