HAL
open science

Dependence modeling between continuous time stochastic processes : an application to electricity markets modeling and risk management

Thomas Deschatre

To cite this version:

Thomas Deschatre. Dependence modeling between continuous time stochastic processes : an application to electricity markets modeling and risk management. General Mathematics [math.GM]. Université Paris sciences et lettres, 2017. English. NNT : 2017PSLED034 . tel-01753215

HAL Id: tel-01753215

https://theses.hal.science/tel-01753215v1

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Université Paris-Dauphine

Dependence modeling between continuous time stochastic processes: an application to electricity markets modeling and risk management

École Doctorale de Dauphine — ED 543

Spécialité **Sciences**

COMPOSITION DU JURY :

M. Marc Hoffmann
Professeur, Université Paris-Dauphine
Directeur de thèse

M. Olivier Féron
Ingénieur, EDF R&D
Directeur de thèse

M. Markus Bibinger
Professeur, Philipps-Universität Marburg
Rapporteur

M. Peter Tankov,
Professeur, ENSAE ParisTech
Rapporteur

M. Jean-David Fermanian
Professeur, ENSAE ParisTech
Président du jury

M. Vincent Rivoirard
Professeur, Université Paris-Dauphine
Membre du jury

Soutenue le 08/12/2017
par Thomas Deschatre

Dirigée par **Marc Hoffmann**

DAUPHINE UNIVERSITÉ PARIS | PSL RESEARCH UNIVERSITY PARIS

Abstract

In this thesis, we study some dependence modeling problems between continuous time stochastic processes. These results are applied to the modeling and risk management of electricity markets.

In a first part, we propose new copulae to model the dependence between two Brownian motions and to control the distribution of their difference. We show that the class of admissible copulae for the Brownian motions contains asymmetric copulae. These copulae allow for the survival function of the difference between two Brownian motions to have higher value in the right tail than in the Gaussian copula case. Results are applied to the joint modeling of electricity and other energy commodity prices.

In a second part, we consider a stochastic process which is a sum of a continuous semimartingale and a mean reverting compound Poisson process and which is discretely observed. An estimation procedure is proposed for the mean reversion parameter of the Poisson process in a high frequency framework with finite time horizon, assuming this parameter is large. Results are applied to the modeling of the spikes in electricity prices time series.

In a third part, we consider a doubly stochastic Poisson process with stochastic intensity function of a continuous semimartingale. A local polynomial estimator is considered in order to infer the intensity function and a method is given to select the optimal bandwidth. An oracle inequality is derived. Furthermore, a test is proposed in order to determine if the intensity function belongs to some parametrical family. Using these results, we model the dependence between the intensity of electricity spikes and exogenous factors such as the wind production.

Keywords

Dependence ; Copula ; Brownian motion ; High frequency statistics ; Semimartingale ; Poisson process ; Stochastic intensity ; Non parametric estimation ; Local polynomial estimation ; Bandwidth selection ; Oracle inequality ; Electricity markets ; Spikes ; Wind production ; Risk management ; Mathematical finance.

Résumé

Cette thèse traite de problèmes de dépendance entre processus stochastiques en temps continu. Ces résultats sont appliqués à la modélisation et à la gestion des risques des marchés de l'électricité.

Dans une première partie, de nouvelles copules sont établies pour modéliser la dépendance entre deux mouvements Browniens et contrôler la distribution de leur différence. On montre que la classe des copules admissibles pour les Browniens contient des copules asymétriques. Avec ces copules, la fonction de survie de la différence des deux Browniens est plus élevée dans sa partie positive qu'avec une dépendance gaussienne. Les résultats sont appliqués à la modélisation jointe des prix de l'électricité et d'autres commodités énergétiques.

Dans une seconde partie, nous considérons un processus stochastique observé de manière discrète et défini par la somme d'une semimartingale continue et d'un processus de Poisson composé avec retour à la moyenne. Une procédure d'estimation pour le paramètre de retour à la moyenne est proposée lorsque celui-ci est élevé dans un cadre de statistique haute fréquence en horizon fini. Ces résultats sont utilisés pour la modélisation des pics dans les prix de l'électricité.

Dans une troisième partie, on considère un processus de Poisson doublement stochastique dont l'intensité stochastique est une fonction d'une semimartingale continue. Pour estimer cette fonction, un estimateur à polynômes locaux est utilisé et une méthode de sélection de la fenêtre est proposée menant à une inégalité oracle. Un test est proposé pour déterminer si la fonction d'intensité appartient à une certaine famille paramétrique. Grâce à ces résultats, on modélise la dépendance entre l'intensité des pics de prix de l'électricité et de facteurs exogènes tels que la production éolienne.

Mots-clés

Dépendance ; Copule ; Mouvement Brownien ; Statistique haute fréquence ; Semimartingale ; Processus de Poisson ; Intensité stochastique ; Estimation non paramétrique ; Estimateur à polynômes locaux ; Sélection de fenêtre ; Inégalité oracle ; Marchés de l'électricité ; Pics ; Production éolienne ; Gestion des risques ; Finance mathématique.

Remerciements

Je remercie tout d'abord Olivier Féron et Marc Hoffmann qui m'ont encadré tout au long de ces trois années. Ce travail n'aurait pu être ce qu'il est sans leur soutien et leurs conseils. Olivier, tu as toujours su mettre mon travail et mes compétences en avant, et tu as toujours su tirer de moi le meilleur. Tu as été un excellent encadrant, à la fois sur le plan technique et surtout le plan humain. Marc, tu as su me pousser au-delà de mes retranchements. Ton exigence sans précédent m'a permis d'aller plus loin dans mes réflexions, d'élever mes travaux à un autre niveau et ainsi de découvrir les subtilités de la statistique des processus (ou du moins une partie). Vous avez renforcé ma passion pour les mathématiques et contribué à la création de celle pour les statistiques. Pour cela, je vous remercie encore.

Je remercie Markus Bibinger et Peter Tankov d'avoir lu de manière très attentive mon manuscrit et d'avoir écrit des rapports très constructifs. Vos remarques sont particulièrement pertinentes. Je remercie aussi Vincent Rivoirard et Jean-David Fermanian qui ont accepté de faire partie de mon jury, j'en suis très honoré.

Je remercie bien sûr Almut Veraart qui m'a accueilli au département statistiques de l'Imperial College. Ce mois passé à Londres a débouché sur une collaboration très fructueuse, dans un cadre très agréable. Cela a été un vrai plaisir de travailler avec vous.

Je souhaite remercier le CEREMADE qui m'a accueilli pendant ces trois ans. En particulier, merci à Christian Robert qui m'a accordé sa confiance pour effectuer les travaux dirigés de statistiques ; ce fut une expérience particulièrement enrichissante. Merci énormément à Gilles Barès et Thomas Duleu qui ont été très réactifs à chacune de mes demandes et qui m'ont permis de résoudre de nombreux problèmes.

Merci à tous les membres du laboratoire FIME, laboratoire qui constitue pour moi un lieu d'échange scientifique et convivial sans égal. Plus particulièrement, je remercie Damien qui m'a accordé une place dans son bureau et avec qui j'ai pu participer à de nombreuses discussions très enrichissantes. Je pense que mon alimentation est de loin beaucoup plus écologique que ce qu'elle n'était au début de ma thèse. Je remercie aussi Clémence qui a toujours été à l'écoute et a pu me donner de nombreux conseils sur mes travaux, et qui m'a laissé la chance d'intervenir dans le cadre du séminaire FIME. Merci à René, qui en plus d'être un très bon chercheur, fut un très bon chef de département. Tu t'es toujours intéressé à ce que chacun faisait. Merci de m'avoir donné l'envie d'effectuer mon doctorat chez EDF, sur les marchés de l'électricité, lors d'un séminaire qui fut l'un des plus passionnants lors de mon année à l'ENSAE le vendredi midi. Je vous remercie aussi de m'avoir invité au séminaire à Florence qui fut un voyage plus qu'agréable et qui restera un très bon souvenir.

N'oublions pas de remercier le département OSIRIS à EDF pour m'avoir accueilli durant ces trois ans et en particulier le groupe R32, faisant volontairement ici un parachronisme. L'ambiance

a toujours été au meilleur et chacun a contribué chaque jour à la bonne cohérence du groupe. Michael, Pierre et Pierre, merci pour "les nombreuses discussions sur le C++" mais aussi pour les différents moments rue des Ecoles, ou place Denfert Rochereau, ou rue Brisemiche ; je dois ainsi inclure Anne-Laure, Emma et Guillaume dans cette phrase. On pourra aussi ajouter Erwan qui a contribué de manière non négligeable à l'ambiance à Vienne. Je souhaite remercier une deuxième fois Guillaume qui m'a inclus au sein du projet MAPE, et qui m'a permis d'augmenter les champs d'applications de mes travaux. Je remercie aussi Pierre une deuxième fois pour tous ses conseils en tant que doctorant puis docteur. Je remercie Hugo Gevret avec qui j'ai eu d'innombrables échanges sur l'optimisation de la production (ϵ% du temps) et sur d'autres sujets ($1 - \epsilon$% du temps), en particulier durant les voyages de département.

Sur un plan plus personnel, je souhaite remercier mes amis qui ont été plus que présents pour moi, durant et avant ces trois années, et avec qui j'ai passé des moments inoubliables, hors du monde des mathématiques. Je souhaite remercier ma famille et ma belle famille, qui m'ont entre autres fourni les nutriments nécessaires à la bonne réussite de cette thèse via de nombreux plats cuisinés. Enfin, merci à Lise pour son soutien chaque jour mais surtout pour sa (très) bonne humeur permanente qui ont été plus qu'indispensables.

Contents

Abstract . . . iii
Résumé . . . v
Remerciements . . . vii

Introduction **1**
1 Context . . . 1
1.1 Motivation . . . 1
1.2 Description of electricity markets . . . 2
1.3 Some modeling aspects . . . 4
2 First part: Dependence modeling for Brownian motions . . . 6
3 Second part: Inference of a spike process in high frequency statistics . . . 11
4 Third part: Non parametric estimation of the intensity of a doubly stochastic Poisson process depending on a covariable . . . 15
5 Structure of the thesis . . . 21

Introduction (French) **27**
1 Contexte . . . 27
1.1 Objet de la thèse . . . 27
1.2 Description des marchés de l'électricité . . . 28
1.3 Quelques aspects de modélisation . . . 30
2 Première partie: Modélisation de la dépendance entre mouvements Browniens . . . 32
3 Deuxième partie: Estimation d'un processus de pics en statistiques haute fréquence 38
4 Troisième partie: Estimation non paramétrique de l'intensité d'un processus de Poisson doublement stochastique fonction d'une covariable . . . 42
5 Structure de la thèse . . . 48

1 On the control of the difference between two Brownian motions : a dynamic copula approach **53**
1 Introduction . . . 54
1.1 Motivation . . . 54
1.2 Objectives and results . . . 55
1.3 Structure of the paper . . . 56
2 Markov Diffusion Copulae . . . 56
2.1 Definition . . . 57
2.2 Brownian motion case . . . 58
3 Reflection Brownian Copula . . . 59
3.1 Closed formula for the copula . . . 59

3.2 Extensions . . . 60
4 Control of the distribution of the difference between two Brownian motions . . . 61
4.1 Impact of symmetry on $S_{\eta,t}$ . . . 62
4.2 The Gaussian Random Variables Case . . . 63
4.3 The Brownian Motion Case . . . 65
5 Proofs . . . 67
5.1 Preliminary results . . . 67
5.2 Proof of Proposition 1.1 . . . 67
5.3 Proof of Proposition 1.2 . . . 68
5.4 Proof of Proposition 1.3 . . . 70
5.5 Proof of Proposition 1.4 . . . 71
5.6 Proof of Proposition 1.5 . . . 72
5.7 Proof of Proposition 1.6 . . . 73

2 On the control of the difference between two Brownian motions: an application to energy markets modeling 77
1 Introduction . . . 78
1.1 Motivation . . . 78
1.2 Objectives and results . . . 79
1.3 Structure of the paper . . . 80
2 A two-state correlation copula . . . 81
2.1 Model . . . 81
2.2 The copula . . . 81
2.3 Distribution of the difference between the two Brownian motions . . . 82
3 Multi-barrier correlation model . . . 83
3.1 Model . . . 84
3.2 Results on the distribution of the difference between the two Brownian motions 86
3.3 A local correlation model . . . 87
4 An application for joint modeling of commodity prices on energy market . . . 90
4.1 Model . . . 90
4.2 Parameters . . . 91
4.3 Numerical results . . . 92
4.4 Pricing of European spread options . . . 94
5 Proofs . . . 96
5.1 Preliminary results . . . 96
5.2 Proof of Proposition 2.2 . . . 97
5.3 Proof of Proposition 2.3 . . . 99
5.4 Proof of Proposition 2.4 and Corollary 2.1 . . . 101
5.5 Proof of Proposition 2.5 . . . 102

3 Estimation of a fast mean reverting jump process with application to spike modeling in electricity prices 105
1 Introduction . . . 106
1.1 Motivation . . . 106
1.2 Main results . . . 108
1.3 Organisation of the paper . . . 111
2 Statistical results . . . 111

2.1 Model assumptions . . . 111
2.2 Estimation of the jumps times and λ_n . . . 113
2.3 Estimator of β_n . . . 114
3 Practical implementation . . . 115
3.1 Choice of the threshold . . . 115
3.2 Numerical illustration . . . 116
3.3 Practical implementation on real data . . . 116
4 Application to electricity spot price . . . 119
4.1 The spot price approach . . . 119
4.2 The forward approach . . . 120
4.3 Change of measure and pricing of options . . . 120
4.4 Application to the French market with real data . . . 122
5 Proofs . . . 123
5.1 Proof of Proposition 3.1 . . . 123
5.2 Estimator of β_n in the case where the jump times and sizes are known . . . 128
5.3 Proof of Theorem 3.1 . . . 145

4 Local polynomial estimation of a doubly stochastic Poisson process 151
1 Introduction . . . 152
1.1 Motivation . . . 152
1.2 Objectives and results . . . 153
2 Statistical setting . . . 154
3 Local polynomial estimation . . . 156
3.1 Method for bandwidth selection . . . 157
3.2 Concentration inequalities . . . 158
3.3 Oracle inequality . . . 159
3.4 Adaptative minimax estimation . . . 160
4 Test for a parametric family . . . 161
5 Dependence between the frequency of electricity spot spikes and temperature . . . 163
5.1 Data . . . 163
5.2 Detection of the jumps . . . 163
5.3 Dependence with temperature . . . 164
6 Numerical results . . . 165
7 Proof of Proposition 4.6 . . . 166
7.1 Proof of Proposition 4.9 . . . 167
7.2 Proof of Proposition 4.10 . . . 171
7.3 Proof of Proposition 4.6 . . . 175
8 Other proofs . . . 177
8.1 Proof of Proposition 4.1 . . . 177
8.2 Proof of Proposition 4.3 . . . 177
8.3 Proof of Proposition 4.7 . . . 179
8.4 Proof of Proposition 4.8 . . . 182

5 A joint model for electricity spot prices and wind penetration with dependence in the extremes **191**
1 Introduction 192
2 Data description and exploratory study 192
2.1 Data description 192
2.2 Exploratory data analysis 193
3 A joint model for the electricity spot price and the wind penetration index 195
3.1 Model for the electricity spot price 196
3.2 Model for the wind penetration index 204
3.3 Dependence modelling 207
4 Application: Impact of the dependence between electricity spot prices and wind penetration on the income of an electricity distributor 208

Introduction

1 Context

1.1 Motivation

This thesis focuses on dependence modeling between continuous time stochastic processes and their statistical estimation for risk management purposes. The different works of this thesis are driven by one application: the modeling of electricity markets. In particular, we are interested in the modeling of the dependence between electricity prices and different risk factors. Electricity being produced using other energy commodities which are traded on financial markets, there exists a dependence between its prices and the other commodity prices. Electricity prices are also strongly related to physical factors (electricity consumption, weather variable like wind, temperature...) which are considered as risk factors. For instance, especially in France, low temperature leads to the use of heating which leads to an increase of the demand and then of the electricity prices. In Germany, high wind leads to an increase of renewable production and then a decrease of electricity prices. These two types of dependence,

1) commodity dependence and

2) physical dependence,

are the main applications to our works.

Taking into account these dependences is necessary and have at least two main interests:

- to better capture the dynamics of the electricity prices and its behavior,
- to have a better quantification of financial risks ;

these two points are of course related. We can for instance take the point of view of an electricity producer which owns a coal plant. If S_t denotes the electricity price at time t, C_t denotes the coal price at time t and K a fixed cost, its incomes at time T can be modeled in a simplified way by $\int_0^T (S_t - HC_t - K)^+ \, dt$ where H is a constant of normalization between electricity prices and coal prices which are not of the same unit. It is then important to have a model that represents well the statistical properties of the two time series S and C but it is also important to model their dependence. The dependence between S and C has an impact on the distribution of the incomes and then on the financial risks of the producer. An other example considered in this thesis is the one of a producer wanting to buy a part Q of the wind production W_t at a fixed price K during a period T. Its incomes at time T are equals to $Q \int_0^T W_t \, (S_t - K) \, dt$ and the dependence between

S and W will have an impact on its distribution. These dependences also have an impact on the way to hedge the different risks.

Continuous time stochastic processes are convenient tools in order to model the dynamics of the prices, especially in a risk management context. Indeed, stochastic calculus is an useful tool for pricing and hedging financial options. Whereas theoretical studies are at the heart of this thesis, operational feasibility can not be neglected:

- proposed models have to be tractable, simulation and pricing of classical options have to be feasible in a reasonable time ;
- estimation problematics have to be considered in a context of discrete time observations.

This context leads us to study in this thesis two types of dependence:

(i) the dependence between two Brownian motions in Chapter 1 and Chapter 2, applied to the dependence modeling between electricity prices and fuel prices,

(ii) the dependence between a point process and a continuous semimartingale in Chapter 4 and Chapter 5, applied to the dependence between spikes frequency appearance of electricity prices and exogenous factors. An estimation procedure for the intensity function depending on an exogenous variable is proposed in Chapter 4.

Before studying the dependence between spikes and exogenous factors, one prerequisite is to propose a model that we can infer for the spike modeling of electricity spot prices and that is adapted for risk management purposes ; this corresponds to Chapter 3 where we propose the estimation of a spike process noised by a continuous semimartingale.

1.2 Description of electricity markets

In order to better understand the modeling, we briefly describe in this section the electricity markets. Electricity markets are local markets, that is there is one market by country and the regulation is different depending on the country. However, equilibrium between consumption and production needs to be satisfied everywhere and we find the same market structure in different countries. There exists three submarkets ordered according to the time horizon.

The intraday market

This market is an over the counter market and corresponds to a maturity less than one day. It insures the security of the system by a balancing mechanism: the different actors can rebalance their offer or demand continuously. We do not not give more details as this market is not considered in this thesis.

The spot market

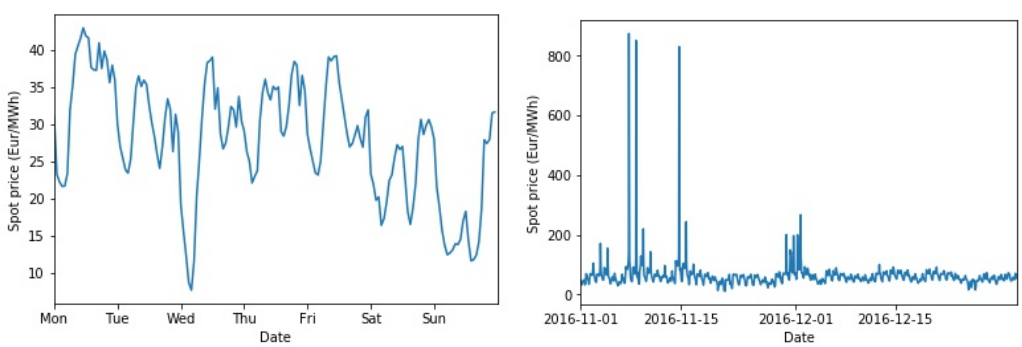
The spot market is a physical market where the electricity is delivered. It is an auction market driven by the equilibrium between demand and supply. The day before delivery, each participant submits a bid curve per hour (or semi-hour). This curve is constructed using merit order: the cheapest mean of production is used first, which is solar and wind production in general. As other energy commodities are used to produce electricity, their price has an impact on this curve and then on spot price ; dependence is then strong. The demand is crossed with the supply for each hour (or semi-hour) giving the spot price.

Remark The spot price is more a day ahead price than a spot price because the price is fixed 24 hours before.

The electricity is not a classical financial asset and presents some particularities. First, it is non storable and can not be traded as any financial asset. Second, prices have a delivery period: the electricity is not delivered instantly but continuously during a period of 1 hour for instance. All these particularities leads to the following features of electricity spot prices time series:

- Seasonality: they exhibit daily, weekly and yearly seasonality. This seasonality is highly related to the one of the consumption due to the non-storability of electricity.
- Spikes: prices jump upward or downward to high values, positive or negative, before to revert quickly to their original level. They can appear when demand is abnormally high or temperature abnormally low or high. In case of high temperature, air conditioning produces these spikes and in case of low temperature, heating is responsible.
- Negative prices: this is a consequence of non storability. If the production is higher than expected, the cost of stopping a production plant may be high and the producer may prefer to pay for consuming electricity. In Germany, unexpected production is caused by the penetration of the renewable energies in the system. For instance, high unexpected wind production may cause negative spikes.
- Mean reversion: mean reversion is present for the spikes where it is very strong but also when the behavior of the price is normal with a lower mean reversion.

Figure 1 shows some of these features on the French market.



(a) *Seasonality (week of the* 04th *of July 2016).*

(b) *Spikes.*

Figure 1: *French spot price illustrations.*

The forward market

The forward market is a classical market with tradable assets and is over the counter. They have a maturity as in classical forward markets but also a delivery period. Contrarily to the spot market, even if there is a delivery period, the electricity is not delivered and the market is then open to everyone. The forward products can be used for hedging purposes but also for speculation.

Different products exist depending on the maturity and the delivery period which can be weeks, months, quarters, seasons or years. If today is the 20th of June 2017, the contract called "July 2017" corresponds to the delivery of electricity continuously during the month of July ; the contract called "Year 2018" corresponds to the delivery of electricity continuously during the year 2018. The first contract is denoted by 1 Month Ahead (1MAH), corresponding to the product starting from the first of next month during one month. In the same way, the second one is denoted by 1 Year Ahead (1YAH). Thus, depending on the date, the 1MAH corresponds to different contracts, same for the 1YAH. The 2MAH corresponds to the product "August 2017". Table 1 gives some examples to understand the nomenclature. Figure 2 represents the 1YAH between January 2011 and March 2017. The decrease of the level of the price is mainly caused by the decrease of the fuel prices.

Product	Contract name	Begin of delivery	End of delivery
1 Month Ahead	July 2017	01/07/2017	31/07/2017
2 Month Ahead	August 2017	01/08/2017	31/08/2017
3 Month Ahead	September 2017	01/09/2017	30/09/2017
1 Quarter Ahead	Q3 2017	01/07/2017	31/09/2017
2 Quarter Ahead	Q4 2017	01/10/2017	31/12/2017
1 Year Ahead	2018	01/01/2018	31/01/2018

Table 1: *Forward products seen from the* 20th *of June 2017.*

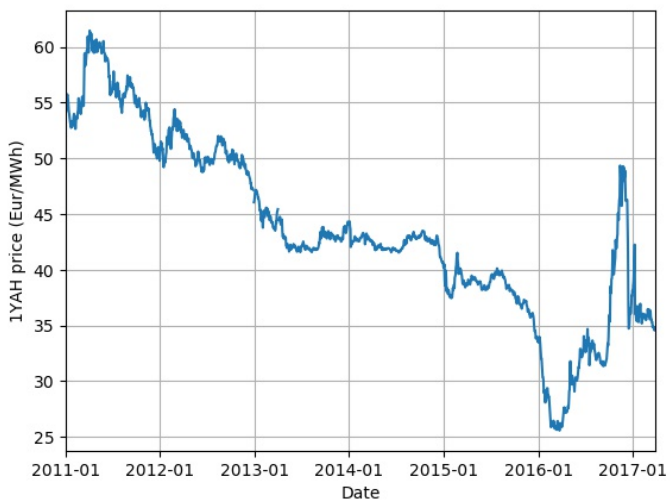


Figure 2: *1 Year Ahead price on the French market between January 2011 and March 2017.*

1.3 Some modeling aspects

Electricity markets modeling is strongly related to the modeling of bond markets in the literature. As for bond markets, one needs to compute the spot price S_t but also the forward curve $f(t,T)$, $t \leq T$ corresponding to the price at time t of the product that delivers at time T 1MWh of electricity during one hour. This product does not exist in reality but is a way to model the

existing product $f(t,T,\theta)$ that delivers 1MWh of electricity between T and $T+\theta$. By absence of arbitrage opportunity, we have

$$f(t,T,\theta) = \frac{1}{\theta}\int_T^{T+\theta} f(t,u)\,du.$$

Joint modeling of spot and forward products prices is necessary because our portfolio is often constituted with all these assets, especially because forward products are a way to hedge financial risks linked to spot price. Two approaches exist, which is also the case in bond markets.

- The first approach consists in modeling directly the spot price, corresponding to the model of the interest rates in bond markets which have first been modeled by Vasicek [57]. In this case, the forward price $f(t,T)$ is equal to $\mathbb{E}^{\mathbb{Q}}(S_T|\mathcal{F}_t)$ where $\mathcal{F}_t$ is the filtration generated by the spot that corresponds to the information provided by the spot until time t and $\mathbb{Q}$ is a risk neutral probability. As the spot is a non tradable asset, the fundamental theorem of asset pricing [31] does not apply and the spot does not need to be martingale even under a risk neutral probability, letting us with some liberty in the modeling. However, the forward products need to be martingales as they are tradable, which can be insured by the martingality of $f(t,T)$; by definition of $f(t,T)$ as a conditional expectation, it is the case. Most common models consist in modeling the spot (or the logarithm of the spot) as the sum of a seasonality function and a multi-factor diffusion part Y_t of the form

  $$Y_t = \sum_{i=1}^{m} w_i Y_t^i,\ dY_t^i = -\beta_i Y_t^i dt + dL_t^i,\ t\in[0,T]$$

  where L^i are Levy processes, often Brownian motions to model the diffusive part and compound Poisson processes to model the spike part. The reader can refer to [20; 9; 44] for more information on this model.

- The second approach is related to the Heath Jarrow Morton approach [32] and consists in modeling directly the forward curve $f(t,T)$. The spot price is then given by $S_t = \lim_{T\to t} f(t,T)$. As said before, the stochastic process $f(t,T)$ is generally modeled by a martingale for each T under $\mathbb{Q}$ in order to have the products $f(t,T,\theta)$ martingales. The most common model consists in modeling the forward curve by

  $$df(t,T) = f(t,T)\left(\sum_{i=1}^{N}\sigma_i(t,T)\,dW_t^i\right)$$

  with $\left(W^i\right)_{i=1..N}$ a multivariate Brownian motion. The reader can refer to the work of Benth and Koekebakker [11] for more information on this model and its applications. In this thesis, for application, we often consider the particular case of the two factor model corresponding to

  $$N=2,\ \sigma_1(t,T) = \sigma_s e^{-\alpha(T-t)},\ \sigma_2(t,T) = \sigma_l.$$

  The dynamics of the curve $T\mapsto f(\cdot,T)$ is driven by two factors: a short term factor $\sigma_s e^{-\alpha(T-t)}dW_t^1$ and a long term factor, $\sigma_l dW_t^2$. The short term factor models the Samuelson effect: the volatility increases when time to maturity decreases. The long term factor models the volatility of long term products: the short term volatility is close to 0 for long term products and we would have a null volatility for them without long term volatility which is not the case.

Concerning the dependence modeling, one approach specific to power prices consists in modeling the electricity spot prices construction (by merit order principle) from fundamental variables (electricity demand, fuel prices, production capacities...). These models are called structural models, see [2; 3; 17; 16] for some examples. The dependence between spot prices and risk factors are in general well captured but computation of forward prices and option prices is costly. We prefer to consider reduced form models, where spot price is modeled by a diffusion. Most common practice is to model the dependence between electricity prices (spot or forward) and other commodity prices (spot or forward) by a correlation matrix between the Brownian motions, see [18] for instance. This is also the case for dependence between temperature and electricity spot prices [12]. More complex structures of dependence in a diffusion modeling context are present in the literature and we present some of them here. In a spot modeling framework, an interesting model for the dependence retaining our attention is the Nakajima and Ohashi model [46] that includes co-integration with other commodity prices in a Brownian motion framework. The electricity spot price revert to a weighted average of all the commodity prices including itself. However, forward prices are hard to compute and do not present long term volatility, implying that products with long term maturity have quasi null volatility. Benth [10] uses time dependent copulae in a discrete time framework in order to model the dependence between electricity spot price and gas spot price. A recent work by Benth proposes to include co-integration [8] in a multi-factor framework with Levy processes ; this model is adapted for both forward modeling and spot modeling.

2 First part: Dependence modeling for Brownian motions

Natural structure of dependence for a multidimensional Brownian motion is the correlation matrix. This correlation matrix corresponds to linear dependence between the Brownian components at each time t and is very easy to manipulate. This structure of dependence is used in most of financial models, the most used being the multivariate Black Scholes model [19]. One of the main application of the multidimensional Brownian motion in finance is pricing and hedging of multi-asset options with payoff $h(S_T)$ where S is a multidimensional diffusion driven by a multidimensional Brownian motion. An important case in two dimensions is $h(x,y) = (x-y-K)^+$ with K a constant, corresponding to a spread option. The price of the option is given by $\mathbb{E}^{\mathbb{Q}}(h(S_T))$ where $\mathbb{Q}$ is a risk neutral probability under which S is martingale. One can see that the price of this option is impacted by the marginal models of each component of S but also by the structure of dependence between the different components, which can not be neglected.

Dependence modeling with correlation matrix presents some limitations and one of them is symmetry. In this part, we only consider the two dimensional case. Let us consider two Brownian motions B^1 and B^2. In the simplified case where volatility of the two assets S^1 and S^2 are the same, the payoff $\left(S_T^1 - S_T^2 - K\right)^+$ of the spread option depends on the distribution of $B_T^1 - B_T^2$, as both assets are martingale under $\mathbb{Q}$ and have null drifts. Our main interest is then the distribution of $B^1 - B^2$. Let us suppose first that their dependence is modeled by a correlation, that is $d\langle B_t^1, B_t^2\rangle = \rho dt$. In this case, we have for $x \in \mathbb{R}$ and $t > 0$,

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \mathbb{P}\left(B_t^1 - B_t^2 \leq -x\right)$$

implying for $x \geq 0$,

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) \leq \frac{1}{2}.$$

The distribution of $B_t^1 - B_t^2$ is then symmetric. This symmetry is due to the symmetric structure of dependence between B_t^1 and B_t^2 at a given time t, corresponding to the copula of $\left(B_t^1, B_t^2\right)$. Let us recall that a (two dimensional) copula is a two increasing function $C : [0,1]^2 \mapsto [0,1]$ having uniform marginal distributions ; the reader can refer to [47] for more information on copulae. Sklar's theorem [54] states that for any two dimensional random variable (X, Y), its distribution is characterized by the marginal distribution of X, the marginal distribution of Y and a copula C that represents the structure of dependence. In the correlation case, at time t, the copula of $\left(B_t^1, B_t^2\right)$ is called a Gaussian copula and is symmetric, that is $C(u,v) = C(v,u)$ for $u,\ v \in [0,1]$. In a more general way, if X and Y have the same distribution function and their copula is symmetric, we prove that the distribution $X - Y$ is symmetric. As the marginal distribution of B_t^1 and B_t^2 are known, the control of the distribution of $B_t^1 - B_t^2$ only depends on the copula of $\left(B_t^1, B_t^2\right)$. In order to have higher values for the quantity $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ for $x \geq 0$ than $\frac{1}{2}$, one has to consider asymmetric copulae, which are not easy to construct, see [40] or [59] for instance. However, copula is a natural tool for random variables but not from stochastic processes. Classical asymmetric copulae might not be adapted for Brownian motions. This leads to the following questions:

Question 1 *Are there asymmetric and admissible copulae for modeling the dependence between two Brownian motions ?*

Question 2 *Are there admissible copulae allowing for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ to have higher values than in the correlated case ?*

A first step to answer Question 1 is to give a suitable definition for admissible copulae for Brownian motions. Literature about copulae for stochastic processes is not large and can be divided in three different topics. First one concerns copulae for discrete time stochastic processes. These copulae have been introduced by Patton [48] and generalized by Fermanian [27]. At each time t, the copula is constructed conditionally on what happens before time $t-1$. In a continuous time framework, [25] considers the copula between X_s and X_t where X is a stochastic process and gives some conditions on this copula for X to be a Markovian process. The last framework is the one we consider and consists in studying the copula between two stochastic processes, and in particular Brownian motions, for each time t [13; 24; 53; 14; 35]. In [14] and [35], the notion of admissible copula is linked to a local correlation function between two Markovian diffusions. In [24], those results are generalized in dimension n. The Markovian framework seems to be the natural framework for Brownian motions and is also considered in [13]. We propose a definition similar to the one of [13] that includes the work of [14], [35] and [24].

Definition (Admissible copula for Markovian diffusions) We say that a collection of copulae $C = (C_t)_{t \geq 0}$ is an admissible copula for the n real valued Markovian diffusions, $n \geq 2$, $\left(X^i\right)_{1 \leq i \leq n}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if there exists a $\mathbb{R}^m$ Markovian diffusion $Z = \left(Z^i\right)_{1 \leq i \leq m}$, $m \geq n$, defined on a probability extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\begin{cases} \mathcal{L}\left(Z^i\right) = \mathcal{L}\left(X^i\right), 1 \leq i \leq n, \\ Z_0^i = X_0^i, 1 \leq i \leq n, \\ \text{for } t \geq 0, \text{ the copula of } \left(Z_t^i\right)_{1 \leq i \leq n} \text{ is } C_t. \end{cases}$$

where $\mathcal{L}(Y)$ is the infinitesimal generator of a Markovian process Y.

This definition includes, in the case of Brownian motions, deterministic correlation, local correlation but also stochastic correlation if this one is a Markov process as it is possible to have $m \geq n$. In

the local correlation model, in the two dimensional case, Jaworski and Krzywda [35] prove that a copula is admissible for Brownian motions if

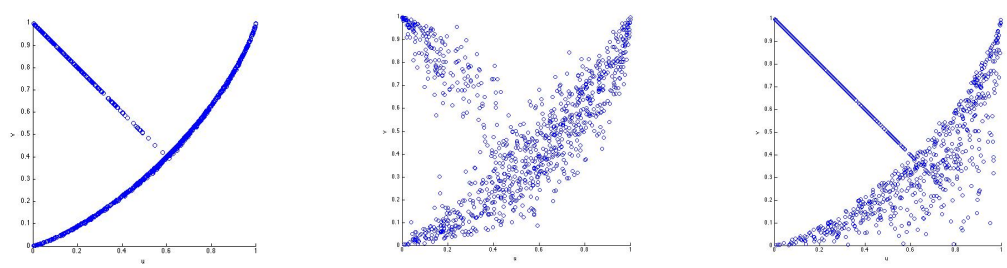
$$\left|\frac{1}{2}e^{\frac{\Phi^{-1}(v)^2-\Phi^{-1}(u)^2}{2}}\frac{\partial^2_{u,u}C(u,v)}{\partial^2_{u,v}C(u,v)}+\frac{1}{2}e^{\frac{\Phi^{-1}(u)^2-\Phi^{-1}(v)^2}{2}}\frac{\partial^2_{v,v}C(u,v)}{\partial^2_{u,v}C(u,v)}\right|<1 \quad \forall(t,u,v)\in\mathbb{R}^+\times[0,1]^2$$

when the copula does not depend on time with Φ the cumulative distribution function of a standard normal random variable. In particular, they prove that the extension of the Farlie-Gumbel-Morgenstern copula in a dynamical framework $C_t(u,v)=uv(1+a(1-u)(1-v)),\ t\geq 0$ with $a\in[-1,1]$ is admissible for Brownian motions. However, this copula is not asymmetric.

Let us consider a Brownian motion B^1 defined on a filtered probability space $(\Omega,\ \mathcal{F},\ (\mathcal{F}_t)_{t\geq 0},\ \mathbb{P})$ with $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual hypothesis (right continuity and completion) with B^1 adapted to $(\mathcal{F}_t)_{t\geq 0}$. While the correlation is the standard way to construct a dependent Brownian motion B^2 from B^1, an other way existing in the literature is to consider the reflection of B^1 on $x=h$ with $h\in\mathbb{R}$ which is also a $\mathcal{F}$ Brownian motion according to the reflection principle (see [36, Theorem 3.1.1.2, p. 137]). The reflected Brownian motion $\tilde{B}^h$ is defined by $\tilde{B}^h_t=-B^1_t+2(B^1_t-B_{\tau^h})\mathbf{1}_{t\geq\tau^h}$ with $\tau^h=\inf\{t\geq 0: B^1_t=h\}$. Let us consider $M(u,v)=\min(u,v)$ and $W(u,v)=\max(u+v-1,0)$ for $u,\ v\in[0,1]$. The copula of $(B^1,\tilde{B}^h)$, $\left(C^{ref,h}_t\right)_{t\geq 0}$, is defined by

$$C^{ref,h}_t(u,v)=\begin{cases} v & \text{if} \quad \Phi^{-1}(u)-\Phi^{-1}(v)\geq\frac{2h}{\sqrt{t}} \\ W(u,v)+\Phi\left(\Phi^{-1}(M(u,1-v))-\frac{2h}{\sqrt{t}}\right) & \text{if} \quad \Phi^{-1}(u)-\Phi^{-1}(v)<\frac{2h}{\sqrt{t}}\end{cases}$$

and is admissible for Brownian motions. An illustration is given in Figure 3a. One way to construct new copulae from this copula is to consider the dependence between B^1 and a correlated Brownian motion to the reflection of B^1, see Figure 3b. An other way is to use a random variable for the barrier, see Figure 3c where the barrier is equal to $h+\mathcal{E}$ with $\mathcal{E}$ following an exponential law.



(a) (b) (c)

Figure 3: *The Reflection Brownian Copula $C^{ref,h}$ and some of its extensions at time $t=1$ with $h=2$. Figure 3a is the Reflection Brownian Copula. Figure 3b is the extension considering a Brownian motion correlated to the reflection of the first Brownian with a correlation $\rho=0.95$. Figure 3c is the extension in the case of a random barrier equals to the sum of h and an exponential random variable with parameter $\lambda=2$.*

Result 1 *The copula $C^{ref,h}$ and its extensions are admissible copulae for Brownian motions and are asymmetric.*

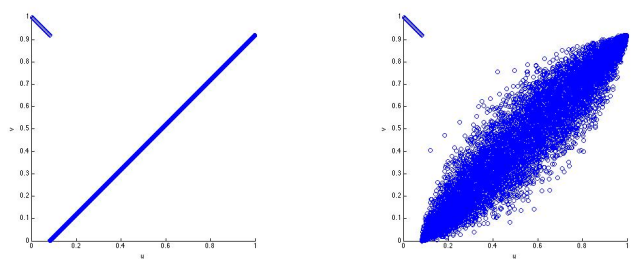
Before answering Question 2, let us look to the range of values achievable for $\mathbb{P}(X - Y \geq x)$ when X and Y are two standard normal random variables, corresponding to the static case of our problem.

Result 2 *The range of values achievable for $\mathbb{P}(X - Y \geq x)$ is $\left[0, \Phi\left(\frac{-x}{2}\right)\right]$ if we restrict the set of copulae to Gaussian copulae and to $\left[0, 2\Phi\left(\frac{-x}{2}\right)\right]$ otherwise.*

The supremum bound $2\Phi\left(\frac{-x}{2}\right)$ is achieved by the copula

$$C^r(u,v) = \begin{cases} M(u-1+r, v) & \text{if } (u,v) \in [1-r,1] \times [0,r], \\ W(u,v) & \text{if } (u,v) \in [0,1]^2 \setminus ([1-r,1] \times [0,r]) \end{cases}$$

with $r = 2\Phi\left(\frac{-\eta}{2}\right)$, see Figure 4a for illustration. The copula presents two states of dependence: the first one corresponds to the countermonotonic copula W, equivalent to a correlation of -1, in the upper left part of the unit square and the second one corresponds to the comonotonic copula M, equivalent to a correlation of 1. The result follows from [47, Section 6.1] and [29; 52; 41] where finding achievable bounds on $\mathbb{P}(X + Y > x)$ is considered. The range of values between $\left[\Phi\left(\frac{-x}{2}\right), 2\Phi\left(\frac{-x}{2}\right)\right]$ is achieved by considering the copula with a relaxed correlation in the countermonotonic part, see Figure 4.



(a) $\rho = 1$. (b) $\rho = 0.95$.

Figure 4: *Patchwork copula $C^r(u,v)$ presenting two states depending on the value of u: the first copula is in the upper left part of the plan and is W ; the second one is the Gaussian copula with correlation equal to ρ, with $\rho = 1$ that is the degenerated copula M or $\rho = 0.95$. r is equal to $2\Phi\left(\frac{-\eta}{2}\right)$ with $\eta = 0.2$.*

Result 3 is the same than Result 2 but in a dynamical framework and gives an answer to Question 2.

Result 3 *The range of values achievable for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ is $\left[0, \Phi\left(\frac{-x}{2\sqrt{t}}\right)\right]$ if we restrict the set of copulae to correlation dependence and to $\left[0, 2\Phi\left(\frac{-x}{2\sqrt{t}}\right)\right]$ if we consider all admissible copulae for Brownian motions. These values can be achieved using the Reflection Brownian Copula.*

The range of values that can be achieved is the same than in the static framework with random variables. An interesting result is that the Reflection Brownian Copula that achieves these values has also two states of correlation: one with correlation equal to 1 and one with correlation equal to -1, as for the copula achieving the upper bound in the static case.

Application to energy commodity prices modeling Spread options are common on energy commodity markets. Let us recall the problem of the producer owning a coal plant with incomes modeled by $(S_t - HC_t - K)^+$. S and C are modeled by diffusions driven by Brownian motions and dependence between those Brownian motions is usually modeled by a correlation, implying symmetry in the distribution of $S_t - HC_t$. Nevertheless, coal is a fuel for electricity and HC_t is more likely to be lower than S_t, which can not be done with a correlation. As marginal models are satisfying, we only want to change the structure dependence between electricity prices and coal prices, corresponding to the copulae between Brownian motions. Results of this part has shown that to capture asymmetry and higher value for the survival function of the difference of two Brownian motions, one has to consider two states of correlation, one negative and one positive, instead of one. This leads to the two following models:

- A multi-barrier model based on the copula of a Brownian motion and its reflection: we define two barriers ν and η with $\nu < \eta$ and we consider two independent Brownian motions X and B^Y. We construct the Brownian motion Y^n that is correlated to $\tilde{X}^n$: $Y^n = \rho\tilde{X}^n + \sqrt{1-\rho^2}B^Y$, with $\tilde{X}^n$ the Brownian motion equal to $-X$ at the beginning and reflecting when $X - Y^n$ hits a two-state barrier equal to η before the first reflection and switching from η to ν or from ν to η at each reflection.

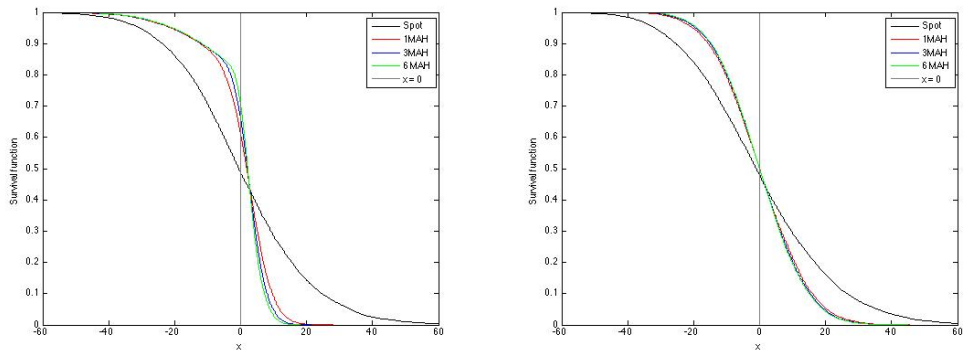
- A local correlation model with two states of correlation: the local correlation function is chosen such that it is Lipschitz and equal to $\rho_1 < 0$ when $x - y \leq \nu$ and to $\rho_2 > 0$ when $x - y \geq \eta$ with $\eta > \nu$.

These two models seem to be equivalent. Concerning the first model, we derive a closed formula for the cumulative distribution function of the difference between the two Brownian motions. While the second one is more easy to use and understand than the first one, it does not give a closed formula for the cumulative distribution function. Higher values than $\frac{1}{2}$ are achieved for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ where $x \geq 0$ in both model.

Each commodity price is modeled by a two factor diffusion

$$df^i(t,T) = f^i(t,T)\left(\sigma_s^i e^{-\alpha^i(T-t)}dW_t^{s,i} + \sigma_l^i dW_t^{l,i}\right),\ i = \{\text{Electricity, Coal}\}.$$

The parameters are estimated on forward markets of electricity and of coal during 2014 in France with daily observations using the method of [28]. We consider the benchmark model corresponding to dependence modeled by a correlation matrix between the four Brownian motions and the multi-barrier model where the dependence between the two long term Brownian factors are modeled using the multi-barrier model. Modeling the two short term factors using the multi-barrier model has no impact because short term volatilities are too different. Figure 5 corresponds to the survival function of the difference between products Spot, 1MAH, 3MAH and 6MAH of each commodity for the two models of dependence. Survival function in the multi-barrier model takes higher values than in the benchmark model for $x \geq 0$ for long term products, which is consistent with the fact that the dependence is changed between the long term Brownian motions. Results are the same considering the local correlation model.



(a) *Multi-barrier model.* (b) *Benchmark model.*

Figure 5: *Empirical survival function of the difference between the price of electricity and the price of coal at time $t = 365$ days estimated with 10000 simulations with a time step of $\frac{1}{24}$ days for different products (Spot, 1MAH, 3MAH, 6MAH) in the multi-barrier model and in the benchmark model.*

One of the main issue of this model is calibration and the choice of the values of the barrier, that must depends on the initial value of $f^{\text{Electricity}}(0,T) - Hf^{\text{Coal}}(0,T)$ if we want to have more impact on the survival function of $f^{\text{Electricity}}(t,T) - Hf^{\text{Coal}}(t,T)$.

3 Second part: Inference of a spike process in high frequency statistics

A spike is defined as a jump, positive or negative, coming back to 0 in a short period of time. A natural stochastic model for it is a mean reverting jump process with strong mean reversion, see Figure 6. Now, let us consider a stochastic process X defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P}\right)$ of the form

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + Z_t^\beta,\ t \geq 0$$

with

$$Z_t^\beta = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}(dt, dx),$$

W a standard Brownian motion, μ and σ two adapted cádlág processes and $\underline{p}$ a Poisson measure on $\mathbb{R}^+ \times \mathbb{R}$ independent from W with compensator $\underline{q} = \lambda dt \otimes \nu(dx)$. X is then the sum of a continuous Itô semimartingale and a mean reverting jump process corresponding to the spike process.
X is observed on a regular grid $\mathcal{M} = \{t_i = i\Delta_n, 0 \leq i \leq \lfloor \frac{T}{\Delta_n} \rfloor\}$ with $\Delta_n = \frac{T}{n}$. We assume that $\Delta_n \to 0$ with T fixed, corresponding to a high frequency framework with fixed time horizon. Our objective is to estimate the parameters of the spike process Z^β and especially the parameter β corresponding to the speed of the mean reversion. If we do not add any further assumptions,

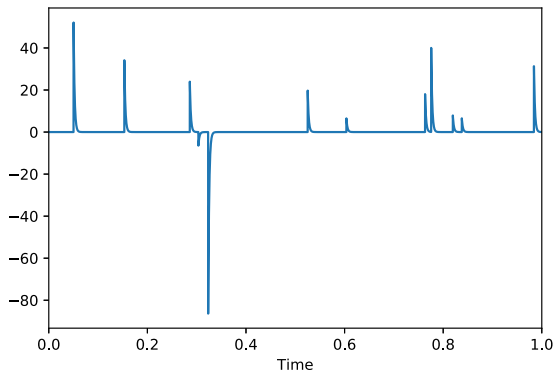
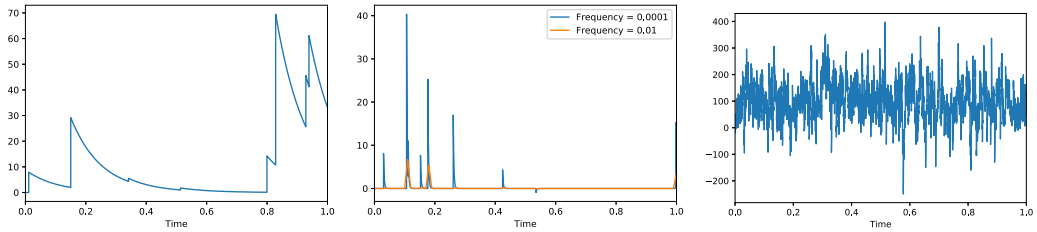


Figure 6: *Illustration of a spike process.*

the parameter β is a drift parameter and is not identifiable if T is fixed ; one can refer to Aït-Sahalia and Jacod [6] for the non identifiability of the drift. Nevertheless, the case in which we are interested is when the mean reversion is strong, corresponding to a spike mode. An illustration is given in Figure 7a: if β is too low compared to Δ_n, we do not observe the spike effect and the process does not even revert to 0 before T. To model this spike effect, we need to add the assumption $\beta = \beta_n \to \infty$. One has to work under assumption $\beta_n \Delta_n \lesssim 1$ in order to observe all the spikes: in the other case, a spike can happen and revert to 0 in a period of Δ_n and it is not possible to observe it, see Figure 7b. An non divergence assumption is also needed in the case $\lambda = \lambda_n$, which is $\frac{\lambda_n}{\beta_n} \lesssim 1$: in the other case, the average number of jumps λ_n is stronger than the speed of mean reversion β_n and the process diverges as illustrated in Figure 7c. An other condition which is classical is $\lambda_n^2 \Delta_n \to 0$, meaning that there is at most one jump in an interval of size Δ_n. In this framework, there are no results about the identifiability or non identifiability of β, raising the following questions.



(a) *Low mean reverting.* (b) *Frequency too high.* (c) *Number of jumps too high.*

Figure 7: *Spikes processes in non considered regimes.*

Question 1 *How to identify the jump times and sizes of the spike process in this new framework ?*

Question 2 *How to estimate the parameter β_n if it is possible and what is the error of estimation ?*

Jump detection in a jump diffusion model with high frequency observations is an important question

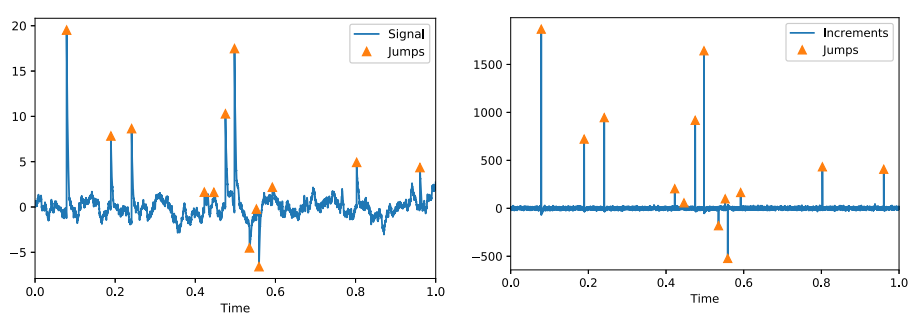
which has been studied by many authors in the literature. One of the main applications is the estimation of volatility in presence of a jump component. The reader can refer to the works of Mancini [42; 43], Aït and Jacod [5; 6] or Lee and Mykland [38]. In all these works, the main idea consists in looking to the size of the increments $\Delta_i^n X = X_{t_i} - X_{t_{i-1}}$. This increment can be written as the sum of the drift increment, the Brownian one and the jump one. As the drift is absolutely continuous with respect to the Lebesgue measure, its increment is of order Δ_n. The Brownian increment is of order $\sqrt{\Delta_n}$. The jump increment is of size 1 when there is a jump between times t_{i-1} and t_i and is equal to 0 when there is no jump in this interval. Thus considering a threshold $v_n \asymp \Delta_n^{-\varpi}$ with $\varpi \in \left(0, \frac{1}{2}\right)$, the quantity $\frac{|\Delta_i^n X|}{\sqrt{\Delta_n} v_n}$ converges to 0 in the absence of jumps but to ∞ in the presence of it.

In our case, the problem is slightly different because there is an extra term caused by the mean reversion after a jump of order $\beta_n \Delta_n$. In the case $\beta_n \Delta_n \to 0$, it is still possible to distinguish the jump increment from the mean reversion increment, see Figure 8. In the case where $\beta_n \Delta_n \asymp 1$, it is not possible to distinguish a mean reversion increment from a jump increment as they have the same size, see Figure 9. However, we can see that after a jump, the next increment which is the mean reversion one has an opposite sign. One can also show that, under suitable conditions, after a mean reversion increment which is over the threshold $v_n \sqrt{\Delta_n}$, if there are not too many jumps, the next increment is of the same sign than the mean reversion one. The following strategy is then adopted: let $\mathcal{I}_n(1) < ... < \mathcal{I}_n\left(\hat{\lambda}_n\right)$ the indices $i \in \{1, ..., n-1\}$ such that

- $|\Delta_i^n X| > v_n \sqrt{\Delta_n}$ if $\beta_n \Delta_n \to 0$,
- $|\Delta_i^n X| > v_n \sqrt{\Delta_n},\ \Delta_i^n X \Delta_{i+1}^n X < 0$ if $\beta_n \Delta_n \asymp 1$.

We have the following result:

Result 1 *With probability converging to 1,* $\hat{\lambda}_n = N_1$ *and* $T_q \in \left(\left(\mathcal{I}_n(q) - 1\right)\Delta_n,\ \mathcal{I}_n(q)\Delta_n\right]$ *with* T_q *the jump times.*

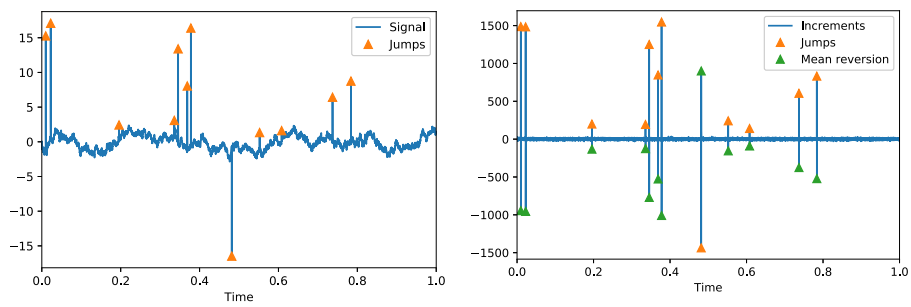


(a) X.

(b) $\frac{\Delta_i^n X}{\Delta_n^{0.49}}$.

Figure 8: *Jump detection in the case* $\beta_n \Delta_n = 0.3$.

Once the spike times are identified, a natural estimator of their size is $\Delta_{\mathcal{I}_n(q)}^n X$ for $q \in \{1, .., \hat{\lambda}_n\}$ which is equal to $\Delta X_{T_q} e^{-\beta_n (T_q - \mathcal{I}_n(q)\Delta_n)}$ plus an error term with ΔX_{T_q} the size of the qth jump.



(a) X. (b) $\frac{\Delta_i^n X}{\Delta_n^{0.49}}$.

Figure 9: *Jump detection in the case* $\beta_n \Delta_n = 1$.

The term $e^{-\beta_n (T_q - \mathcal{I}_n(q)\Delta_n)}$ includes a bias caused by the fact that the jump happens between times $(\mathcal{I}_n(q) - 1)\Delta_n$ and $\mathcal{I}_n(q)\Delta_n$ and has already started to mean revert at the moment of observation. If $\beta_n \Delta_n \to 0$, this term is equal to $1 + O(\beta_n \Delta_n)$ and it is possible to identify the size of the jump. If $\beta_n \Delta_n \asymp 1$, it is not possible to identify the jump size because we do not know exactly the jump time. However, if $\lambda_n \to \infty$, it is possible to average this error and we have the following result:

$$\frac{m\beta_n \Delta_n}{(1 - e^{-m\beta_n \Delta_n})\hat{\lambda}_n} \sum_{q=1}^{\hat{\lambda}_n} \left(\Delta^n_{\mathcal{I}_n(q)} X\right)^m \to \int_{\mathbb{R}} x^m \nu(dx)$$

for every integer $m > 0$ such that $\int_{\mathbb{R}} x^m \nu(dx) < \infty$. This estimator differs from the classical estimator $\frac{1}{\hat{\lambda}_n} \sum_{q=1}^{\hat{\lambda}_n} \left(\Delta^n_{\mathcal{I}_n(q)} X\right)^m$ by the term $\frac{m\beta_n \Delta_n}{(1 - e^{-m\beta_n \Delta_n})}$: this correction corresponds to the average error of the bias caused by the mean reversion. To estimate the moments of the jump sizes, one needs to have a consistent estimator of β_n.

To estimate the parameter β_n, we consider the slope of the process after a jump

$$-\mathrm{sgn}\left(\Delta^n_{\mathcal{I}_n(q)}\right) \Delta^n_{\mathcal{I}_n(q)+1} X$$

which is of order $\left(1 - e^{-\beta_n \Delta_n}\right) |\Delta X_{T_q}|$ where sgn is the sign function. Averaging these quantities over all the jumps allows to average the noise caused by the Brownian motion. Dividing by an approximation of $\sum_{q=1}^{\hat{\lambda}_n} |\Delta X_{T_q}|$, and taking a logarithm transformation of the average, we obtain the following estimator

$$\widehat{\beta}_n = \frac{-1}{\Delta_n} \log\left(\left(1 + \frac{\sum_{q=1}^{\hat{\lambda}_n} \mathrm{sgn}(\Delta^n_{\mathcal{I}_n(q)} X)\left(\Delta^n_{\mathcal{I}_n(q)+1} X + 2\Delta_n \sum_{j=1}^{q-1} \Delta^n_{\mathcal{I}_n(j)} X\right)}{\sum_{q=1}^{\hat{\lambda}_n} |\Delta^n_{\mathcal{I}_n(q)} X|} \mathbf{1}_{\hat{\lambda}_n > 0}\right) \vee \Delta_n\right).$$

where a correction term $2\Delta_n \sum_{j=1}^{q-1} \Delta^n_{\mathcal{I}_n(j)} X$ is added in order to avoid a bias term of order $\frac{\lambda_n}{\beta_n}$.

Result 2 *The error* $\frac{\widehat{\beta}_n - \beta_n}{\beta_n}$ *is equal to* $O_p\left(\lambda_n \Delta_n + \min\left(\frac{\lambda_n}{\beta_n}, \beta_n^{-\frac{1}{2}}\right) + \left(\beta_n \sqrt{\lambda_n \Delta_n}\right)^{-1}\right)$ *and a central limit theorem is given under some asymptotic assumptions.*

The first error term is a bias term, the second one is due to noise caused by previous jumps that have not revert to 0 entirely and the third one is caused by the Brownian motion. In order for the estimator to be consistent, we need to have the Brownian motion error $\left(\beta_n\sqrt{\lambda_n\Delta_n}\right)^{-1} \to 0$ as the two first errors terms converge to 0. This condition can be explained by the fact that the size of the noise due to an increment of the Brownian motion is of order $\sqrt{\Delta_n}$, the one due to the average of increments happening after a jump is then of order $\sqrt{\frac{\Delta_n}{\lambda_n}}$ and the size of the mean reversion if of order $\beta_n\Delta_n$. For the noise to be negligible compared to the estimator which is the slope, it is needed to have $\sqrt{\frac{\Delta_n}{\lambda_n}} = o\left(\beta_n\Delta_n\right)$, corresponding to our condition.

Application to spike modeling in electricity time series Let us generalize the two factor model in order to have a spike component. We model the forward price $f\left(t,T\right)$ under the historical probability by

$$f\left(t,T\right) = \int_0^t \mu_s ds + f^c\left(t,T\right) + \int_0^t \int_{\mathbb{R}} x e^{-\beta(T-s)} \underline{p}\left(ds,dx\right)$$

where

$$df^c\left(t,T\right) = f^c\left(t,T\right)\left(\sigma_s e^{-\alpha(T-t)} dW_s + \sigma_l dW_l\right)$$

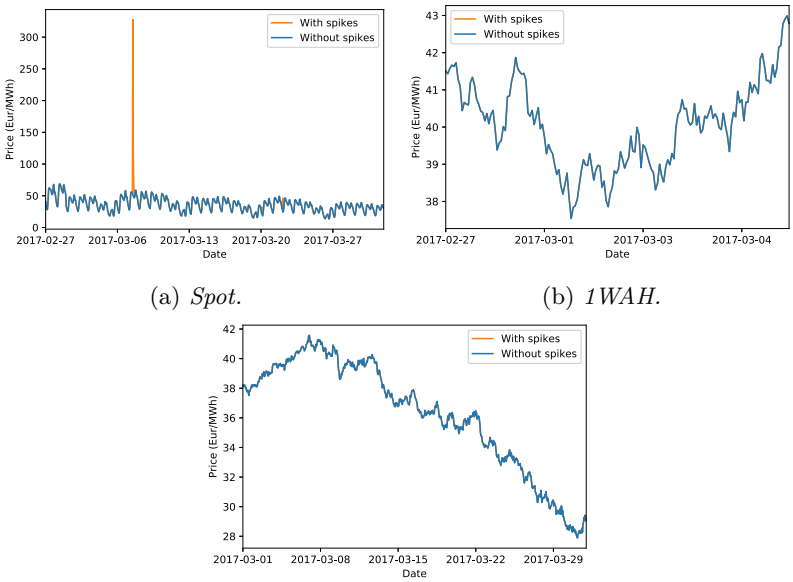
corresponds to the classical forward dynamics. The spot price is then equal to

$$S_t = \int_0^t \mu_s ds + S_t^c + \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}\left(ds,dx\right)$$

where S^c is the equivalent spot model of the two factor model and is a semimartingale. We have established a simple model on the forward and the spot, that differs only slightly from the classical models by adding a spike component. We apply our estimation procedure on French, German and Australian spot prices to estimate the spike component. When we compute the forward products $f\left(t,T,\theta\right)$, the component due to f^β is of order $\frac{\lambda}{\beta}$ which is negligible compared to the continuous part of the forward. The spot factor has not impact on the forward prices, which is consistent with the data. The parameters of the continuous part of the model can then be estimated on the forward products as there were no spikes. The continuous part is calibrated on the French forward products. Figure 10 corresponds to one simulation of the models with and without spikes using estimated parameters on French market and shows that the spike factor has no impact on forward products prices but only on the spot. We also show that spike modeling has also a strong impact on strips of call options pricing of the form $\int_0^T \left(S_t - K\right)^+ dt$.

4 Third part: Non parametric estimation of the intensity of a doubly stochastic Poisson process depending on a covariable

In this part, we are interested in a continuous semimartingale X and in a doubly stochastic Poisson process N defined on the common filtered probability space $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_t\right)_{0\le t\le T}, \mathbb{P}\right)$. The law of the doubly stochastic Poisson process is entirely determined by its intensity function λ which is also a stochastic process. Modeling the dependence between N and X is then the same as modeling



(a) *Spot.* (b) *1WAH.*

(c) *1MAH.*

Figure 10: *Simulation of different products in a two factor model with and without spikes between the* 27th *of February 2017 and the* 31st *of March 2017. We illustrate the spot, the 1WAH starting the* 27th *of February 2017 and the 1MAH starting the* 01st *of March 2017.*

the dependence between λ and X. We assume that X and N are observed continuously on a time horizon $[0, T]$. We assume that

$$\lambda_s = nq\left(X_s\right), \; s \in [0, T]$$

where $n \in \mathbb{N}$, $n \geq 1$ corresponds to the asymptotic. Conditionally on $(X_t)_{0 \leq t \leq T}$, N is an inhomogeneous Poisson process with intensity at time t $nq\left(X_t\right)$. Our objective is to estimate the function q on a given interval I of $\mathbb{R}$. The literature on non parametric estimation of intensity function for Poisson processes is large. The simplest case corresponds to the inhomogeneous Poisson process where the intensity is a deterministic function of the time: [50; 51] use model selection techniques and projection estimators in a non asymptotic framework. A penalization function is proposed in oder to select the optimal model. [26; 15] use kernel estimators in an asymptotic framework ; in [15], a method to select the bandwidth is proposed. Most used doubly stochastic Poisson process models are the one of Aalen and Cox [1; 23]. In the Aalen model, the intensity function is of the form $\alpha_t Y_t$ with α_t a function of time and Y_t a stochastic process. In the Cox model, it is of the form $\alpha_t \exp\left(\beta^T Z\right)$ with Z a multi-dimensional random variable or stochastic process in some cases (see [45] for instance). Again, projection estimators are used by [22] and local polynomial estimators which are generalization of kernel estimators are used by [21] for those models. In [60], the intensity of a doubly stochastic Poisson process is inferred as a function of time using kernel methods in an asymptotic framework. To our knowledge, non parametric methods of estimation in our framework is less common in the literature except for [56] that proposes a kernel estimator of the function q in the case where T goes to ∞ and when X satisfies some conditions, which can be for instance stationarity. We want to work in a more general framework when X does not need

to be stationary.

First, in order to estimate the function q at some point $x \in I$, we need for X to be close to x a certain amount of time before T. One way to evaluate the time spend by X around x when X is a semimartingale is the local time l_T^x. We consider the natural local time of X being the measure verifying the occupation time formula

$$\int_0^t f(X_s)\, ds = \int_{\mathbb{R}} f(x)\, l_t^x dx,\ 0 \leq t \leq T$$

for any measurable function f on $\Omega \times \mathbb{R}$. It differs from the classical local time used in the literature where the integration of the left hand side is with respect to $d\langle X \rangle_s$ [49, Chapter 6] but both are linked. l_T^x can also be defined by

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^T \mathbf{1}_{|X_s - x| \leq \epsilon} ds.$$

If we write $X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$, a sufficient condition for the existence of l_T^x is $\inf_{s \in [0,T]} \sigma_s \geq \underline{\sigma}$ almost surely with $\underline{\sigma} > 0$ a constant. Adding the assumption $\mathbb{E}\left(\int_0^T |\mu_s| ds + \sup_{0 \leq t \leq T} |\int_0^T \sigma_s dW_s|\right) < \infty$, we have $\mathbb{E}\left(\sup_{x \in \mathbb{R}} l_T^x\right) < \infty$ which is needed in this part. All these results can be derived easily using [49, Exercise 1.15] and [7, Equation $(\mathrm{III})_\gamma$]. We also consider the degenerated case $X_t = t$ corresponding to the inhomogeneous Poisson process. We then work under one of the following assumptions:

(i) $\inf_{0 \leq s \leq T} \sigma_s \geq \underline{\sigma}$ with $\underline{\sigma} > 0$ a deterministic constant and

$$\mathbb{E}\left(\int_0^T |\mu_s| ds + \sup_{0 \leq t \leq T} |\int_0^T \sigma_s dW_s|\right) < \infty,$$

(ii) $X_t = t$ for all t in $[0, T]$.

In order to estimate q at point $x \in I$, we then need to have $l_T^x > 0$. We then choose to work conditionally on the event $D(I, \nu)$ defined by

$$D(I, \nu) = \{\omega \in \Omega,\ \inf_{x \in I} l_T^x(\omega) \geq \frac{\nu T}{|I|}\}$$

with $\nu \in (0, 1]$. This framework is the same as the one of [33] in the context of non parametric estimation of the volatility function $\sigma(Y)$ of a diffusion Y.

Let K be a positive kernel function with bounded support $[-1, 1]$, $\|K\|_\infty = \sup_{x \in \mathbb{R}} K(x) < \infty$, $K_h(x) = h^{-1} K\left(\frac{x}{h}\right)$ for $x \in \mathbb{R}$ and $h > 0$ the bandwidth parameter. We consider the local polynomial estimator of q with degree m for $h > 0$ and $x \in \mathbb{R}$

$$\hat{q}_h(x) = \frac{1}{n} \int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dN_s$$

with

$$U(x) = \left(1, x, \frac{x^2}{2!}, .., \frac{x^m}{m!}\right)^T, \; w(x,h,z) = U^T(0) B(x,h)^{-1} U(z) \mathbf{1}_{B(x,h)\in\mathcal{S}_{m+1}^+}, \; z \in \mathbb{R}$$

and

$$B(x,h) = \int_0^T U\left(\frac{X_s - x}{h}\right) U^T\left(\frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s \in I} ds.$$

If $B(x,h) \in \mathcal{S}_{m+1}^+$, this estimator is equal to $U^T(0)\hat{\theta}_h(x)$ with

$$\hat{\theta}_h(x) = \underset{\theta\in\mathbb{R}^{m+1}}{argmin} \; -\frac{2}{n}\theta^T \int_0^T U\left(\frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s\in I} dN_s$$
$$+ \theta^T \int_0^T U\left(\frac{X_s - x}{h}\right) U^T\left(\frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s\in I} ds\theta.$$

We denote by q_h the conditional expectation of $\hat{q}_h$ given X:

$$q_h(x) = \int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s\in I} q(X_s) ds.$$

On the event $D(I,\nu)$, if there exists $\Delta > 0$ and $K_{\min} > 0$ such that $K(x) \geq K_{\min}\mathbf{1}_{|x|\leq\Delta}$ for $x \in \mathbb{R}$, we prove that $B(x,h) \in \mathcal{S}_{m+1}^+$ for $x \in I$ and $0 < h \leq \frac{2}{3}\Delta|I|$.

In order to evaluate the performance of our estimator, we choose to work with the integrated quadratic loss on I, conditionally on $D(I,\nu)$, that is with the quantity $\mathbb{E}\left(\|q - \hat{q}_h\|_I^2 | D(I,\nu)\right)$ where $\|f\|_I^2$ is equal to $\int_I f(x)^2 dx$. We want to answer the following questions:

Question 1 *How to choose the bandwidth parameter h in an optimal way ?*

Question 2 *What is the speed of convergence of our estimator and is it optimal ?*

Question 3 *Is the function q belongs to some parametric family ?*

Question 1 is central because the bandwidth parameter has an impact on the quality of our estimator and Question 2 will assess its quality in terms of speed convergence. Question 3 has operational purposes: it is easier in terms of comprehension and modeling to work with a parametric function than a non parametric one.

The loss function can be written as the sum of a bias term $\mathbb{E}\left(\|q - q_h\|_I^2 | D(I,\nu)\right)$ decreasing with h and a variance term $\mathbb{E}\left(\|q_h - \hat{q}_h\|_I^2 | D(I,\nu)\right)$ increasing with h. The bandwidth h that minimizes the sum of the bias and the variance is then optimal in the sense that the loss function is minimized. However, this optimal bandwidth depends on q that we do not know and is called the oracle bandwidth. One wants to choose a bandwidth h such that the value of the loss function is close to the oracle one. The same issue exists for density estimation where the observations are i.i.d. random variables, see the discussion in [55, Section 1.8].

One solution is to give unbiased estimators of the bias and the variance terms and to choose for the bandwidth h the one minimizing the sum of the two. While an unbiased estimator of the variance term is easily to find, the main issue is the approximation of the bias term. In the i.i.d case, the variance term is deterministic and known but in our case, an unbiased estimator of it is given by

$$\hat{V}_h = \frac{1}{n^2} \int_0^T \int_I \left(w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\right)^2 \mathbf{1}_{X_s\in I} dx dN_s.$$

Concerning the estimator of the bias term, in a context of i.i.d observations with a kernel estimator, [37] proposes to approximate the bias $\|q-q_h\|_I^2$ by $\|q-q_{h_{\min}}\|_I^2$ with $h_{\min}$ sufficiently small. If $h_{\min}\to 0$, the bias $\|q_{h_{\min}}-q\|_I^2\approx 0$ and $\|q_{h_{\min}}-q_h\|_I^2\approx\|q-q_h\|_I^2$. This method is derived from the classical Goldenshluger Lepski method [30; 39]. We adapt the method of [37] to the Poisson process framework but also to the local polynomial framework. An unbiased estimator of $\|q_{h_{\min}}-q_h\|_I^2$ is given by

$$\|\hat{q}_{h_{\min}}-\hat{q}_h\|_I^2-\hat{V}_h-\hat{V}_{h_{\min}}+2\hat{V}_{h,h_{\min}}$$

with $\hat{V}_{h,h_{\min}}$ equals to

$$\frac{1}{n^2}\int_0^T\int_I w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)w\left(x,h_{\min},\frac{X_s-x}{h_{\min}}\right)K_{h_{\min}}\left(X_s-x\right)\mathbf{1}_{X_s\in I}dxdN_s.$$

At the end, we select the following bandwidth

$$\hat{h}=\underset{h\in\mathcal{H}}{argmin}\ \|\hat{q}_{h_{\min}}-\hat{q}_h\|_I^2-\hat{V}_h+2\hat{V}_{h,h_{\min}}+\lambda\hat{V}_h$$

among a finite set $\mathcal{H}$ included in $(0,\infty)$ with $\lambda>0$ an hyper parameter chosen by the statistician. We assume that $\min\mathcal{H}=h_{\min}\geq\frac{\|K\|_\infty\|K\|_1|I|}{n}$, with $\|K\|_1=\int_{\mathbb{R}}|K\left(u\right)|du$ and $\max\mathcal{H}\leq\frac{2}{3}|I|\Delta$. This choice of bandwidth leads to the following result.

Result 1 *Assume* $\sup_{x\in I}q\left(x\right)<\infty$. *The loss* $\mathbb{E}\left(\|q-q_h\|_I^2|D\left(I,\nu\right)\right)$ *is bounded by the sum of*

$$\left(\lambda\vee\frac{1}{\lambda}+O\left(\log\left(n\right)^{-1}\right)\right)\min_{h\in\mathcal{H}}\mathbb{E}\left(||\hat{q}_h-q||_I^2|D\left(I,\nu\right)\right)$$

and

$$O\left(\log\left(n\right)\right)\mathbb{E}\left(||q_{h_{\min}}-q||_I^2|D\left(I,\nu\right)\right)+\frac{1}{\nu^2}O\left(\frac{\log\left(n\vee|\mathcal{H}|\right)^6}{n}\right).$$

Result 1 consists in an oracle inequality that is derived from the two concentration inequalities [51, Equation (2.2)] and [34, Theorem 4.2]. The error term of order $\log\left(n\right)\mathbb{E}\left(||q_{h_{\min}}-q||_I^2|D\left(I,\nu\right)\right)$ is caused by the approximation of $\|q-q_h\|_I^2$ by $\|q-q_{h_{\min}}\|_I^2$. The error coming from this term depends of the regularity of q. The term $\frac{1}{\nu^2}$ corresponds to an error caused by the quantity of observations of X in I and if it is small, it leads to more error. When n is large, and if $\lambda=1$, our choice of bandwidth gives values of the loss close to the optimal one if $\log\left(n\right)\mathbb{E}\left(||q_{h_{\min}}-q||_I^2|D\left(I,\nu\right)\right)$ is enough small.

Result 1 does not give information about the quality of our estimator. For $\rho,\ \beta,\ L>0$, let $\Lambda_{\rho,\beta}=\{f:I\to\mathbb{R}:\ f\left(x\right)\geq\rho,\ \sup_{x\in I}f\left(x\right)<\infty\}\cap\Sigma\left(\beta,L,I\right)$ where $\Sigma\left(\beta,L,I\right)$ is the Hölder class of order β on I with bounding constant L. Result 2 and Result 3 assess the performance of $\hat{q}_{\hat{h}}$ in the minimax sense over $\Lambda_{\rho,\beta}$ and answer to Question 2. We recall that m is the degree of the estimator polynomial.

Result 2 *The sequence* $\mathbb{E}\left(\varphi_n^2\|q-\hat{q}_{\hat{h}}\|_I^2|D\left(I,\nu\right)\right)$ *is bounded uniformly over* $\Lambda_{\rho,\beta}$ *with* φ_n *equals to* $n^{\frac{-\beta}{2\beta+1}}$ *if* $m\geq\lfloor\beta\rfloor$ *and* $n^{\frac{-m}{2m+1}}$ *if* $m<\lfloor\beta\rfloor$.

Result 3 *The rate of convergence $n^{\frac{-\beta}{2\beta+1}}$ is a lower bound in the minimax sense.*

Then, if $m \geq \lfloor \beta \rfloor$, our estimator in optimal in the minimax sense with rate of convergence $n^{\frac{-\beta}{2\beta+1}}$.

To answer Question 3, we propose to test

$$\begin{cases} H_0 : \ \exists \theta_0 \in \Theta, \ q = g_\theta \text{ against} \\ H_1 : \ \forall \theta \in \Theta, \ q \neq g_\theta \end{cases}$$

with $\Theta \subset \mathbb{R}^d$, $d \geq 1$ and g_θ a function parametrized by θ. Let us consider the contrast

$$\begin{aligned} M_n(\theta) = \| \hat{q}_{\hat{h}}(\cdot) - \int_0^T w\left(\cdot, \hat{h}, \frac{X_s - \cdot}{\hat{h}}\right) K_{\hat{h}}(X_s - \cdot) \mathbf{1}_{X_s \in I} g_\theta(X_s) \, ds \|_I^2 \\ - \frac{1}{n^2} \int_I \int_0^T w^2\left(x, \hat{h}, \frac{X_s - x}{\hat{h}}\right) K_{\hat{h}}^2(X_s - x) \mathbf{1}_{X_s \in I} dN_s dx. \end{aligned}$$

The second term is a bias correction term in order to have an asymptotic unbiased estimator. The first term measures the distance between $\hat{q}_{\hat{h}}$ which is an estimator of q under both hypothesis and g_θ, or rather a biased version of it allowing to avoid a bias term in this distance. This contrast converges in probability to $\|q - g_\theta\|_I^2$ and a natural estimator of $\hat{\theta}_n$ of θ_0 under H_0 is

$$\hat{\theta}_n = \inf_{\theta \in \Theta} M_n(\theta).$$

A way to test H_0 is to look at $M_n\left(\hat{\theta_n}\right)$ that is small under H_0 but diverges under H_1.

Result 4 *Under H_0, $\hat{\theta}_n$ converges to θ_0 at the rate $n^{-\frac{1}{2}}$ and a critical region of the test at level α is*

$$|M_n\left(\hat{\theta}_n\right)| \geq \hat{c}(\alpha) = n^{-1} \hat{h}^{-\frac{1}{2}} \sqrt{\hat{V}_n} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

where

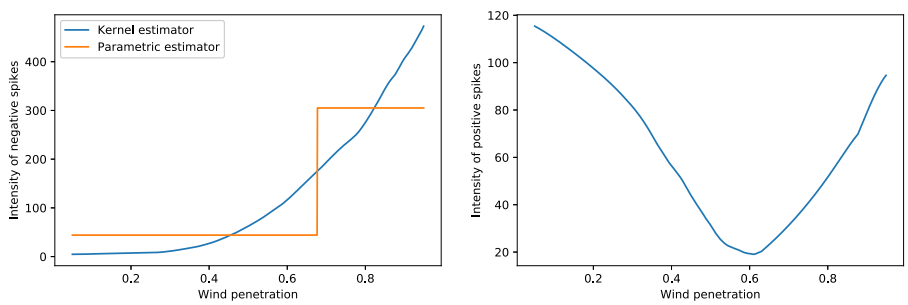
$$\hat{V}_n = C(K) \int_I \left(\frac{g_{\hat{\theta}_n}(y)}{\int_0^T K_{\hat{h}}(y - X_s) \mathbf{1}_{X_s \in I} ds} \right)^2 dy,$$

$C(K)$ is a constant depending only of K and Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ random variable.

A central theorem is also given for $\hat{\theta}_n$. Result 4 indicates that $M_n\left(\hat{\theta}_n\right)$ converges to 0 in probability at the convergence rate $n^{-1}\hat{h}^{-\frac{1}{2}}$ under H_0 and to ∞ under H_1. This test is similar to the one of [4] used to test if the drift and volatility of a diffusion belong to some parametric family.

Application to dependence modeling between electricity spot prices and wind production Following the ideas of [58], we study the dependence between the electricity spot price and the wind penetration index in Germany. The wind penetration index is defined as the ratio between the wind production and the total electricity production. Data considered are the hourly German spot price and hourly wind penetration index between year 2012 and 2016, both included. Our intuition is that high wind penetration index leads to negative spikes in electricity spot price time

series. Spikes are modeled by a strong mean reverting Poisson process, as in Chapter 3. Method of Chapter 3 is then used in order to detect the spikes in the spot time series. We distinguish positive and negative spikes and estimate the intensity of the point process as a function of the wind penetration index for each process using the local polynomial estimator of Chapter 4 of order 0, that is with kernel, see Figure 11a and Figure 11b. Using the parametrical test, we test if the two intensity functions are constant: the test is not rejected for positive spikes but is for negative ones with a confidence level at 95%. We also find an increasing function of the wind penetration for the negative spikes intensity.



(a) *For negative spikes.* (b) *For positive spikes.*

Figure 11: *Kernel estimators of the intensity of the spot spikes as a function the wind penetration.*

Based on these results, the spot price is modeled as the sum of a seasonality function, a continuous autoregressive process and two spikes process: one for the positive spikes and one for the negative ones, both having the same mean reversion. The intensity of the positive spike process is modeled by a constant but the one of the negative one has two states: a low intensity for low wind penetration values and a high intensity for high wind penetration values, see Figure 11a. Concerning the wind penetration index, as its values lie between 0 and 1, its logit is modeled by the sum of a seasonality function and a continuous autoregressive process of order 24. Methods of estimation are provided for both models.

In order to study the impact of our modeling, one consider the point of view of an electricity company buying electricity to an wind producer at a fixed price K. The wind producer produces $Q\%$ of the total wind production. The incomes of the electricity company over a period T are then equals to $Q\int_0^T C_t WP_t\,(S_t - K)\,dt$ where C_t is the total load. Value at Risk and Expected Shortfall of this model are compared to the ones of the model where the intensity of negative spikes is constant: the difference is significant between the two models.

5 Structure of the thesis

The thesis is composed of five chapters based on the following works:

- [Chapter 1] *On the control of the difference between two Brownian motions: a dynamic copula approach*, published in *Dependence Modeling*.

- [Chapter 2] *On the control of the difference between two Brownian motions: an application to energy markets modeling*, published in *Dependence Modeling*.

- [Chapter 3] *Estimation of a fast mean reverting jump process with application to spike modeling in electricity prices*, joint work with O. Féron and M. Hoffmann.

- [Chapter 4] *Local polynomial estimation of the intensity of a doubly stochastic Poisson process.*

- [Chapter 5] *A joint model for electricity and wind penetration with dependence in the electricity spikes*, joint work with A. Veraart, submitted in *Forecasting and Risk Management for Renewable Energy 2017: Conference proceedings.*

Bibliography

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[2] René Aid, Luciano Campi, Adrien Nguyen Huu, and Nizar Touzi. A structural risk-neutral model of electricity prices. *International Journal of Theoretical and Applied Finance*, 12(07):925–947, 2009.

[3] René Aid, Luciano Campi, and Nicolas Langrené. A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*, 23(3):387–438, 2013.

[4] Yacine Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial studies*, 9(2):385–426, 1996.

[5] Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009.

[6] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[7] Martin T Barlow and Marc Yor. Semi-martingale inequalities via the garsia-rodemich-rumsey lemma, and applications to local times. *Journal of functional Analysis*, 49(2):198–229, 1982.

[8] Fred Espen Benth. Cointegrated commodity markets and pricing of derivatives in a non-gaussian framework. In *Advanced Modelling in Mathematical Finance*, pages 477–496. Springer, 2016.

[9] Fred Espen Benth, Jan Kallsen, and Thilo Meyer-Brandis. A non-gaussian ornstein–uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2):153–169, 2007.

[10] Fred Espen Benth and Paul C Kettler. Dynamic copula models for the spark spread. *Quantitative Finance*, 11(3):407–421, 2011.

[11] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

[12] Fred Espen Benth, Nina Lange, and Tor Age Myklebust. Pricing and hedging quanto options in energy markets. *Journal of Energy Markets*, 2015.

[13] Tomasz R Bielecki, Jacek Jakubowski, Andrea Vidozzi, and Luca Vidozzi. Study of dependence for some stochastic processes. *Stochastic analysis and applications*, 26(4):903–924, 2008.

[14] Damien Bosc. *Three essays on modeling the dependence between financial assets.* PhD thesis, Ecole Polytechnique X, 2012.

[15] Maria Mori Brooks and J Stephen Marron. Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1):157–165, 1991.

[16] René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer, 2014.

[17] René Carmona, Michael Coulon, and Daniel Schwarz. Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2):167–202, 2013.

[18] René Carmona and Valdo Durrleman. Pricing and hedging spread options. *Siam Review*, 45(4):627–685, 2003.

[19] René Carmona and Valdo Durrleman. Generalizing the black-scholes formula to multivariate contingent claims. *Journal of computational finance*, 9(2):43, 2005.

[20] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[21] Feng Chen, Paul SF Yip, and KF Lam. On the local polynomial estimators of the counting process intensity function and its derivatives. *Scandinavian Journal of Statistics*, 38(4):631–649, 2011.

[22] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l'institut Henri Poincaré (B)*, volume 47, pages 1171–1196, 2011.

[23] David R Cox. Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer, 1992.

[24] Glenis Crane. Time and space varying copulas. *arXiv preprint arXiv:0812.3208*, 2008.

[25] William F Darsow, Bao Nguyen, and Elwood T Olsen. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

[26] Peter Diggle. A kernel method for smoothing point process data. *Applied statistics*, pages 138–147, 1985.

[27] Jean-David Fermanian and Marten Wegkamp. Time dependent copulas. *Preprint INSEE, Paris, France*, 2004.

[28] Olivier Féron and Elias Daboussi. *Commodities, Energy and Environmental Finance*, chapter Calibration of electricity price models, pages 183–207. Springer, 2015.

[29] Maurice J Frank, Roger B Nelsen, and Berthold Schweizer. Best-possible bounds for the distribution of a sum - a problem of kolmogorov. *Probability Theory and Related Fields*, 74(2):199–211, 1987.

[30] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.

[31] J Michael Harrison and Stanley R Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260, 1981.

[32] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105, 1992.

[33] Marc Hoffmann. On estimating the diffusion coefficient: parametric versus nonparametric. In *Annales de l'IHP Probabilités et statistiques*, volume 37, pages 339–372, 2001.

[34] Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer, 2003.

[35] Piotr Jaworski and Marcin Krzywda. Coupling of wiener processes by using copulas. *Statistics & Probability Letters*, 83(9):2027–2033, 2013.

[36] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical methods for financial markets*. Springer, 2009.

[37] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, pages 1–38, 2016.

[38] Suzanne S Lee and Per A Mykland. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563, 2007.

[39] Matthieu Lerasle, Nelo Molter Magalhães, and Patricia Reynaud-Bouret. Optimal kernel selection for density estimation. In *High Dimensional Probability VII*, pages 425–460. Springer, 2016.

[40] Eckhard Liebscher. Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99(10):2234–2250, 2008.

[41] GD Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4):803–806, 1982.

[42] Cecilia Mancini. Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52, 2004.

[43] Cecilia Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296, 2009.

[44] Thilo Meyer-Brandis and Peter Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(05):503–528, 2008.

[45] Susan Allbritton Murphy and Pranab Kumar Sen. Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1):153–180, 1991.

[46] Katsushi Nakajima and Kazuhiko Ohashi. A cointegrated commodity pricing model. *Journal of Futures markets*, 32(11):995–1033, 2012.

[47] Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2006.

[48] Andrew J Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556, 2006.

[49] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[50] Patricia Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126(1):103–153, 2003.

[51] Patricia Reynaud-Bouret. Concentration inequalities, counting processes and adaptive statistics. In *ESAIM: Proceedings*, volume 44, pages 79–98. EDP Sciences, 2014.

[52] Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, pages 623–632, 1982.

[53] Carlo Sempi. Coupled brownian motion. In *Combining Soft Computing and Statistical Methods in Data Analysis*, pages 569–574. Springer, 2010.

[54] M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[55] Alexandre B Tsybakov. Introduction to nonparametric estimation, 2009.

[56] Klaus J Utikal. Nonparametric inference for a doubly stochastic poisson process. *Stochastic processes and their applications*, 45(2):331–349, 1993.

[57] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.

[58] Almut ED Veraart. Modelling the impact of wind power production on electricity prices by regime-switching lévy semistationary processes. In *Stochastics of Environmental and Financial Economics*, pages 321–340. Springer, 2016.

[59] Shaomin Wu. Construction of asymmetric copulas and its application in two-dimensional reliability modelling. *European Journal of Operational Research*, 238(2):476–485, 2014.

[60] Tingting Zhang and SC Kou. Nonparametric inference of doubly stochastic poisson process data via the kernel method. *The annals of applied statistics*, 4(4):1913, 2010.

Introduction

1 Contexte

1.1 Objet de la thèse

Cette thèse traite de problèmes de modélisation de la dépendance entre processus stochastiques en temps continu et de leur estimation statistique pour une application en gestion des risques. Les différents travaux de cette thèse sont motivés par une application : la modélisation des marchés de l'électricité. En particulier, nous sommes intéressés par la modélisation de la dépendance entre les prix de l'électricité et différents facteurs de risque. L'électricité étant produite par différentes commodités énergétiques vendues sur des marchés financiers, une dépendance existe entre son prix et le prix des autres commodités. Les prix de l'électricité sont aussi fortement liés à des aléas physiques (consommation d'électricité, variable météorologique telle que le vent, la température...) considérés comme des facteurs de risque. Par exemple, en particulier en France, une basse température entraîne une utilisation des chauffages qui fait augmenter la demande et donc les prix de l'électricité. En Allemagne, un fort vent entraîne une augmentation de la production des énergies renouvelables et donc une diminution des prix de l'électricité. Ces deux types de dépendance,

1) dépendance avec les autres commodités et

2) dépendance physique,

constituent les applications principales de nos travaux.

Il est nécessaire de prendre en compte ces dépendances qui ont au moins deux principaux intérêts :

- mieux capturer la dynamique des prix de l'électricité et son comportement,
- avoir une meilleure quantification des risques financiers ;

ces deux points sont bien sûr liés. Prenons par exemple le point de vue d'un producteur d'électricité possédant une centrale à charbon. Si S_t dénote le prix de l'électricité à l'instant t, C_t le prix du charbon à l'intant t et K un coût fixe, ses revenus à la date T peuvent être modélisés de manière simplifiée par $\int_0^T (S_t - HC_t - K)^+ \, dt$ où H est une constante de normalisation entre les prix de l'électricité et les prix du charbon qui n'ont pas la même unité. Il est donc important d'avoir un modèle qui représente bien les propriétés statistiques des deux séries temporelles S et C mais il est aussi important de modéliser leur dépendance. La dépendance entre S et C a un impact sur la distribution des revenus et donc des risques financiers du producteur. Un autre exemple considéré dans cette thèse est celui d'un producteur voulant acheter un pourcentage Q de la production éolienne W_t à un prix fixe K pendant une période T. Ses revenus à la date T sont

égaux à $Q \int_0^T W_t (S_t - S_0)\, dt$ et la dépendance entre S et W a un impact sur sa distribution. Ces dépendances ont aussi un impact sur la manière de couvrir ces différents risques.

Les processus stochastiques en temps continu sont des outils pratiques pour la modélisation de la dynamique des prix, en particulier dans un contexte de gestion des risques. En effet, le calcul stochastique est un outil utile pour la valorisation et la couverture d'options financières. Bien que les études théoriques soient au coeur de cette thèse, la faisabilité opérationnelle ne peut être négligée :

- les modèles proposés doivent être "maniables", les simulations et la valorisation d'options classiques doivent être réalisables en un temps raisonnable ;
- les problématiques d'estimation doivent être effectuées dans un cadre d'observations discrètes.

Ce contexte nous amène à étudier dans cette thèse deux types de dépendance :

(i) la dépendance entre deux mouvements Browniens dans le Chapitre 1 et dans le Chapitre 2, appliquée à la modélisation de la dépendance entre les prix de l'électricité et les prix des combustibles,

(ii) la dépendance entre un processus ponctuel et une semimartingale continue dans le Chapitre 4 et le Chapitre 5, appliquée à la modélisation de la dépendance entre la fréquence d'apparition des pics dans les prix de l'électricité et des facteurs exogènes. Une méthode d'estimation de la fonction d'intensité du processus de Poisson fonction d'une variable exogène est proposée dans le Chapitre 4.

Avant d'étudier la dépendance entre les pics de prix et des facteurs exogènes, il est nécessaire de proposer un modèle que l'on peut inférer pour la modélisation des pics de prix de l'électricité et qui est adapté à un contexte de gestion des risques ; cela correspond au Chapitre 3 où l'on propose l'estimation d'un processus de pics bruité par une semimartingale continue.

1.2 Description des marchés de l'électricité

Pour mieux comprendre les aspects de modélisation, nous décrivons brièvement dans cette section les marchés de l'électricité. Les marchés de l'électricité sont des marchés locaux : il y a un marché par pays et la régulation est différente selon le pays. Cependant, l'équilibre entre consommation et production doit être satisfait partout et l'on trouve la même structure de marché dans les différents pays. Il existe trois sous marchés ordonnés selon leur horizon de temps.

Le marché intraday

Ce marché est un marché gré à gré et correspond à une maturité inférieure à un jour. Il assure la sécurité du système par un mécanisme de balancement : les différents acteurs peuvent rebalancer leur offre et leur demande de manière continue. Ce marché n'étant pas considéré dans cette thèse, nous ne donnons pas plus de détails.

Le marché spot

Le marché spot est un marché physique où l'électricité est livrée. C'est un marché d'enchères. Le jour précédant la livraison, chaque participant soumet une courbe d'offre par heure (ou par demi-heure). Cette courbe est construite selon l'ordre au mérite : le moyen de production le moins

cher est utilisé en premier, correspondant à la production solaire et éolienne en général. Comme les autres commodités énergétiques servent à produire l'électricité, leur prix a un impact sur cette courbe et donc sur le prix spot ; la dépendance est forte. La demande est croisée avec l'offre pour chaque heure (ou demi-heure) donnant ainsi le prix spot.

Remarque Le prix spot est plus proche d'un prix day ahead que d'un prix spot car le prix est fixé 24 heures avant.

L'électricité n'est pas un actif financier classique et présente quelques particularités. Premièrement, l'électricité n'est pas stockable et ne peut pas être échangée comme n'importe quel actif financier. Deuxièmement, les prix sont associés à une période de livraison : l'électricité n'est pas livrée instantanément mais de manière continue sur une période d'une heure par exemple. Toutes ces particularités font que les prix de l'électricité possèdent les caractéristiques suivantes :

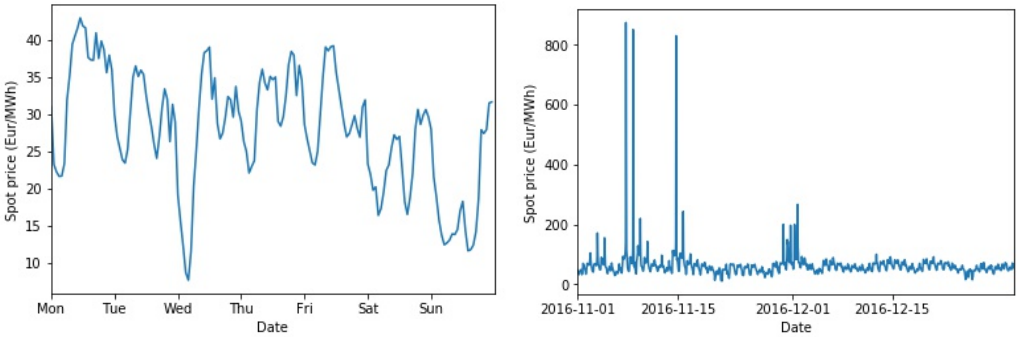
- Saisonnalité : ils possèdent une saisonnalité horaire, hebdomadaire et annuelle. Cette saisonnalité est fortement liée à celle de la consommation à cause de la non stockabilité de l'électricité.
- Pics de prix : les prix sautent vers le haut ou vers le bas prenant de fortes valeurs, positives ou négatives, avant de revenir très rapidement vers leur niveau d'origine. Ces pics apparaissent par example quand la demande est anormalement haute ou que la température est anormalement haute ou basse. Dans le cas d'une forte température, l'air conditionné produit ces pics et dans le cas d'une basse température, c'est le chauffage.
- Prix négatifs : c'est une conséquence de la non stockabilité. Si la production est plus forte que prévue, le coût d'arrêter une centrale peut être élevé et le producteur peut préférer payer pour que l'électricité soit consommée. En Allemagne, une production imprévue est causée par la pénétration des énergies renouvelables dans le système. Par exemple, une forte production éolienne non prévue peut entraîner des pics de prix négatifs.
- Retour à la moyenne : le retour à l'équilibre est présent pour les pics où il est très fort mais aussi quand le comportement des prix est normal avec un retour moins fort.

La Figure 1 montre quelques-unes de ces caractéristiques pour le marché Français.

Le marché forward

Le marché forward est un marché classique avec des actifs échangeables ; c'est un marché gré à gré. Les actifs ont une maturité comme pour les marchés forward classiques mais aussi une période de livraison. Contrairement au marché spot, même si il existe une période de livraison, l'électricité n'est pas livrée et le marché est ouvert à tout le monde. Les produits forward peuvent être utilisés dans un but de couverture mais aussi de spéculation.

Différents produits existent en fonction de la maturité et de la période de livraison qui peuvent être de plusieurs semaines, mois, trimestres, saisons ou années. Si nous sommes le 20 Juin 2017 aujourd'hui, le contrat dénommé "Juillet 2017" correspond à une livraison de l'électricité de manière continue durant le mois de Juillet ; le contrat "Année 2018" correspond à une livraison durant l'année 2018. Le premier contrat est noté One Month Ahead (1MAH) et correspond au produit qui délivre à partir du premier du mois suivant pendant tout le mois. De la même manière, le deuxième contrat est noté One Year Ahead (1YAH). Selon la date, le 1MAH correspond à plusieurs contrats,



(a) *Saisonnalité (semaine du* 04 *Juillet 2016).* (b) *Pics.*

Figure 1: *Illustration du prix spot Français.*

de même pour le 1YAH. Le 2MAH correspond au produit "Août 2017". Le Tableau 2 donne quelques exemples pour comprendre la nomenclature. La Figure 2 represente le 1YAH entre Janvier 2011 et Mars 2017. La diminution du niveau du prix est principalement due à la diminution du prix des combustibles.

Produit	Contrat	Début de la livraison	Fin de la livraison
1 Month Ahead	Juillet 2017	01/07/2017	31/07/2017
2 Month Ahead	Août 2017	01/08/2017	31/08/2017
3 Month Ahead	Septembre 2017	01/09/2017	30/09/2017
1 Quarter Ahead	Q3 2017	01/07/2017	31/09/2017
2 Quarter Ahead	Q4 2017	01/10/2017	31/12/2017
1 Year Ahead	2018	01/01/2018	31/01/2018

Table 2: *Produits forward vus du* 20 *Juin 2017.*

1.3 Quelques aspects de modélisation

Dans la littérature, la modélisation des marchés de l'électricité est fortement liée à celle des marchés d'obligations. Comme pour les marchés obligataires, il est nécessaire de calculer le prix spot S_t mais aussi la courbe forward $f(t,T)$, $t \leq T$ qui est le prix à la date t du produit qui délivre à la date T 1MWh d'électricité durant une heure. Ce produit n'existe pas mais est un moyen de modéliser le produit existant $f(t,T,\theta)$ qui délivre 1MWh d'électricité entre T et $T+\theta$. Par absence d'opportunité d'arbitrage, nous avons

$$f(t,T,\theta) = \frac{1}{\theta}\int_T^{T+\theta} f(t,u)\,du.$$

La modélisation jointe du prix spot et du prix des produits forward est nécessaire car notre portefeuille contient souvent tous ces actifs, et surtout car les produits forward permettent de couvrir les risques financiers liés au prix spot. Deux approches existent, ce qui est aussi le cas pour les marchés obligataires.

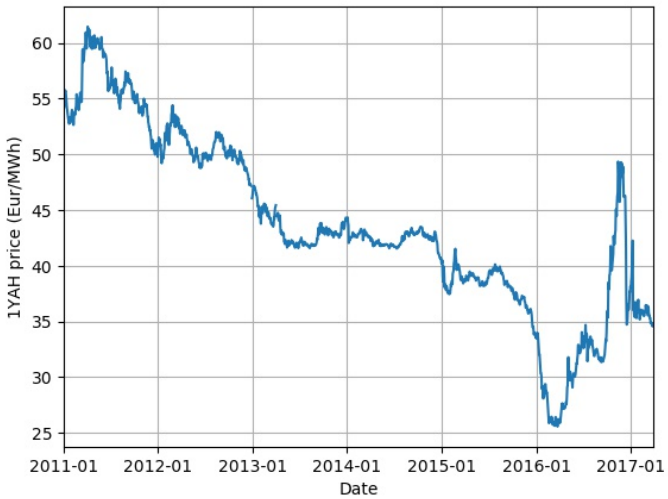


Figure 2: *Prix du 1 Year Ahead sur le marché Français entre Janvier 2011 et Mars 2017.*

- La première approche consiste à modéliser directement le prix spot, ce qui est équivalent à la modélisation des taux d'intérêt pour les marchés d'obligations ; Vasicek [57] a proposé le premier modèle pour ces marchés. Dans ce cas, le prix forward est égal à $\mathbb{E}^{\mathbb{Q}}\left(S_T|\mathcal{F}_t\right)$ où $\mathcal{F}_t$ est la filtration générée par le prix spot correspondant à l'information donnée par le spot jusqu'à la date t et $\mathbb{Q}$ est une probabilité risque neutre. Comme le spot n'est pas un actif échangeable, le théorème fondamental de la valorisation d'actifs [31] ne s'applique pas et le spot n'a pas besoin d'être modélisé par une martingale même sous une probabilité risque neutre, ce qui nous laisse une certaine liberté pour la modélisation. Cependant, comme les produits forward sont échangeables, leur prix doit être martingale : ceci est assuré par la martingalité de $f\left(t,T\right)$ qui est défini comme une espérance conditionnelle. Les modèles les plus communs consistent en une modélisation du spot (ou du logarithme du spot) par une somme d'une fonction de saisonnalité et d'une diffusion multi-facteurs de la forme

$$Y_t = \sum_{i=1}^{m} w_i Y_t^i,\ dY_t^i = -\beta_i Y_t^i dt + dL_t^i,\ t \in [0,T]$$

  où les L^i sont des processus de Levy, souvent des mouvements Browniens pour modéliser la partie diffusive et des processus de Poisson composés pour modéliser la partie pic. Le lecteur peut se référer à [20; 9; 44] pour plus d'informations sur ce modèle.

- La seconde approche est liée à celle de Heath Jarrow Morton [32] et consiste à modéliser directement la courbe forward $f\left(t,T\right)$. Le prix spot est alors donné par $S_t = \lim\limits_{T\to t} f\left(t,T\right)$. Comme mentionné précédemment, le processus stochastique $f\left(t,T\right)$ est généralement modélisé par une martingale pour chaque T sous $\mathbb{Q}$ pour avoir les produits $f\left(t,T,\theta\right)$ martingales. Le modèle le plus présent consiste en une modélisation de la courbe par

$$df\left(t,T\right) = f\left(t,T\right)\left(\sum_{i=1}^{N} \sigma_i\left(t,T\right) dW_t^i\right)$$

où $\left(W^i\right)_{i=1..N}$ est un mouvement Brownien multidimensionnel. Le lecteur peut se référer au travail de Benth et Koekebakker [11] pour plus d'informations sur ce modèle et ses applications. Dans cette thèse, nous considérons souvent pour les applications le modèle deux facteurs qui correspond au cas

$$N = 2,\ \sigma_1(t,T) = \sigma_s e^{-\alpha(T-t)},\ \sigma_2(t,T) = \sigma_l.$$

La dynamique de la courbe $T \mapsto f(\cdot, T)$ est contrôlée par deux facteurs : un facteur court terme $\sigma_s e^{-\alpha(T-t)} dW_t^1$ et un facteur long terme $\sigma_l dW_t^2$. Le facteur court terme modélise l'effet Samuelson : la volatilité augmente quand le temps à maturité diminue. Le facteur long terme modélise la volatilité des produits long terme : la volatilité court terme est proche de 0 pour ces produits et nous aurions une volatilité nulle sans la volatilité long terme, ce qui n'est pas le cas.

Concernant la modélisation de la dépendance, une approche spécifique consiste en la modélisation de la construction des prix spot (par ordre au mérite par exemple) par les variables fondamentales (demande, prix des combustibles, capacités de production...). Ces modèles sont appelés modèles structurels, voir [2; 3; 17; 16] pour quelques exemples. La dépendance entre les prix de l'électricité et les facteurs de risque est en général bien capturée mais le calcul des prix forward et du prix des options est coûteux. Nous préférons considérer des modèles à forme réduite, où le prix spot est modélisé par une diffusion. La pratique la plus courante est de modéliser la dépendance entre les prix de l'électricité (spot ou forward) et les prix des autres commodités énergétiques (spot ou forward) par une corrélation entre les mouvements Browniens, voir [18] par exemple. C'est aussi le cas pour la modélisation de la dépendance entre la température et les prix spot [12]. Nous pouvons trouver dans la littérature des modèles de dépendances plus complexes dans un cadre de modélisation avec des diffusions et nous en présentons quelques uns ici. Dans un contexte de modélisation des prix spot, un modèle intéressant est celui de Nakajima et Ohashi [46] qui incorpore de la co-intégration avec les autres prix des commodités dans un cadre de diffusions Browniennes. Le prix spot de l'électricité retourne vers une moyenne pondérée des prix des commodités, incluant lui-même. Cependant, les prix forward sont durs à calculer et ne présentent pas de volatilité long terme, ce qui implique que les produits à maturité longue ont une volatilité quasi nulle. Benth [10] utilise des copules dépendantes du temps pour la modélisation de la dépendance entre les prix de l'électricité et ceux du gaz en temps discret. Un récent travail de Benth propose d'inclure de la co-intégration dans un cadre multi-facteurs avec des processus Levy [8] ; ce modèle peut être utilisé pour les prix forward et les prix spot.

2 Première partie : Modélisation de la dépendance entre mouvements Browniens

La structure naturelle pour la modélisation de la dépendance pour les composantes d'un mouvement Brownien multidimensionnel est l'utilisation d'une matrice de corrélation. Cette matrice de corrélation correspond à une dépendance linéaire entre les composantes du mouvement Brownien à chaque instant t et est très facile à manipuler. Cette structure de dépendance est utilisée dans la plupart des modèles financiers, le plus utilisé étant le modèle de Black Scholes multivarié [19]. Une des principales applications du mouvement Brownien multidimensionnel en finance est la valorisation et la couverture d'options indexées sur plusieurs actifs avec un payoff $h(S_T)$ où S est une diffusion multidimensionnelle menée par un mouvement Brownien multidimensionnel. Un

cas important en deux dimensions est le cas $h(x,y) = (x-y-K)^+$ avec K une constante, ce qui correspond à une option spread. Le prix de l'option est donné par $\mathbb{E}^{\mathbb{Q}}(h(S_T))$ où $\mathbb{Q}$ est une probabilité risque neutre sous laquelle S est martingale. Nous pouvons voir que le prix de l'option est impacté par les modèles marginaux pour chaque composante de S mais aussi par la structure de dépendance entre les différentes composantes, qui ne peut être négligée.

La modélisation de la dépendance par une matrice de corrélation présente quelques limites et une d'entre elle est la symétrie. Dans cette partie, nous considérons seulement le cas en deux dimensions. Soient B^1 and B^2 deux mouvements Browniens. Dans le cas simplifié où les volatilités des deux actifs S^1 and S^2 sont les mêmes, le payoff de l'option spread $\left(S_T^1 - S_T^2 - K\right)^+$ dépend de la distribution de $B_T^1 - B_T^2$ comme les deux actifs sont martingales sous $\mathbb{Q}$ et ont des tendances nulles. Le principal sujet d'intérêt est donc la distribution de $B^1 - B^2$. Supposons dans un premier temps que leur dépendance est modélisée par une corrélation, c'est-à-dire que $d\langle B_t^1, B_t^2\rangle = \rho dt$. Dans ce cas, nous avons pour $x \in \mathbb{R}$ and $t > 0$,

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \mathbb{P}\left(B_t^1 - B_t^2 \leq -x\right)$$

impliquant pour $x \geq 0$,

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) \leq \frac{1}{2}.$$

La distribution de $B_t^1 - B_t^2$ est donc symétrique. Cette symétrie est causée par la structure de la dépendance entre B_t^1 et B_t^2 qui est symétrique pour un temps donné t et qui correspond à la copule de $\left(B_t^1, B_t^2\right)$. Rappelons qu'une copule (en dimension deux) est une fonction deux-croissante $C : [0,1]^2 \mapsto [0,1]$ à marges uniformes ; le lecteur peut se référer à [47] pour plus d'informations sur les copules. Le théorème de Sklar [54] indique que pour n'importe quel couple de variables aléatoires (X,Y), sa distribution est caractérisée par la distribution marginale de X, la distribution marginale de Y et une copule C représentant la structure de dépendance. Dans le cas d'une corrélation, la copule de $\left(B_t^1, B_t^2\right)$ est appelée copule gaussienne et est symétrique, c'est-à-dire que $C(u,v) = C(v,u)$ pour $u,\ v \in [0,1]$. Dans un cas plus général, nous prouvons que si X et Y ont la même distribution et que leur copule est symétrique, alors la distribution de $X - Y$ est symétrique. Comme les distributions marginales de B_t^1 and B_t^2 sont connues, le contrôle de la distribution de $B_t^1 - B_t^2$ dépend seulement de la copule de $\left(B_t^1, B_t^2\right)$. Pour avoir des valeurs supérieures $\frac{1}{2}$ pour la quantité $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right),\ x \geq 0$, il est nécessaire de considérer des copules asymétriques, qui ne sont pas faciles à construire, voir [40] ou [59] par exemple. Cependant, les copules sont un outil naturel pour les variables aléatoires mais ce n'est plus le cas pour les processus stochastiques. Les copules asymétriques classiques ne sont peut être pas adaptées pour les mouvements Browniens. Cela mène aux questions suivantes :

Question 1 *Existe-t-il des copules asymétriques et admissibles pour modéliser la dépendance entre deux mouvements Browniens ?*

Question 2 *Y-a-t-il des copules admissibles permettant à* $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ *de prendre des valeurs plus élevées que dans le cas où les Browniens sont corrélés ?*

Une première étape pour répondre à la Question 1 est de définir convenablement ce qu'est une copule admissible pour des mouvements Browniens. La littérature concernant les copules pour les processus stochastiques n'est pas très étoffée et peut être divisée en trois sujets différents. Le premier sujet concerne les copules pour les processus stochastiques en temps discret. Ces copules ont été introduites par Patton [48] et généralisées par Fermanian [27]. A chaque temps t, la copule est construite conditionnellement à ce qu'il s'est passé au temps $t-1$. Dans un cadre en temps

continu, [25] considère la copule entre X_s et X_t avec X un processus stochastique et donne des conditions sur cette copule pour que X soit un processus Markovien. Le dernier cadre est celui dans lequel on se place et consiste à étudier la copule entre deux processus stochastiques, et en particulier deux mouvement Browniens, pour chaque temps t [13; 24; 53; 14; 35]. Dans [14] et [35], la notion de copule admissible est liée à une fonction de corrélation locale entre deux processus Markoviens. Dans [24], ces résultats sont généralisés en dimension n. Le cadre Markovien semble naturel pour les mouvements Browniens et est aussi considéré dans [13]. Nous proposons une définition similaire à celle de [13] incluant le travail de [14], [35] and [24].

Définition (Copule admissible pour les diffusions Markoviennes) Une collection de copules $C = (C_t)_{t\geq 0}$ est admissible pour les n diffusions Markoviennes à valeurs dans $\mathbb{R}$, $n \geq 2$, $\left(X^i\right)_{1\leq i\leq n}$ définies sur un espace probabilisé commun $(\Omega, \mathcal{F}, \mathbb{P})$ si il existe une diffusion Markovienne dans $\mathbb{R}^m$ $Z = \left(Z^i\right)_{1\leq i\leq m}$, $m \geq n$, définie sur une extension de $(\Omega, \mathcal{F}, \mathbb{P})$ telle que

$$\begin{cases} \mathcal{L}\left(Z^i\right) = \mathcal{L}\left(X^i\right), 1 \leq i \leq n, \\ Z_0^i = X_0^i, 1 \leq i \leq n, \\ \text{pour } t \geq 0, \text{ la copule de } \left(Z_t^i\right)_{1\leq i\leq n} \text{ est } C_t. \end{cases}$$

avec $\mathcal{L}(Y)$ le générateur infinitésimal d'un processus Markovien Y.

Cette définition inclut, dans le cas des mouvements Browniens, la corrélation déterministe, la corrélation locale mais aussi la corrélation stochastique si cette dernière est un processus Markovien comme il est possible d'avoir $m \geq n$. Dans le modèle à corrélation locale, dans le cas de la dimension deux, Jaworski and Krzywda [35] prouvent qu'une copule est admissible pour deux mouvements Browniens si

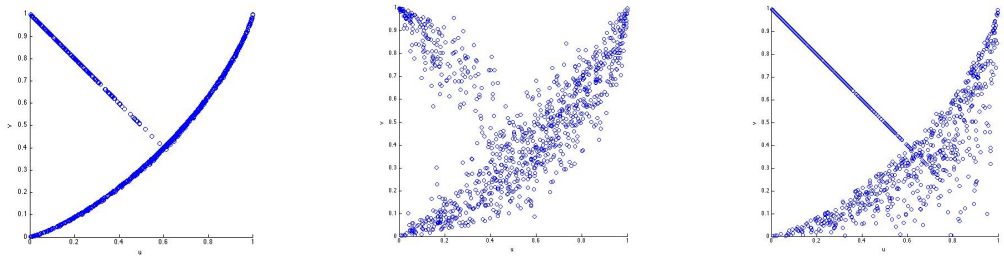
$$\left|\frac{1}{2}e^{\frac{\Phi^{-1}(v)^2-\Phi^{-1}(u)^2}{2}}\frac{\partial^2_{u,u}C(u,v)}{\partial^2_{u,v}C(u,v)} + \frac{1}{2}e^{\frac{\Phi^{-1}(u)^2-\Phi^{-1}(v)^2}{2}}\frac{\partial^2_{v,v}C(u,v)}{\partial^2_{u,v}C(u,v)}\right| < 1 \quad \forall (t,u,v) \in \mathbb{R}^+ \times [0,1]^2$$

quand la copule ne dépend pas du temps avec Φ la fonction de répartition d'une loi normale standard. En particulier, ils prouvent que l'extension de la copule Farlie-Gumbel-Morgenstern dans un cadre dynamique $C_t(u,v) = uv(1 + a(1-u)(1-v))$, $t \geq 0$ avec $a \in [-1,1]$ est admissible pour les mouvements Browniens. Cependant, cette copule n'est pas asymétrique.

Soit B^1 un mouvement Brownien défini sur une espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ avec $(\mathcal{F}_t)_{t\geq 0}$ satisfaisant les conditions usuelles (continuité à droite et complétude) avec B^1 adapté à $(\mathcal{F}_t)_{t\geq 0}$. Alors que la corrélation est le moyen standard de construire un mouvement Brownien B^2 dépendant de B^1 à partir de B^1, un autre moyen existant dans la littérature est de considérer la réflexion de B^1 sur $x = h$ avec $h \in \mathbb{R}$ qui est aussi un $\mathcal{F}$ mouvement Brownien selon le principe de réflexion (voir [36, Theorem 3.1.1.2, p. 137]). Le Brownien réfléchi $\tilde{B}^h$ est défini par $\tilde{B}^h_t = -B^1_t + 2(B^1_t - B_{\tau^h})\mathbf{1}_{t\geq\tau^h}$ avec $\tau^h = \inf\{t \geq 0 : B^1_t = h\}$. Soient $M(u,v) = \min(u,v)$ et $W(u,v) = \max(u+v-1, 0)$ pour $u,\ v \in [0,1]$. La copule de $(B^1, \tilde{B}^h)$, $\left(C_t^{ref,h}\right)_{t\geq 0}$, est définie par

$$C_t^{ref,h}(u,v) = \begin{cases} v & \text{if} \quad \Phi^{-1}(u) - \Phi^{-1}(v) \geq \frac{2h}{\sqrt{t}} \\ W(u,v) + \Phi\left(\Phi^{-1}(M(u,1-v)) - \frac{2h}{\sqrt{t}}\right) & \text{if} \quad \Phi^{-1}(u) - \Phi^{-1}(v) < \frac{2h}{\sqrt{t}} \end{cases}$$

et est admissible pour les mouvements Browniens. Une illustration est donnée dans la Figure 3a. Un moyen de construire de nouvelles copules à partir de celle-ci est de considérer la dépendance entre B^1 et un Brownien corrélé à la réflexion de B^1, voir Figure 3b. Un autre moyen est de considérer une barrière stochastique, voir Figure 3c où la barrière est égale à $h + \mathcal{E}$ et $\mathcal{E}$ suit une loi exponentielle.



(a) (b) (c)

Figure 3: *La Reflection Brownian Copula $C^{ref,h}$ et certaines de ses extensions au temps $t = 1$ avec $h = 2$. La Figure 3a est la Reflection Brownian Copula. La Figure 3b est la l'extension considérant un Brownien corrélé à la réflexion du premier avec une corrélation $\rho = 0.95$. La Figure 3c est l'extension dans le cas où la barrière est stochastique et est égale à la somme de h et d'une variable aléatoire exponentielle de paramètre $\lambda = 2$.*

Résultat 1 *La copule $C^{ref,h}$ et ses extensions sont admissibles pour les mouvements Browniens et sont asymétriques.*

Avant de répondre à la Question 2, intéressons nous à la plage de valeurs pouvant être prises par $\mathbb{P}(X - Y \geq x)$ quand X et Y sont deux variables aléatoires suivant une loi normale standard, ce qui correspond au cas statique de notre problème.

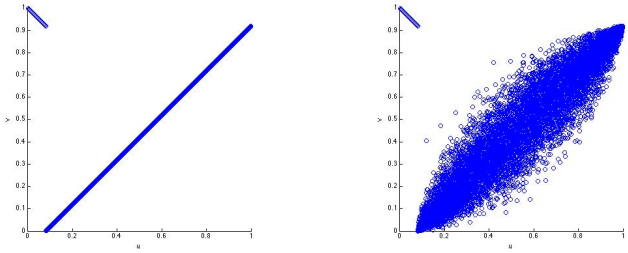
Résultat 2 *La plage de valeurs pouvant être prises par $\mathbb{P}(X - Y \geq x)$ est $\left[0, \Phi\left(\frac{-x}{2}\right)\right]$ si l'on se limite à l'ensemble des copules gaussiennes et $\left[0, 2\Phi\left(\frac{-x}{2}\right)\right]$ sinon.*

La borne supérieure $2\Phi\left(\frac{-x}{2}\right)$ est achevée par la copule

$$C^r(u,v) = \begin{cases} M(u - 1 + r, v) & \text{if } (u,v) \in [1-r,1] \times [0,r], \\ W(u,v) & \text{if } (u,v) \in [0,1]^2 \setminus ([1-r,1] \times [0,r]) \end{cases}$$

avec $r = 2\Phi\left(\frac{-\eta}{2}\right)$, voir Figure 4a. La copule présente deux états de dépendance : le premier correspond à la copule contre monotone W, équivalente à une corrélation de -1, en haut à gauche du carré unitaire et le second correspond à la copule monotone M, équivalente à une corrélation de 1. Le résultat provient de [47, Section 6.1] et [29; 52; 41] où les bornes de $\mathbb{P}(X + Y > x)$ sont étudiées. La plage de valeurs entre $\left[\Phi\left(\frac{-x}{2}\right), 2\Phi\left(\frac{-x}{2}\right)\right]$ est atteinte en considérant la copule avec une corrélation relâchée dans la partie contre monotonique, voir Figure 4.

Le Résultat 3 est le même que le Résultat 2 dans un cadre dynamique et donne une réponse à la Question 2.



(a) $\rho = 1$. (b) $\rho = 0.95$.

Figure 4: *Copule patchwork $C^r(u,v)$ présentant deux états selon la valeur de u : la première copule est en haut à gauche du carré unitaire et est W ; la seconde est la copule gaussienne avec une corrélation égale à ρ, avec $\rho = 1$ correspondant à la copule dégénérée M ou $\rho = 0.95$. r est égal à $2\Phi\left(\frac{-\eta}{2}\right)$ avec $\eta = 0.2$.*

Résultat 3 *La plage de valeurs pouvant être prises par $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ est $\left[0, \Phi\left(\frac{-x}{2\sqrt{t}}\right)\right]$ si l'on se limite à une dépendance avec corrélation et $\left[0, 2\Phi\left(\frac{-x}{2\sqrt{t}}\right)\right]$ si l'on considère toutes les copules admissibles pour les mouvements Browniens. Ces valeurs peuvent être atteintes en utilisant la Reflection Brownian Copula.*

La plage de valeurs pouvant être prises est la même que dans le cadre statique où l'on considère des variables aléatoires. Un résultat intéressant est que la Reflection Brownian Copula qui atteint la borne supérieure présente aussi deux états de corrélation : un avec une corrélation égale à 1 et un avec une corrélation égale à -1, comme pour la copule achevant la borne supérieure dans le cas statique.

Application à la modélisation des prix des commodités énergétiques Les options spread sont courantes sur les marchés des commodités. Rappelons ici le problème du producteur possédant une centrale à charbon dont les revenus sont modélisés par $(S_t - HC_t - K)^+$. S et C sont modélisés par des diffusions menées par des mouvements Browniens et la dépendance entre ces Browniens est souvent modélisée par une corrélation, impliquant de la symétrie dans la distribution de $S_t - HC_t$. Cependant, le charbon est un combustible pour l'électricité et HC_t a plus de probabilité d'être inférieur, ce qui ne peut pas être modélisé avec une corrélation. Les modèles marginaux étant satisfaisant, nous souhaitons seulement changer la structure de dépendance entre les prix de l'électricité et ceux du charbon, ce qui revient à changer la copule entre les deux mouvements Browniens. Les résultats de cette partie ont montré que pour capturer de l'asymétrie et des plus grandes valeurs pour la fonction de survie de la différence des deux Browniens, il faut considérer deux états de corrélation, un négatif et un positif, à la place de un. Ce résultat mène aux deux modèles suivants :

- Un modèle multi-barrière basé sur la copule entre un mouvement Brownien et sa réflexion : nous définissons deux barrières ν et η avec $\nu < \eta$ et nous considérons deux mouvements Browniens indépendants X et B^Y. Nous construisons le mouvement Brownien Y^n qui est corrélé à $\tilde{X}^n$: $Y^n = \rho\tilde{X}^n + \sqrt{1-\rho^2}B^Y$, avec $\tilde{X}^n$ le mouvement Brownien égal à $-X$ au début et réfléchissant quand $X - Y^n$ tape une barrière à deux états égale à η avant la première réflexion et passant de η à ν ou de ν à η à chaque réflexion.

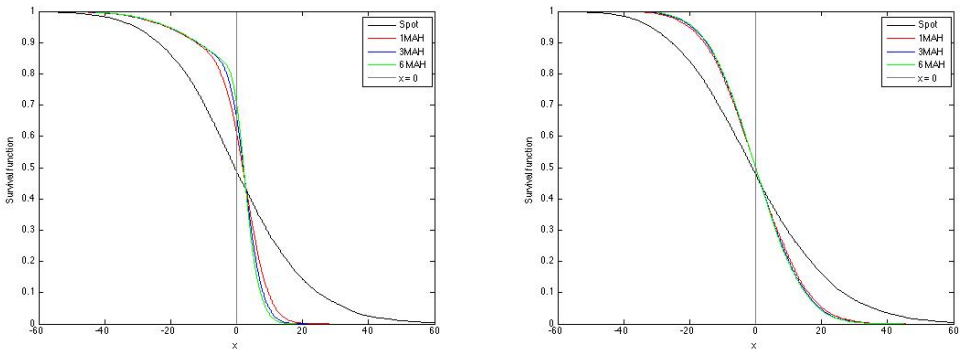
- Un modèle à corrélation locale avec deux états de corrélation : la fonction de corrélation locale est choisie Lipschitz et égale à $\rho_1 < 0$ quand $x - y \leq \nu$ et à $\rho_2 > 0$ quand $x - y \geq \eta$ avec $\eta > \nu$.

Ces deux modèles semblent être équivalents. Concernant le premier modèle, une formule fermée est établie pour la fonction de répartition de la différence entre les deux mouvements Browniens. Le second modèle est plus facile à utiliser mais ne dispose pas d'une formule fermée pour cette fonction. Des valeurs supérieures à $\frac{1}{2}$ sont atteintes par $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ avec $x \geq 0$ dans chacun des modèles.

Chaque prix de commodité est modélisé par une diffusion à deux facteurs

$$df^i(t,T) = f^i(t,T)\left(\sigma_s^i e^{-\alpha^i(T-t)} dW_t^{s,i} + \sigma_l^i dW_t^{l,i}\right),\ i = \{\text{Electricité, Charbon}\}.$$

Les paramètres sont estimés sur les marché forward électricité et charbon durant 2014 en France avec des observations journalières en utilisant la méthode de [28]. Nous comparons le modèle benchmark utilisé qui correspond à une dépendance modélisée par une matrice de corrélation entre les quatre mouvements Browniens et un modèle où la dépendance entre les deux facteurs long terme est modélisée avec le modèle multi-barrière. Modéliser la dépendance entre les facteurs court terme avec un modèle multi-barrière n'a pas d'impact sur les résultats car les volatilités court terme sont trop différentes. La Figure 5 représente la fonction de survie de la différence entre les produits Spot, 1MAH, 3MAH et 6MAH de chaque commodité pour chaque modèle de dépendance. La fonction de survie dans le modèle multi-barrière prend des valeurs plus hautes que dans le modèle benchmark pour $x \geq 0$ et pour des produits long terme, ce qui est cohérent avec le fait que la dépendance est modélisée sur les Browniens long terme. Les résultats sont les mêmes si l'on considère le modèle à corrélation locale.



(a) *Modèle multi-barrière.* (b) *Modèle benchmark.*

Figure 5: *Fonction de survie empirique de la différence entre le prix de l'électricité et le prix du charbon au temps $t = 365$ jours estimée avec 10000 simulations et un pas de temps égal à $\frac{1}{24}$ jour pour différents produits (Spot, 1MAH, 3MAH, 6MAH) dans le modèle multi-barrière et dans le modèle benchmark.*

Un des principaux problèmes de ce modèle est la calibration et le choix des valeurs de la barrière, qui doit dépendre de la valeur initial de $f^{\text{Electricité}}(0,T) - Hf^{\text{Charbon}}(0,T)$ si l'on veut avoir plus d'impact sur la fonction de survie de $f^{\text{Electricité}}(t,T) - Hf^{\text{Charbon}}(t,T)$.

3 Deuxième partie : Estimation d'un processus de pics en statistiques haute fréquence

Un pic est défini comme un saut, positif ou négatif, revenant à la valeur 0 en une courte période de temps. Un modèle naturel pour celui-ci est un processus de saut avec un fort retour à la moyenne, voir Figure 6. Considérons un processus stochastique X défini sur un espace de probabilité filtré $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P}\right)$ de la forme

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + Z_t^\beta,\ t \geq 0$$

avec

$$Z_t^\beta = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}\left(dt, dx\right),$$

W un mouvement Brownien standard, σ et μ deux processus càdlàg adaptés et $\underline{p}$ une mesure de Poisson sur $\mathbb{R}^+ \times \mathbb{R}$ indépendante de W de compensateur $\underline{q} = \lambda dt \otimes \nu\left(dx\right)$. X est donc la somme d'une semimartingale d'Itô et d'un processus de saut avec retour à la moyenne correspondant au processus de pics.

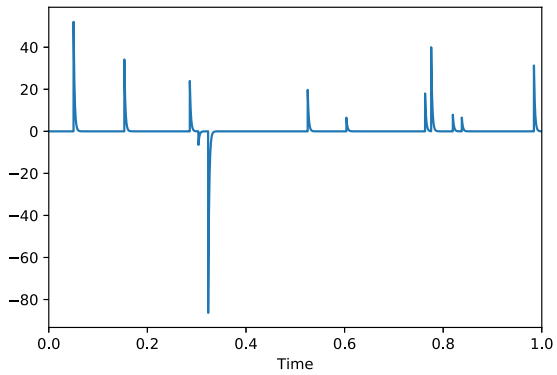
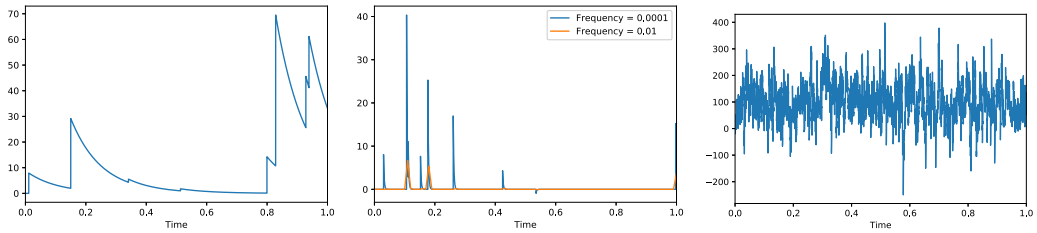


Figure 6: *Illustration d'un processus de pics.*

X est observé sur une grille régulière $\mathcal{M} = \{t_i = i\Delta_n, 0 \leq i \leq \lfloor \frac{T}{\Delta_n} \rfloor\}$ avec $\Delta_n = \frac{T}{n}$. Nous supposons que $\Delta_n \to 0$ avec T fixé, correspondant à un cadre haute fréquence avec horizon fini. Notre objectif est d'estimer les paramètres du processus de pics Z^β et plus particulièrement le paramètre β correspondant à la vitesse de retour à la moyenne. Si nous n'ajoutons pas d'autres hypothèses, le paramètre β est un paramètre de drift et est non identifiable si T est fixé ; le lecteur peut se référer à Aït-Sahalia et Jacod [6] pour la non identifiabilité du drift. Néanmoins, nous sommes intéressés par le cas où la vitesse de retour à la moyenne est forte. Une illustration est donnée en Figure 7a : si β est trop petit par rapport à Δ_n, l'effet pic n'est pas observé et le processus ne revient pas à 0 avant l'instant T. Pour modéliser cet effet pic, il est nécessaire d'ajouter l'hypothèse $\beta = \beta_n \to \infty$. Il faut travailler cependant sous l'hypothèse $\beta_n \Delta_n \lesssim 1$ pour observer tous les pics : dans le cas contraire, un pic peut avoir lieu et revenir à 0 en un temps Δ_n et il n'est pas possible de l'observer, voir Figure 7b. Une hypothèse de non divergence est aussi nécessaire dans le cas $\lambda = \lambda_n$ qui est $\frac{\lambda_n}{\beta_n} \lesssim 1$: dans le cas contraire, le nombre moyen de sauts λ_n

est plus grand que la vitesse de retour à la moyenne β_n et le processus diverge comme illustré dans la Figure 7c. Une autre condition qui est classique est $\lambda_n^2 \Delta_n \to 0$, supposant qu'il y a au plus un saut dans un intervalle de temps de taille Δ_n. Dans ce cadre, il n'y a plus de résultats concernant l'identifiabilité ou non du paramètre β, amenant aux questions suivantes.



(a) *Retour à la moyenne faible.* (b) *Fréquence trop forte.* (c) *Nombre de sauts trop élevé.*

Figure 7: *Processus de pics dans des régimes non considérés.*

Question 1 *Comment identifier les temps et les tailles de sauts du processus de pics dans ce nouveau cadre ?*

Question 2 *Comment peut-on estimer le paramètre β_n si cela est possible et quelle est l'erreur d'estimation ?*

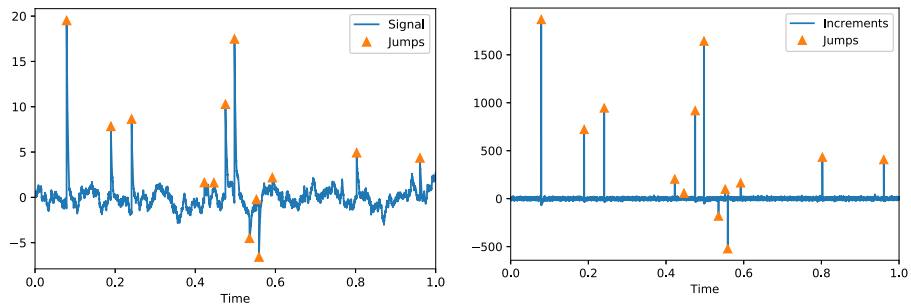
La détection de sauts dans un modèle de diffusion à sauts avec des observations haute fréquence est une question importante qui a été étudiée par plusieurs auteurs dans la littérature. Une des principales applications est l'estimation de la volatilité en présence d'une composante de saut. Le lecteur peut se référer aux travaux de Mancini [42; 43], Aït et Jacod [5; 6] ou Lee et Mykland [38]. Dans tous ces travaux, l'idée principale consiste à regarder la taille d'un incrément $\Delta_i^n X = X_{t_i} - X_{t_{i-1}}$. Cet incrément peut être écrit comme la somme d'un incrément de tendance, d'un incrément de Brownien et d'un incrément de saut. Comme la tendance est absolument continue par rapport à la mesure de Lebesgue, son incrément est d'ordre Δ_n. L'incrément du Brownien est d'ordre $\sqrt{\Delta_n}$. L'incrément du saut est d'ordre 1 quand il y a un saut entre les instants t_{i-1} et t_i et égal à 0 si il n'y en a pas. Ainsi, si l'on considère un seuil $v_n \asymp \Delta_n^{-\varpi}$ avec $\varpi \in \left(0, \frac{1}{2}\right)$, la quantité $\frac{|\Delta_i^n X|}{\sqrt{\Delta_n} v_n}$ converge vers 0 en l'absence de sauts mais vers ∞ lorsqu'il y en a.

Dans notre cas, le problème est légèrement différent à cause du terme supplémentaire dû au retour à la moyenne après un saut qui est d'ordre $\beta_n \Delta_n$. Lorsque $\beta_n \Delta_n \to 0$, il est encore possible de différencier l'incrément du saut de l'incrément du retour à la moyenne, voir Figure 8. Lorsque $\beta_n \Delta_n \asymp 1$, il n'est plus possible de distinguer les deux comme ils ont le même ordre de grandeur, voir Figure 9. Cependant, après un saut, l'incrément qui est celui de retour à la moyenne a le signe opposé de celui du saut. Nous pouvons montrer, sous certaines conditions, qu'après un incrément de retour à la moyenne qui est supérieur au seuil $v_n \sqrt{\Delta_n}$, si il n'y a pas trop de sauts, l'incrément suivant aura le même signe que celui de retour à la moyenne. La stratégie suivante est donc adoptée : soient $\mathcal{I}_n(1) < ... < \mathcal{I}_n\left(\hat{\lambda}_n\right)$ les indices $i \in \{1, ..., n-1\}$ tels que

- $|\Delta_i^n X| > v_n \sqrt{\Delta_n}$ si $\beta_n \Delta_n \to 0$,
- $|\Delta_i^n X| > v_n \sqrt{\Delta_n},\ \Delta_i^n X \Delta_{i+1}^n X < 0$ si $\beta_n \Delta_n \asymp 1$.

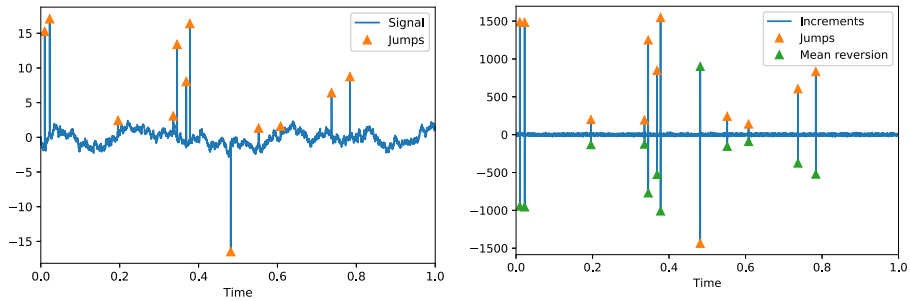
On a le résultat suivant :

Résultat 1 *Avec une probabilité convergeant vers 1,* $\hat{\lambda}_n = N_1$ *et* $T_q \in ((\mathcal{I}_n(q) - 1)\Delta_n,\ \mathcal{I}_n(q)\Delta_n]$ *avec* T_q *les instants de sauts.*



(a) X. (b) $\frac{\Delta_i^n X}{\Delta_n^{0.49}}$.

Figure 8: *Détection de sauts dans le cas* $\beta_n \Delta_n = 0.3$.



(a) X. (b) $\frac{\Delta_i^n X}{\Delta_n^{0.49}}$.

Figure 9: *Détection de sauts dans le cas* $\beta_n \Delta_n = 1$.

Une fois que les temps de pics sont identifiés, un estimateur naturel de leur taille est $\Delta^n_{\mathcal{I}_n(q)} X$ pour $q \in \{1, .., \hat{\lambda}_n\}$ qui est égal à $\Delta X_{T_q} e^{-\beta_n (T_q - \mathcal{I}_n(q)\Delta_n)}$ plus une erreur avec ΔX_{T_q} la taille du $q^{\text{ème}}$ saut. Le terme $e^{-\beta_n (T_q - \mathcal{I}_n(q)\Delta_n)}$ inclut un biais dû au fait que le saut apparaît entre les instants $(\mathcal{I}_n(q) - 1)\Delta_n$ et $\mathcal{I}_n(q)\Delta_n$ et a déjà commencé à retourner vers 0 au moment de l'observation. Si $\beta_n \Delta_n \to 0$, ce terme est égal à $1 + O(\beta_n \Delta_n)$ et il est possible d'identifier la taille du saut. Si $\beta_n \Delta_n \asymp 1$, il n'est pas possible de l'identifier car nous n'avons pas accès au temps de saut exact. Cependant, si $\lambda_n \to \infty$, il est possible de moyenner cette erreur et l'on a le résultat suivant :

$$\frac{m\beta_n \Delta_n}{(1 - e^{-m\beta_n \Delta_n})\hat{\lambda}_n} \sum_{q=1}^{\hat{\lambda}_n} \left(\Delta^n_{\mathcal{I}_n(q)} X\right)^m \to \int_{\mathbb{R}} x^m \nu(dx)$$

en probabilité pour tout entier $m > 0$ tel que $\int_{\mathbb{R}} x^m \nu(dx) < \infty$. Cet estimateur diffère de l'estimateur classique $\frac{1}{\hat{\lambda}_n} \sum_{q=1}^{\hat{\lambda}_n} \left(\Delta^n_{\mathcal{I}_n(q)} X\right)^m$ par le terme $\frac{m\beta_n \Delta_n}{(1 - e^{-m\beta_n \Delta_n})}$: cette correction correspond

à l'erreur moyenne du biais causée par le retour à la moyenne. Pour estimer les moments de la distribution des sauts, il est nécessaire d'avoir un estimateur consistant de β_n.

Pour estimer le paramètre β_n, nous considérons la pente du processus après un saut

$$-\mathrm{sgn}\left(\Delta^n_{\mathcal{I}_n(q)}\right)\Delta^n_{\mathcal{I}_n(q)+1}X$$

qui est de l'ordre de $\left(1-e^{-\beta_n\Delta_n}\right)|\Delta X_{T_q}|$ avec sgn la fonction signe. Moyenner ces quantités sur tous les sauts permet de moyenner le bruit causé par le mouvement Brownien. En divisant par une approximation de $\sum_{q=1}^{\hat{\lambda}_n}|\Delta X_{T_q}|$ et en prenant le logarithme de la moyenne, nous obtenons l'estimateur suivant :

$$\widehat{\beta}_n=\frac{-1}{\Delta_n}\log\left(\left(1+\frac{\sum_{q=1}^{\hat{\lambda}_n}\mathrm{sgn}(\Delta^n_{\mathcal{I}_n(q)}X)\left(\Delta^n_{\mathcal{I}_n(q)+1}X+2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{\mathcal{I}_n(j)}X\right)}{\sum_{q=1}^{\hat{\lambda}_n}|\Delta^n_{\mathcal{I}_n(q)}X|}\mathbf{1}_{\hat{\lambda}_n>0}\right)\vee\Delta_n\right).$$

où un terme de correction $2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{\mathcal{I}_n(j)}X$ est ajouté pour éviter un terme de biais d'ordre $\frac{\lambda_n}{\beta_n}$.

Résultat 2 *L'erreur* $\frac{\widehat{\beta}_n-\beta_n}{\beta_n}$ *est égale à* $O_p\left(\lambda_n\Delta_n+\min\left(\frac{\lambda_n}{\beta_n},\beta_n^{-\frac{1}{2}}\right)+\left(\beta_n\sqrt{\lambda_n\Delta_n}\right)^{-1}\right)$ *et un théorème central limite est donné sous certaines hypothèses.*

Le premier terme d'erreur est un terme de bias, le second est dû au bruit causé par les sauts qui ne sont pas encore retournés à 0 complètement et le troisième est dû au mouvement Brownien. Pour que l'estimateur soit consistent, il faut que l'erreur du mouvement Brownien $\left(\beta_n\sqrt{\lambda_n\Delta_n}\right)^{-1}\to 0$ comme les deux premiers termes convergent vers 0. Cette condition peut être expliquée par le fait que la taille du bruit dû à un incrément du mouvement Brownien est de l'ordre de $\sqrt{\Delta_n}$, celle dû au moyennage de ces incréments arrivant après un saut est de l'ordre de $\sqrt{\frac{\Delta_n}{\lambda_n}}$ et la taille d'un incrément de retour à la moyenne est de l'ordre de $\beta_n\Delta_n$. Pour que le bruit soit négligeable devant l'estimateur qui est la pente, il faut avoir $\sqrt{\frac{\Delta_n}{\lambda_n}}=o\left(\beta_n\Delta_n\right)$, ce qui correspond à notre condition.

Application à la modélisation des pics dans les prix de l'électricité Généralisons le modèle deux facteurs dans le but d'intégrer un terme de pic. Le prix forward est modélisé sous la probabilité historique par

$$f\left(t,T\right)=\int_0^t\mu_s ds+f^c\left(t,T\right)+\int_0^t\int_{\mathbb{R}}xe^{-\beta(T-s)}\underline{p}\left(ds,dx\right)$$

où

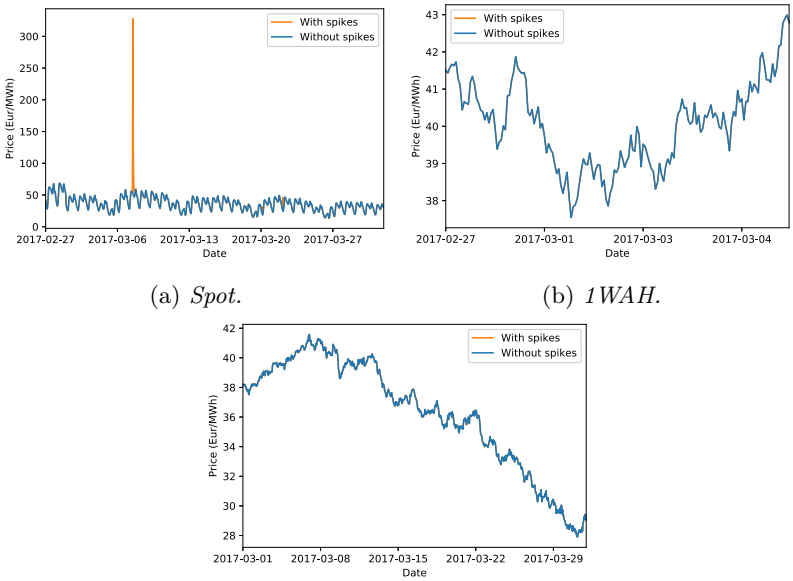
$$df^c\left(t,T\right)=f^c\left(t,T\right)\left(\sigma_s e^{-\alpha(T-t)}dW_s+\sigma_l dW_l\right)$$

correspond à la dynamique classique des forward. Le prix spot est donc égal à

$$S_t=\int_0^t\mu_s ds+S_t^c+\int_0^t\int_{\mathbb{R}}xe^{-\beta(t-s)}\underline{p}\left(ds,dx\right)$$

où S^c est le modèle spot équivalent au modèle deux facteurs et est une semimartingale. Nous aboutissons à un modèle simple sur le forward et le spot, qui ne diffère que très légèrement des

modèles classiques par l'ajout d'un terme. Notre procédure d'estimation du facteur de pics est utilisé sur les données spot françaises, allemandes et australiennes. Lors du calcul des produits $f(t,T,\theta)$, la composante dû à f^{β} est d'ordre $\frac{\lambda}{\beta}$ ce qui est négligeable devant la partie continue du forward. Le facteur de pics a très peu d'impact sur la partie forward, ce qui est consistant avec les données. Les paramètres de la partie continue des prix spot et forward peuvent donc être calibrés sur les produits forwards de la même manière que sans la composante de pics. Nous calibrons la partie continue sur les données forward françaises. Figure 10 représente une simulation des modèles avec et sans pics avec les paramètres du marché français ; le facteur pic est bien négligeable pour les produits forward. La composante de pics n'a qu'un impact sur le prix spot. Nous montrons que la modélisation des pics a un fort impact sur la valorisation d'options strip de la forme $\int_0^T (S_t - K)^+ \, dt$.



(a) *Spot.* (b) *1WAH.*

(c) *1MAH.*

Figure 10: *Simulation de différents produits dans un modèle deux facteurs avec et sans pics entre le* 27 *Février 2017 et le* 31 *Mars 2017. Nous illustrons le spot, le 1WAH commençant le* 27 *Février 2017 et le 1MAH commençant le* 01 *Mars 2017.*

4 Troisième partie : Estimation non paramétrique de l'intensité d'un processus de Poisson doublement stochastique fonction d'une covariable

Dans cette partie, nous nous intéressons à une semimartingale continue X et à un processus de Poisson doublement stochastique N définis sur un espace de probabilité filtré $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P}\right)$ commun. La loi du processus de Poisson doublement stochastique est entièrement déterminé par sa fonction d'intensité λ qui est aussi un processus stochastique. Modéliser la dépendance entre N

et X est donc équivalent à modéliser la dépendance entre λ et X. Nous supposons que λ et X sont observés de manière continue sur un horizon de temps $[0,T]$. Nous supposons que

$$\lambda_s = nq\left(X_s\right),\ s \in [0,T]$$

où $n \in \mathbb{N},\ n \geq 1$ correspond à une asymptotique. Conditionnellement à $(X_t)_{0\leq t\leq T}$, N est un processus de Poisson inhomogène d'intensité $nq\left(X_t\right)$ à l'instant t. Notre objectif est d'estimer la fonction q sur un intervalle donné I de $\mathbb{R}$. La littérature sur l'estimation non paramétrique de l'intensité d'un processus de Poisson est large. Le cas le plus simple est le cas d'un processus de Poisson inhomogène où l'intensité est une fonction déterministe du temps : [50; 51] utilisent des techniques de sélection de modèle et des estimateurs par projection dans un cadre non asymptotique. Une fonction de pénalité est proposé pour sélectionner le modèle optimal. [26; 15] utilisent des estimateurs à noyau dans un cadre asymptotique ; dans [15], une méthode de sélection de la fenêtre est proposée. Les processus de Poisson doublement stochastiques les plus utilisés sont ceux de Cox et d'Aalen [1; 23]. La fonction d'intensité du modèle de Aalen est de la forme $\alpha_t Y_t$ avec α_t une fonction du temps et Y_t un processus stochastique. Celle du modèle de Cox est de la forme $\alpha_t \exp\left(\beta^T Z\right)$ avec Z un vecteur aléatoire ou un processus stochastique dans certains cas (voir [45] par example). Encore une fois, des estimateurs par projection sont utilisés par [22] et des estimateurs à polynômes locaux qui généralisent les estimateurs à noyau sont utilisés par [21]. Dans [60], l'intensité d'un processus de Poisson doublement stochastique est estimé comme une fonction du temps avec des estimateurs à noyau dans un cadre asymptotique. A notre connaissance, les méthodes d'estimation non paramétrique dans notre cadre sont moins répandues dans la littérature à part pour [56] qui propose un estimateur à noyau de la fonction q dans le cas où T tend vers ∞ et quand X satisfait certaines conditions, qui peuvent être par exemple la stationnarité de X. Nous voulons travailler dans un cadre plus général où X n'est pas forcément stationnaire.

Tout d'abord, pour estimer la fonction q en un point $x \in I$, il faut que X soit proche de x un certain nombre de fois avant l'instant T. Un moyen d'évaluer le temps passé par X autour de x quand X est une semimartingale est le temps local l_T^x. Nous considérons le temps local naturel de X qui est la mesure vérifiant la formule de temps d'occupation

$$\int_0^t f\left(X_s\right) ds = \int_{\mathbb{R}} f\left(x\right) l_t^x dx,\ 0 \leq t \leq T$$

pour n'importe quelle fonction mesurable f sur $\Omega \times \mathbb{R}$. Il diffère du temps local classique utilisé dans la littérature où l'intégration se fait sur le terme de gauche par rapport à $d\langle X\rangle_s$[49, Chapter 6] mais les deux sont liés. l_T^x peut aussi être défini par

$$\lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^T \mathbf{1}_{|X_s - x| \leq \epsilon} ds.$$

Si nous écrivons $X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$, une condition suffisante pour l'existence de l_T^x est $\inf\limits_{s\in[0,T]} \sigma_s \geq \underline{\sigma}$ presque sûrement avec $\underline{\sigma} > 0$ une constante. En ajoutant la condition $\mathbb{E}\left(\int_0^T |\mu_s| ds + \sup\limits_{0\leq t\leq T} |\int_0^T \sigma_s dW_s|\right) < \infty$, nous avons $\mathbb{E}\left(\sup\limits_{x\in\mathbb{R}} l_T^x\right) < \infty$ qui est nécessaire dans cette partie. Tous ces résultats peuvent être obtenus facilement en utilisant [49, Exercise 1.15] et [7, Equation $(\mathrm{III})_\gamma$]. Nous considérons aussi le cas dégénéré $X_t = t$ correspondant au processus de Poisson inhomogène. Nous travaillons donc sous l'une des hypothèses suivantes :

(i) $\inf_{0\leq s\leq T}\sigma_s \geq \underline{\sigma}$ avec $\underline{\sigma} > 0$ une constante déterministe et

$$\mathbb{E}\left(\int_0^T |\mu_s|ds + \sup_{0\leq t\leq T}|\int_0^T \sigma_s dW_s|\right) < \infty,$$

(ii) $X_t = t$ pour tout t dans $[0,T]$.

Pour estimer la fonction q au point $x \in I$, il est donc nécessaire d'avoir $l_T^x > 0$. Nous choisissons de travailler conditionnellement à l'évènement $D(I,\nu)$ défini par

$$D(I,\nu) = \{\omega \in \Omega,\ \inf_{x\in I} l_T^x(\omega) \geq \frac{\nu T}{|I|}\}$$

avec $\nu \in (0,1]$. The cadre est le même que celui de [33] dans le contexte de l'estimation non paramétrique d'une fonction de volatilité $\sigma(Y)$ d'une diffusion Y.

Soient K une fonction noyau positive à support borné $[-1,1]$, $\|K\|_\infty = \sup_{x\in\mathbb{R}} K(x) < \infty$, $K_h(x) = h^{-1}K\left(\frac{x}{h}\right)$ pour $x \in \mathbb{R}$ avec $h > 0$ le paramètre de fenêtre. Nous considérons l'estimateur à polynômes locaux de q de degré m pour $h > 0$ et $x \in \mathbb{R}$

$$\hat{q}_h(x) = \frac{1}{n}\int_0^T w\left(x,h,\frac{X_s - x}{h}\right) K_h(X_s - x)\mathbf{1}_{X_s\in I}dN_s$$

avec

$$U(x) = \left(1, x, \frac{x^2}{2!}, .., \frac{x^m}{m!}\right)^T,\ w(x,h,z) = U^T(0)\,B(x,h)^{-1}\,U(z)\,\mathbf{1}_{B(x,h)\in\mathcal{S}_{m+1}^+},\ z \in \mathbb{R}$$

et

$$B(x,h) = \int_0^T U\left(\frac{X_s - x}{h}\right)U^T\left(\frac{X_s - x}{h}\right)K_h(X_s - x)\mathbf{1}_{X_s\in I}ds.$$

Si $B(x,h) \in \mathcal{S}_{m+1}^+$, cet estimateur est égal à $U^T(0)\,\hat{\theta}_h(x)$ avec

$$\begin{aligned}\hat{\theta}_h(x) = \underset{\theta\in\mathbb{R}^{m+1}}{argmin}\ &-\frac{2}{n}\theta^T\int_0^T U\left(\frac{X_s - x}{h}\right)K_h(X_s - x)\mathbf{1}_{X_s\in I}dN_s \\ &+ \theta^T\int_0^T U\left(\frac{X_s - x}{h}\right)U^T\left(\frac{X_s - x}{h}\right)K_h(X_s - x)\mathbf{1}_{X_s\in I}ds\theta.\end{aligned}$$

Soit q_h l'espérance conditionnelle de $\hat{q}_h$ sachant X :

$$q_h(x) = \int_0^T w\left(x,h,\frac{X_s - x}{h}\right)K_h(X_s - x)\mathbf{1}_{X_s\in I}q(X_s)\,ds.$$

Sur l'évènement $D(I,\nu)$, si il existe $\Delta > 0$ et $K_{\min} > 0$ tels que $K(x) \geq K_{\min}\mathbf{1}_{|x|\leq\Delta}$ pour $x \in \mathbb{R}$, nous prouvons que $B(x,h) \in \mathcal{S}_{m+1}^+$ pour $x \in I$ et $0 < h \leq \frac{2}{3}\Delta|I|$.

Pour évaluer les performances de notre estimateur, nous choisissons de travailler avec la perte quadratique intégrée sur I, conditionnellement à l'évènement $D(I,\nu)$, c'est-à-dire avec la quantité $\mathbb{E}\left(\|q - \hat{q}_h\|_I^2|D(I,\nu)\right)$ avec $\|f\|_I^2$ égal à $\int_I f(x)^2\,dx$. Nous voulons répondre aux questions suivantes :

Question 1 *Comment choisir la fenêtre h de manière optimale ?*

Question 2 *Quelle est la vitesse de convergence de notre estimateur et est-elle optimale ?*

Question 3 *Est-ce que la fonction q appartient à une famille paramétrique ?*

La Question 1 est centrale car le paramètre de fenêtre a un impact sur la qualité de notre estimateur et la Question 2 permettra de vérifier sa qualité en terme de vitesse de convergence. La Question 3 a des objectifs opérationnels : il est plus facile en terme de compréhension et de modélisation de travailler avec une fonction paramétrique plutôt qu'avec une fonction non paramétrique.

La fonction de perte peut être écrite comme la somme d'un terme de biais $\mathbb{E}\left(\|q-q_h\|_I^2|D\left(I,\nu\right)\right)$ qui décroît avec h et d'un terme de variance $\mathbb{E}\left(\|q_h-\hat{q}_h\|_I^2|D\left(I,\nu\right)\right)$ qui croît avec h. La fenêtre h minimisant la somme du biais et de la variance est donc optimale dans le sens où la fonction de perte est minimisée. Cependant, cette fenêtre optimale dépend de q que nous ne connaissons pas et est appelée fenêtre oracle. Nous voulons déterminer une fenêtre h telle que la valeur de la fonction de coût est proche de celle de l'oracle. Le même problème existe pour l'estimation de la densité lorsque les observations sont des variables aléatoires i.i.d., voir la discussion dans [55, Section 1.8].

Une solution est de proposer des estimateurs non biaisés des termes de biais et de variance et de choisir la fenêtre h minimisant la somme des deux. Alors qu'un estimateur de la variance est facile à trouver, le problème principal est d'approcher le terme de biais. Dans le cas i.i.d., le terme de variance est déterministe et connu mais dans notre cas, un estimateur non biaisé est donné par

$$\hat{V}_h=\frac{1}{n^2}\int_0^T\int_I\left(w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)\right)^2\mathbf{1}_{X_s\in I}dxdN_s.$$

Concernant l'estimateur du terme de biais, dans un contexte d'observations i.i.d. avec un estimateur à noyau, [37] propose d'approcher le biais $\|q-q_h\|_I^2$ par $\|q-q_{h_{\min}}\|_I^2$ avec $h_{\min}$ suffisamment petit. Si $h_{\min}\to 0$, le biais $\|q_{h_{\min}}-q\|_I^2\approx 0$ et $\|q_{h_{\min}}-q_h\|_I^2\approx\|q-q_h\|_I^2$. Cette méthode est une extension de la méthode classique de Goldenshluger Lepski [30; 39]. Nous adaptons la méthode de [37] au cadre de processus de Poisson mais aussi au cadre d'estimateurs à polynômes locaux. Un estimateur non biaisé de $\|q_{h_{\min}}-q_h\|_I^2$ est donné par

$$\|\hat{q}_{h_{\min}}-\hat{q}_h\|_I^2-\hat{V}_h-\hat{V}_{h_{\min}}+2\hat{V}_{h,h_{\min}}$$

avec $\hat{V}_{h,h_{\min}}$ égal à

$$\frac{1}{n^2}\int_0^T\int_I w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)w\left(x,h_{\min},\frac{X_s-x}{h_{\min}}\right)K_{h_{\min}}\left(X_s-x\right)\mathbf{1}_{X_s\in I}dxdN_s.$$

Au final, nous sélectionnons la fenêtre suivante

$$\hat{h}=\underset{h\in\mathcal{H}}{argmin}\,\|\hat{q}_{h_{\min}}-\hat{q}_h\|_I^2-\hat{V}_h+2\hat{V}_{h,h_{\min}}+\lambda\hat{V}_h$$

parmi un ensemble fini $\mathcal{H}$ inclus dans $(0,\infty)$ avec $\lambda>0$ un hyperparamètre choisi par le statisticien. Nous supposons que $\min\mathcal{H}=h_{\min}\geq\frac{\|K\|_\infty\|K\|_1|I|}{n}$, avec $\|K\|_1=\int_{\mathbb{R}}|K\left(u\right)|du$ et $\max\mathcal{H}\leq\frac{2}{3}|I|\Delta$. Ce choix de fenêtre amène au résultat suivant.

Résultat 1 *Supposons* $\sup_{x\in I} q(x) < \infty$. *La perte* $\mathbb{E}\left(\|q-q_h\|_I^2 | D(I,\nu)\right)$ *est bornée par la somme de*

$$\left(\lambda \vee \frac{1}{\lambda} + O\left(\log(n)^{-1}\right)\right) \min_{h\in\mathcal{H}} \mathbb{E}\left(||\hat{q}_h - q||_I^2 | D(I,\nu)\right)$$

et

$$O(\log(n))\,\mathbb{E}\left(||q_{h_{\min}} - q||_I^2 | D(I,\nu)\right) + \frac{1}{\nu^2} O\left(\frac{\log(n\vee|\mathcal{H}|)^6}{n}\right).$$

Le Résultat 1 est une inégalité oracle qui est déduite des deux inégalités de concentration [51, Equation (2.2)] et [34, Theorem 4.2]. Le terme d'erreur d'ordre $\log(n)\,\mathbb{E}\left(||q_{h_{\min}} - q||_I^2 | D(I,\nu)\right)$ est dû à l'approximation de $\|q-q_h\|_I^2$ par $\|q-q_{h_{\min}}\|_I^2$. L'erreur provenant de ce terme dépend de la régularité de q. Le terme $\frac{1}{\nu^2}$ correspond à l'erreur due à la quantité d'observations de X dans I et si celle-ci est petite, cela entraîne plus d'erreurs. Quand n est grand, et si $\lambda = 1$, notre choix de fenêtre donne des valeurs de la perte proches de la perte optimale si $\log(n)\,\mathbb{E}\left(||q_{h_{\min}} - q||_I^2 | D(I,\nu)\right)$ est suffisamment petit.

Le Résultat 1 ne donne pas d'informations sur la qualité de notre estimateur. Pour $\rho,\ \beta,\ L > 0$, soit $\Lambda_{\rho,\beta} = \{f : I \to \mathbb{R} :\ f(x) \geq \rho,\ \sup_{x\in I} f(x) < \infty\} \cap \Sigma(\beta, L, I)$ où $\Sigma(\beta, L, I)$ est la classe de Hölder d'ordre β sur I avec une borne L. Les Résultats 2 et 3 évaluent la performance de $\hat{q}_{\hat{h}}$ au sens minimax sur $\Lambda_{\rho,\beta}$ et répondent à la Question 2. Nous rappelons que m est le degré du polynôme de l'estimateur.

Résultat 2 *La suite* $\mathbb{E}\left(\varphi_n^2 \|q - \hat{q}_{\hat{h}}\|_I^2 | D(I,\nu)\right)$ *est bornée uniformément sur* $\Lambda_{\rho,\beta}$ *avec* φ_n *égal à* $n^{\frac{-\beta}{2\beta+1}}$ *si* $m \geq \lfloor\beta\rfloor$ *et* $n^{\frac{-m}{2m+1}}$ *si* $m < \lfloor\beta\rfloor$.

Résultat 3 *La vitesse de convergence* $n^{\frac{-\beta}{2\beta+1}}$ *est une borne inférieure au sens minimax.*

Si $m \geq \lfloor\beta\rfloor$, notre estimateur est donc optimal au sens minimax avec une vitesse de convergence $n^{\frac{-\beta}{2\beta+1}}$.

Pour répondre à la Question 3, nous proposons de tester

$$\begin{cases} H_0 :\ \exists\theta_0 \in \Theta,\ q = g_\theta \text{ contre} \\ H_1 :\ \forall\theta \in \Theta,\ q \neq g_\theta \end{cases}$$

avec $\Theta \subset \mathbb{R}^d$, $d \geq 1$ et g_θ une fonction paramétrisée par θ. Considérons le constraste

$$\begin{aligned} M_n(\theta) = \|\hat{q}_{\hat{h}}(\cdot) &- \int_0^T w\left(\cdot, \hat{h}, \frac{X_s - \cdot}{\hat{h}}\right) K_{\hat{h}}(X_s - \cdot)\, \mathbf{1}_{X_s\in I} g_\theta(X_s)\, ds\|_I^2 \\ &- \frac{1}{n^2}\int_I \int_0^T w^2\left(x, \hat{h}, \frac{X_s - x}{\hat{h}}\right) K_{\hat{h}}^2(X_s - x)\, \mathbf{1}_{X_s\in I} dN_s dx. \end{aligned}$$

Le deuxième terme est un terme de correction du biais qui permet d'avoir un estimateur non biaisé asymptotiquement. Le premier terme mesure la distance entre $\hat{q}_{\hat{h}}$ qui est un estimateur de q sous les deux hypothèses et g_θ, ou plutôt une version biaisé de ce dernier permettant d'éviter un terme

de biais dans la distance. Ce contraste converge en probabilité vers $\|q - g_\theta\|_I^2$ et un estimateur naturel de θ_0 sous H_0 est

$$\hat{\theta}_n = \inf_{\theta \in \Theta} M_n(\theta).$$

Un moyen de tester H_0 est de regarder $M_n\left(\hat{\theta_n}\right)$ qui est petit sous H_0 mais qui diverge sous H_1.

Résultat 4 *Sous H_0, $\hat{\theta}_n$ converge vers θ_0 à la vitesse $n^{-\frac{1}{2}}$ et une région critique du test au niveau α est*

$$|M_n\left(\hat{\theta}_n\right)| \geq \hat{c}(\alpha) = n^{-1}\hat{h}^{-\frac{1}{2}}\sqrt{\hat{V}_n}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

où

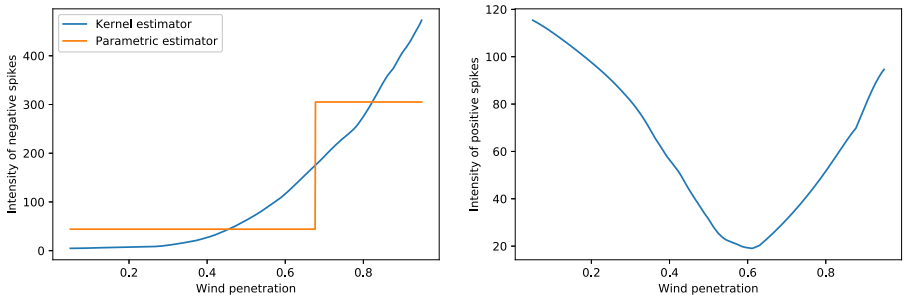
$$\hat{V}_n = C(K)\int_I \left(\frac{g_{\hat{\theta}_n}(y)}{\int_0^T K_{\hat{h}}(y - X_s)\mathbf{1}_{X_s \in I} ds}\right)^2 dy,$$

$C(K)$ est une constante dépendant uniquement de K et Φ est la fonction de répartition d'une variable aléatoire $\mathcal{N}(0,1)$.

Un théorème central limite est aussi donné pour $\hat{\theta}_n$. Le Résultat 4 indique que $M_n\left(\hat{\theta}_n\right)$ converge vers 0 en probabilité à la vitesse $n^{-1}\hat{h}^{-\frac{1}{2}}$ sous H_0 et vers ∞ sous H_1. Ce test est similaire à celui de [4] utilisé pour tester si le drift et la diffusion d'une volatilité appartiennent à une certaine famille paramétrique.

Application à la modélisation de la dépendance entre les prix spot de l'électricité et la production éolienne Suivant les idées de [58], nous étudions la dépendance entre le prix spot de l'électricité et l'indice de pénétration éolien en Allemagne. L'indice de pénétration éolien est défini comme le ratio entre la production éolienne et la production totale d'électricité. Les données considérées sont le prix spot et l'indice de pénétration éolien horaires Allemands entre les années 2012 et 2016, les deux années étant incluses. Notre intuition est qu'un indice de pénétration fort entraîne des pics de prix négatifs dans les prix spot de l'électricité. Les pics sont modélisés par une processus de Poisson avec un fort retour à la moyenne, comme dans le Chapitre 3. La méthode du Chapitre 3 est alors utilisée pour détecter les pics dans la série temporelle des prix. Nous distinguons les pics négatifs des pics positifs et nous estimons l'intensité du processus de comptage comme une fonction de l'indice de pénétration pour chaque processus en utilisant notre estimateur à polynômes locaux d'ordre 0, c'est-à-dire à noyau, voir Figure 11a et Figure 11b. Nous testons si les deux fonctions d'intensité sont constantes avec notre test paramétrique : le test n'est pas rejeté pour les pics positifs mais l'est pour les pics négatifs avec un niveau de confiance à 95%. Nous trouvons aussi que l'intensité des pics négatifs est une fonction croissante de l'indice de pénétration éolien.

Basé sur ces résultats, le prix spot est modélisé comme la somme d'une fonction de saisonnalité, d'un processus continu autorégressif d'ordre 24 et de deux processus de pics : un pour les pics positifs et un pour les négatifs, les deux ayant le même paramètre de retour à la moyenne. L'intensité du processus de pics positifs est modélisée par une constante mais celle du processus de pics négatifs possède deux états : un état de faible intensité pour les faibles valeurs de l'indice et un état de forte intensité pour les fortes valeurs de l'indice, voir Figure 11a. L'indice de pénétration éolien prenant ses valeurs entre 0 et 1, nous modélisons son logit par la somme d'une fonction de saisonnalité et



(a) *Pour les pics négatifs.* (b) *Pour les pics positifs.*

Figure 11: *Estimateurs à noyau de l'intensité des pics de prix spot fonction de la pénétration éolienne.*

d'un processus continu autorégressif d'ordre 24. Des méthodes d'estimation sont données pour les deux modèles.

Pour étudier l'impact de notre modèle, nous considérons le point de vue d'un fournisseur d'électricité achetant de l'électricité à un producteur éolien à un prix fixe K. Le producteur éolien produit $Q\%$ de la production éolienne totale. Les revenus du fournisseur d'électricité sur une période T sont donc égaux à $Q\int_0^T C_t W P_t \left(S_t - K\right) dt$ avec C_t la consommation totale. La Value at Risk et l'Expected Shortfall de ce modèle sont comparés à celles du modèle où l'intensité des pics de prix négatifs est constante : la différence est signifiante entre les deux modèles.

5 Structure de la thèse

Cette thèse est composée de cinq chapitres basés sur les travaux suivants :

- [Chapitre 1] *On the control of the difference between two Brownian motions : a dynamic copula approach*, publié dans *Dependence Modeling.*

- [Chapitre 2] *On the control of the difference between two Brownian motions : an application to energy markets modeling*, publié dans *Dependence Modeling.*

- [Chapitre 3] *Estimation of a fast mean reverting jump process with application to spike modeling in electricity prices*, travail en collaboration avec O. Féron et M. Hoffmann.

- [Chapitre 4] *Local polynomial estimation of the intensity of a doubly stochastic Poisson process.*

- [Chapitre 5] *A joint model for electricity and wind penetration with dependence in the electricity spikes*, travail en collaboration avec A. Veraart, soumis dans *Forecasting and Risk Management for Renewable Energy 2017 : Conference proceedings.*

Bibliography

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[2] René Aid, Luciano Campi, Adrien Nguyen Huu, and Nizar Touzi. A structural risk-neutral model of electricity prices. *International Journal of Theoretical and Applied Finance*, 12(07) :925–947, 2009.

[3] René Aid, Luciano Campi, and Nicolas Langrené. A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*, 23(3) :387–438, 2013.

[4] Yacine Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial studies*, 9(2) :385–426, 1996.

[5] Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009.

[6] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[7] Martin T Barlow and Marc Yor. Semi-martingale inequalities via the garsia-rodemich-rumsey lemma, and applications to local times. *Journal of functional Analysis*, 49(2) :198–229, 1982.

[8] Fred Espen Benth. Cointegrated commodity markets and pricing of derivatives in a non-gaussian framework. In *Advanced Modelling in Mathematical Finance*, pages 477–496. Springer, 2016.

[9] Fred Espen Benth, Jan Kallsen, and Thilo Meyer-Brandis. A non-gaussian ornstein–uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2) :153–169, 2007.

[10] Fred Espen Benth and Paul C Kettler. Dynamic copula models for the spark spread. *Quantitative Finance*, 11(3) :407–421, 2011.

[11] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3) :1116–1157, 2008.

[12] Fred Espen Benth, Nina Lange, and Tor Age Myklebust. Pricing and hedging quanto options in energy markets. *Journal of Energy Markets*, 2015.

[13] Tomasz R Bielecki, Jacek Jakubowski, Andrea Vidozzi, and Luca Vidozzi. Study of dependence for some stochastic processes. *Stochastic analysis and applications*, 26(4) :903–924, 2008.

[14] Damien Bosc. *Three essays on modeling the dependence between financial assets.* PhD thesis, Ecole Polytechnique X, 2012.

[15] Maria Mori Brooks and J Stephen Marron. Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1) :157–165, 1991.

[16] René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer, 2014.

[17] René Carmona, Michael Coulon, and Daniel Schwarz. Electricity price modeling and asset valuation : a multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2) :167–202, 2013.

[18] René Carmona and Valdo Durrleman. Pricing and hedging spread options. *Siam Review*, 45(4) :627–685, 2003.

[19] René Carmona and Valdo Durrleman. Generalizing the black-scholes formula to multivariate contingent claims. *Journal of computational finance*, 9(2) :43, 2005.

[20] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets : a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4) :313–335, 2005.

[21] Feng Chen, Paul SF Yip, and KF Lam. On the local polynomial estimators of the counting process intensity function and its derivatives. *Scandinavian Journal of Statistics*, 38(4) :631–649, 2011.

[22] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l'institut Henri Poincaré (B)*, volume 47, pages 1171–1196, 2011.

[23] David R Cox. Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer, 1992.

[24] Glenis Crane. Time and space varying copulas. *arXiv preprint arXiv :0812.3208*, 2008.

[25] William F Darsow, Bao Nguyen, and Elwood T Olsen. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4) :600–642, 1992.

[26] Peter Diggle. A kernel method for smoothing point process data. *Applied statistics*, pages 138–147, 1985.

[27] Jean-David Fermanian and Marten Wegkamp. Time dependent copulas. *Preprint INSEE, Paris, France*, 2004.

[28] Olivier Féron and Elias Daboussi. *Commodities, Energy and Environmental Finance*, chapter Calibration of electricity price models, pages 183–207. Springer, 2015.

[29] Maurice J Frank, Roger B Nelsen, and Berthold Schweizer. Best-possible bounds for the distribution of a sum - a problem of kolmogorov. *Probability Theory and Related Fields*, 74(2) :199–211, 1987.

[30] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.

[31] J Michael Harrison and Stanley R Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3) :215–260, 1981.

[32] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates : A new methodology for contingent claims valuation. *Econometrica : Journal of the Econometric Society*, pages 77–105, 1992.

[33] Marc Hoffmann. On estimating the diffusion coefficient : parametric versus nonparametric. In *Annales de l'IHP Probabilités et statistiques*, volume 37, pages 339–372, 2001.

[34] Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer, 2003.

[35] Piotr Jaworski and Marcin Krzywda. Coupling of wiener processes by using copulas. *Statistics & Probability Letters*, 83(9) :2027–2033, 2013.

[36] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical methods for financial markets*. Springer, 2009.

[37] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection : a new method with applications to kernel density estimation. *Sankhya A*, pages 1–38, 2016.

[38] Suzanne S Lee and Per A Mykland. Jumps in financial markets : A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6) :2535–2563, 2007.

[39] Matthieu Lerasle, Nelo Molter Magalhães, and Patricia Reynaud-Bouret. Optimal kernel selection for density estimation. In *High Dimensional Probability VII*, pages 425–460. Springer, 2016.

[40] Eckhard Liebscher. Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99(10) :2234–2250, 2008.

[41] GD Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4) :803–806, 1982.

[42] Cecilia Mancini. Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1) :42–52, 2004.

[43] Cecilia Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2) :270–296, 2009.

[44] Thilo Meyer-Brandis and Peter Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(05) :503–528, 2008.

[45] Susan Allbritton Murphy and Pranab Kumar Sen. Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1) :153–180, 1991.

[46] Katsushi Nakajima and Kazuhiko Ohashi. A cointegrated commodity pricing model. *Journal of Futures markets*, 32(11) :995–1033, 2012.

[47] Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2006.

[48] Andrew J Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2) :527–556, 2006.

[49] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[50] Patricia Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126(1) :103–153, 2003.

[51] Patricia Reynaud-Bouret. Concentration inequalities, counting processes and adaptive statistics. In *ESAIM : Proceedings*, volume 44, pages 79–98. EDP Sciences, 2014.

[52] Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, pages 623–632, 1982.

[53] Carlo Sempi. Coupled brownian motion. In *Combining Soft Computing and Statistical Methods in Data Analysis*, pages 569–574. Springer, 2010.

[54] M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[55] Alexandre B Tsybakov. Introduction to nonparametric estimation, 2009.

[56] Klaus J Utikal. Nonparametric inference for a doubly stochastic poisson process. *Stochastic processes and their applications*, 45(2) :331–349, 1993.

[57] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2) :177–188, 1977.

[58] Almut ED Veraart. Modelling the impact of wind power production on electricity prices by regime-switching lévy semistationary processes. In *Stochastics of Environmental and Financial Economics*, pages 321–340. Springer, 2016.

[59] Shaomin Wu. Construction of asymmetric copulas and its application in two-dimensional reliability modelling. *European Journal of Operational Research*, 238(2) :476–485, 2014.

[60] Tingting Zhang and SC Kou. Nonparametric inference of doubly stochastic poisson process data via the kernel method. *The annals of applied statistics*, 4(4) :1913, 2010.

Chapter 1

On the control of the difference between two Brownian motions: a dynamic copula approach

Abstract

We propose new copulae to model the dependence between two Brownian motions and to control the distribution of their difference. Our approach is based on the copula between the Brownian motion and its reflection. We show that the class of admissible copulae for the Brownian motions are not limited to the class of Gaussian copulae and that it also contains asymmetric copulae. These copulae allow for the survival function of the difference between two Brownian motions to have higher value in the right tail than in the Gaussian copula case. Considering two Brownian motions B_t^1 and B_t^2, the main result is that the range of possible values for $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ with $\eta > 0$ is the same for Markovian pairs and all pairs of Brownian motions, that is $\left[0, 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$ with Φ being the cumulative distribution function of a standard Gaussian random variable.

1 Introduction

1.1 Motivation

Modeling dependence between risks has become an important problem in insurance and finance. An important application in risk management for commodity energy markets is the pricing of multi-asset options, and in particular the pricing of spread options. Spread options are used to model the incomes of a plant, such as coal plant. A review on the spread options and on the pricing and hedging models is done by Carmona [4]. The simplest model used for derivative pricing and hedging on several underlying is the multivariate Black and Scholes model [5]. Each price is modeled by a geometric Brownian motion and the dependence between the different Brownian motions is modeled by a constant correlation matrix. The copula between the Brownian motions when they are linked by correlation is called a Gaussian copula. Copulae have many applications in finance and insurance, especially in credit derivative modeling. For instance, Li [17] used the Gaussian copula to model the dependence between time until default of different financial instruments. For more information on the use of copulae in finance, the reader can refer to [7].

Let X_t be the price of electricity at time t, Y_t the price of coal and H the heat rate (conversion factor) between the two. The income of the coal plant at time t can be modeled by $(X_t - HY_t - K)^+$ where K is a constant and corresponds to a fixed cost (we have neglected the price of carbon emissions). Coal is a combustible used to produce electricity and H is the cost of one unit of coal used to produce one unit of electricity. Thus we expect to have $X_t > HY_t$, i.e. the price of electricity greater than the price of the coal used to produce it, with a probability greater than $\frac{1}{2}$. Let us consider that the two commodities are modeled by an arithmetic Brownian motion with a zero drift under a risk neutral probability $\mathbb{P}$: $X_t = \sigma^X B_t^1$ and $HY_t = \sigma^Y B_t^2$ and we suppose that $\langle dB^1, dB^2 \rangle_t = \rho dt$. The dependence between the two Brownian motions is modeled by a correlation, i.e a Gaussian copula. For $x \in \mathbb{R}$, we have

$$\mathbb{P}\left(X_t - HY_t \geq x\right) = \mathbb{P}\left(X_t - HY_t \leq -x\right)$$

and then, if $x \geq 0$,

$$\mathbb{P}\left(X_t - HY_t \geq x\right) \leq \frac{1}{2}.$$

The distribution of the difference between the two prices is symmetric and moreover, the value of its survival function is limited to $\frac{1}{2}$ in the right tail. We would like to have higher values for this probability in order to enrich our modeling. The modeling of the dependence with a constant correlation does not allow to capture the asymmetry in the distribution of the difference of the prices and limits the values that can be achieved by its survival function. Today, it is common practice to use a factorial model [1] to model prices of commodities which is based on Brownian motions. Marginal models, i.e. when we consider only one commodity at the time, are enough performant for risk management. However, the dependence between them is modeled by a Gaussian copula, which is not enough to capture the asymmetry and the values taken by the survival function of their difference. Sklar's Theorem [24] states that the structure of dependence can be separated from the modeling of the marginals with the copula. Studying the impact of the structure of dependence on the modeling is equivalent to studying the impact of the copula.

Whereas copulae are very useful in a static framework where random variables are modeled, modeling with copulae is much more difficult in a dynamic framework, that is when processes are involved. In a discrete time framework, Patton [21] introduces the conditional copula which is a

copula at time t defined conditionally on the information at time $t-1$. Fermanian and Wegkamp [12] generalize the concept of conditional copula. In a continuous time framework, Darsow et al. [9] consider the modeling of the time dependence by a copula. They give sufficient and necessary conditions for a copula to be the copula of a Markov process $X = (X_t)_{t\geq 0}$ between times t and s, i.e. the copula of (X_t, X_s), using the Chapman-Kolmogorov equation. We are more interested in the space dependence, that is the dependence between two different processes at a given time t. The question is studied by Jaworski and Krzywda [15]. They consider two Brownian motions and they are interested in copulae that make the bivariate process self-similar. They find necessary and sufficient conditions for the copula to be suitable for the Brownian motions deriving the Kolmogorov forward equation. The copula is linked by a local correlation function into a partial derivative equation. Further work has been done in the thesis of Bosc [3] where there are no constraints of self-similarity and it is not only limited to Brownian motions ; a more general partial derivative equation is found. More details about their work are given in Section 2.2. However, conditions for the copula to be suitable for the Brownian motions are very restrictive. An equivalent approach to the copula one is the coupling approach. A coupling of two stochastic processes is a bi-dimensional measure on the product space such that the marginal measures correspond to the ones of the stochastic processes. For more information on coupling, the reader can refer to [6]. One of the most important coupling is the coupling by reflection [18], based on the reflection of the Brownian motion which has some importance in this article.

1.2 Objectives and results

The objective of this article is to control the distribution of the difference between two Brownian motions at a given time t. The distribution of the difference between two Brownian motions B^1 and B^2 can be described by $x \mapsto \mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$, $x \in \mathbb{R}$, $t \geq 0$. If $B_t^1 - B_t^2$ has a continuous cumulative distribution function, this function is the survival function of $B_t^1 - B_t^2$ at point x. In particular, we want to find asymmetric distributions for $B^1 - B^2$ with more weight in the positive part than in the Gaussian copula case, i.e. $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ greater than $\frac{1}{2}$ for a given $\eta > 0$. Since distributions of B_t^1 and B_t^2 are known, we control this distribution with the copula of $\left(B_t^1, B_t^2\right)$. One of the main issues is to work in a dynamical framework ; we then first need to extend the definition of copulae to Markovian diffusions. If we denote by $\mathcal{C}_B$ the set of admissible copulae for Brownian motions, which is properly defined in Section 2.2, our main goal is to study the range of the function

$$\begin{array}{rccc} S_{\eta,t} & : & \mathcal{C}_B & \to & [0,1] \\ & & C & \mapsto & \mathbb{P}_C\left(B_t^1 - B_t^2 \geq \eta\right) \end{array}$$

denoted by $Ran\left(S_{\eta,t}\right)$ with $\mathbb{P}_C$ the probability measure associated to $\left(B^1, B^2\right)$ when $C \in \mathcal{C}_B$ and with $\eta > 0$ and $t \geq 0$ given. Our problem is related to the one consisting in finding bounds of $\mathbb{P}\left(X + Y < \eta\right)$ with X and Y two random variables with given distributions, see [20, Section 6.1] and [13; 22; 19]. In our case, we add the constraint that the copulae are in $\mathcal{C}_B$.

Considering the set of Gaussian copulae, it is easy to prove that $\left[0, \Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right] \subset Ran\left(S_{\eta,t}\right)$ by controlling the correlation between the two Brownian motions with Φ the cumulative distribution function of a standard normal random variable. Furthermore, if we consider the restriction of $S_{\eta,t}$ to the set of Gaussian copulae $S_{\eta,t}|_{C_G^d}$, we have $Ran\left(S_{\eta,t}|_{C_G^d}\right) = \left[0, \Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$, see Proposition 1.6 (i) below.

Our major contribution is to construct a family of dynamic copulae in $\mathcal{C}_B$ that can achieve all the

values between 0 and the supremum of $S_{\eta,t}$ on $\mathcal{C}_B$. We first prove that

$$\sup_{C\in\mathcal{C}_B} S_{\eta,t}\left(C\right) = 2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big)$$

in Proposition 1.6 (ii), implying that the Gaussian copulae can not describe all the values that can be achieved by $S_{\eta,t}$. This supremum is achieved with the copula of the Brownian motion and its reflection, which we call the Reflection Brownian Copula, and which a closed formula is given in Proposition 1.1. In the case where X and Y are two normal random variables $\mathcal{N}\left(0,\sqrt{t}\right)$, the supremum of $\mathbb{P}\left(X-Y\geq\eta\right)$ without constraint on the copula is also equal to $2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big)$, see Proposition 1.5 (ii) below that follows from [20, Section 6.1, Example 6.2]. The supremum for $\mathbb{P}_C\left(B_t^1-B_t^2\geq\eta\right)$ is then the same for Markovian pairs and all pairs of Brownian motions. Deriving a new family of copulae that is described in Proposition 1.3 from the Reflection Brownian Copula, it is possible to achieve all the value between 0 and $2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big)$, which means that

$$Ran\left(S_{\eta,t}\right) = \Big[0, 2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big)\Big];$$

this is the result of Proposition 1.6 (iii). The range of possible values for $\mathbb{P}_C\left(B_t^1-B_t^2\geq\eta\right)$ is the same for Markovian pairs and all pairs of Brownian motions. Copulae used to achieve values in $Ran\left(S_{\eta,t}\right)$ present two states depending on the value of $B_t^1-B_t^2$: one of positive correlation and one of negative one. These copulae are asymmetric and to our knowledge, these are the only asymmetric copulae suitable for Brownian motions available in the literature.

1.3 Structure of the paper

In Section 2, we define the notion of dynamic copulae for Markovian diffusion processes and in particular for the case of two Brownian motions. We show that our definition includes several model of dependence present in the literature such as stochastic correlation models. In Section 3, we compute a copula called the Reflection Brownian Copula based on the dependence between a Brownian motion and its reflection and we derive new families of asymmetric copulae based on this copula. In Section 4, after showing the limitations of modeling the dependence between two random variables with symmetric copulae, we establish the results on the range of the function $S_{\eta,t}$, first in a static framework and then in a dynamical framework with Brownian motions.

2 Markov Diffusion Copulae

In finance and insurance, modeling of two dimensional processes is usually based on a 2 dimensional Brownian motion, that is when the structure of dependence between two 1 dimensional Brownian motions is modeled by a correlation. The copula of the two Brownian motions at a given time then belongs to the class of Gaussian copulae.

Let us recall that a function $C:[0,1]^2\mapsto[0,1]$ is a copula if:

(i) C is 2-increasing, i.e. $C\left(u_2,v_2\right)-C\left(u_1,v_2\right)+C\left(u_1,v_1\right)-C\left(u_2,v_1\right)\geq 0$ for $u_2\geq u_1, v_2\geq v_1$ and $u_1,u_2,v_1,v_2\in[0,1]$,

(ii) $C\left(u,0\right)=C\left(0,v\right)=0,\ u,v\in[0,1]$,

(iii) $C(u,1) = u, C(1,u) = u,\ u \in [0,1]$.

We denote by $\mathcal{C}$ the set of copulae and by $\mathcal{C}_G$ the set of Gaussian copulae completed by the upper and lower Frechet copulae $M(u,v) = \min(u,v)$ and $W(u,v) = \max(u+v-1,0)$ corresponding to the limit cases $\rho = 1$ and $\rho = -1$. $\mathcal{C}_G = \{C \in \mathcal{C} : \exists \rho \in (-1,1), C = C_{G,\rho}\} \cup \{M, W\}$ where $C_{G,\rho}$ denote the Gaussian copula with parameter ρ. We have

$$C_{G,\rho}(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

with Φ the cumulative distribution function of a standard normal random variable and Φ_ρ the cumulative distribution function of a bivariate normal random variable with correlation ρ:

$$\Phi_\rho(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} du dv.$$

In the following, a Gaussian copula will abusively refer to an element of $\mathcal{C}_G$.

In this section, we want to generalize the concept of copula which is adapted for random variables to a dynamical framework. We want to define the notion of copula for Markov diffusions in Section 2.1. In particular, we are interested in copulae suitable for Brownian motions in Section 2.2.

2.1 Definition

In order to work in a dynamical framework, we need to extend the concept of copula to Markovian diffusions. Our definition is based on the work of Bielecki et al. [2] and gives a more general definition.

We recall that if $P = (P_t)_{t\geq 0}$ is a Markovian diffusion solution of the stochastic differential equation

$$dP_t = \mu(P_t)\, dt + \sigma(P_t)\, dW_t,$$

with $W = (W_t)_{t\geq 0}$ a standard Brownian motion, the infinitesimal generator $\mathcal{L}$ of P is the operator defined by

$$\mathcal{L}f(x) = \frac{1}{2}\sigma^2(x) f''(x) + \mu(x) f'(x)$$

for f in a suitable space of functions including $\mathcal{C}^2$.

Definition 1.1 (Admissible copula for Markovian diffusions)**.** *We say that a collection of copula $C = (C_t)_{t\geq 0}$ is an admissible copula for the n real valued Markovian diffusions, $n \geq 2$, $\left(X^i\right)_{1\leq i\leq n}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if there exists a $\mathbb{R}^m$ Markovian diffusion $Z = \left(Z^i\right)_{1\leq i\leq m}$, $m \geq n$, defined on a probability extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

$$\begin{cases} \mathcal{L}\left(Z^i\right) = \mathcal{L}\left(X^i\right), 1 \leq i \leq n, \\ Z_0^i = X_0^i, 1 \leq i \leq n, \\ \text{for } t \geq 0, \text{ the copula of } \left(Z_t^i\right)_{1\leq i\leq n} \text{ is } C_t. \end{cases}$$

The strongest constraint to be admissible is that Z has to be a Markovian diffusion. Without this constraint, all the copulae are admissible. Sempi [23] studies the Brownian motions linked by a copula without this constraint. Definition 1.1 is consistent with the approach of [15] or [3] consisting of modeling dependence by a local correlation function. However, our approach is totally different.

2.2 Brownian motion case

From now on, we work in a 2 dimensional framework and we denote by $\mathcal{C}_B$ the set of admissible copulae for Brownian motions, that is when X^1 and X^2 are Brownian motions. The only well known suitable copulae for Brownian motion are the Gaussian copulae.

We can extend the definition of $\mathcal{C}_G$ to a dynamical framework by defining

$$\mathcal{C}_G^d = \{(C_t)_{t\geq 0} : \ \forall t \in \mathbb{R}^+, \ C_t \in \mathcal{C}_G\} \cap \mathcal{C}_B.$$

It is necessary to take the intersection with $\mathcal{C}_B$ because we do not know if conditions are needed on C_t for the copula to be admissible. We are not interested in this question in this paper. However, we know this intersection is not empty because $\{(C_t)_{t\geq 0} : \exists \rho \in (-1,1)\,, \ \forall t \in \mathbb{R}^+ \ C_t = C_{G,\rho}\} \subset \mathcal{C}_B$. One of our objective is to find copulae that are admissible for Brownian motion but that are not Gaussian copulae.

Jaworski and Krzywda [15] prove that the set of admissible copulae for Brownian motions was not reduced to the Gaussian copulae. By linking local correlation and copula with the Kolmogorov backward equation, they find that a sufficient condition to be admissible is

$$\left| \frac{1}{2} e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}} \frac{\partial^2_{u,u} C(u,v)}{\partial^2_{u,v} C(u,v)} + \frac{1}{2} e^{\frac{\Phi^{-1}(u)^2 - \Phi^{-1}(v)^2}{2}} \frac{\partial^2_{v,v} C(u,v)}{\partial^2_{u,v} C(u,v)} \right| < 1 \quad \forall (t,u,v) \in \mathbb{R}^+ \times [0,1]^2 \tag{1.1}$$

when the copula does not depend on time. In particular, they prove that the extension of the FGM copula $C^{FGM}(u,v) = uv\,(1 + a\,(1-u)\,(1-v))\,, a \in [-1,1]$ in a dynamical framework defined by $C_t(u,v) = C^{FGM}(u,v)\,, \ t \geq 0$, is an admissible copula for Brownian motions. Bosc [3] has also found admissible copulae.

Let us consider two independent Brownian motions B^1 and Z defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Definition 1.1 includes several models for Brownian motions used in the literature.

Deterministic correlation Let us consider a function $t \mapsto \rho(t)$ defined on $\mathbb{R}^+$ with values in $[-1,1]$. Let $B_t^2 = \int_0^t \rho(s)\, dB_s^1 + \int_0^t \sqrt{1-\rho(s)^2} dZ_s$.

B^2 is a Brownian motion and the dynamic copula defined at each time t by the copula of $\left(B_t^1, B_t^2\right)$ is in $\mathcal{C}_B$.

Local correlation Let us consider a function $(x,y) \mapsto \rho(x,y)$ defined on $\mathbb{R}^+$ with values in $[-1,1]$ and measurable. If the stochastic differential equation

$$dB_s^2 = \rho\left(B_t^1, B_t^2\right) dB_s^1 + \sqrt{1 - \rho\left(B_t^1, B_t^2\right)^2} dZ_s$$

has a strong solution, the dynamic copula defined at each time t by the copula of $\left(B_t^1, B_t^2\right)$ is in $\mathcal{C}_B$ by the Lévy characterization of Brownian motion.

Stochastic correlation Let us consider a Markovian diffusion $\rho = (\rho_s)_{s\geq 0}$ independent of $\left(B^1, Z\right)$ locally square integrable and with values in $[-1, 1]$.

We can extend the probability space and the filtration generated by $\left(B^1, Z\right)$. The stochastic process B^2 defined by $B_t^2 = \int_0^t \rho(s)\, dB_s^1 + \int_0^t \sqrt{1-\rho(s)^2} dZ_s$ is a Brownian motion and the dynamic copula defined at each time t by the copula of $\left(B_t^1, B_t^2\right)$ is in $\mathcal{C}_B$.

We can also consider a correlation diffusion driven by B^1, Z and an independent Brownian motion. If the system of stochastic differential equations has a strong solution, the copula is still in $\mathcal{C}_B$.

Contrary to the approaches of Jaworski and Krzywda [15], Bosc [3] or Bielecki et al. [2], Definition 1.1 includes stochastic correlation models. However, we need for the stochastic correlation to be a Markovian diffusion which is not needed in a general case ; the stochastic correlation has only to be progressively measurable.

3 Reflection Brownian Copula

In this section, our objective is to construct Markov Diffusion Copulae defined in Section 2. We construct a new copula based on the reflection of the Brownian motion. We show that the copula between the Brownian motion and its reflection is adapted to a dynamical framework and is a suitable copula for Brownian motions. Furthermore, we give a closed formula of this copula in Section 3.1. To our knowledge, this copula has not been studied in detail and it is the new copula suitable for Brownian motions. We also construct new families of copulae by extension of the Reflection Brownian Copula in Section 3.2.

3.1 Closed formula for the copula

In this section, we study the copula between the Brownian motion and its reflection. Since its reflection is also a Brownian motion, the copula is a good candidate for being in $\mathcal{C}_B$.

Let us consider a filtered probability space $(\Omega,\, \mathcal{F},\, (\mathcal{F}_t)_{t\geq 0},\, \mathbb{P})$ with $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual hypothesis (right continuity and completion) and $B = (B_t)_{t\geq 0}$ a Brownian motion adapted to $(\mathcal{F}_t)_{t\geq 0}$. We denote by $\tilde{B}^h$ the Brownian motion reflection of B on $x = h$ with $h \in \mathbb{R}$, i.e. $\tilde{B}_t^h = -B_t + 2(B_t - B_{\tau^h})\mathbf{1}_{t\geq \tau^h}$ with $\tau^h = \inf\{t \geq 0 : B_t = h\}$. Thus, $\tilde{B}^h$ is a $\mathcal{F}$ Brownian motion according to the reflection principle (see [16, Theorem 3.1.1.2, p. 137]). Proposition 1.1 gives the copula of $\left(B, \tilde{B}^h\right)$.

We recall that $M(u,v) = \min(u,v)$, $W(u,v) = \max(u+v-1, 0)$, $u, v \in [0,1]$ and that Φ denotes the cumulative distribution function of a standard normal random variable.

Proposition 1.1. *Let $h > 0$. The copula of $(B, \tilde{B}^h)$, $\left(C_t^{ref,h}\right)_{t\geq 0}$, is defined by*

$$C_t^{ref,h}(u,v) = \begin{cases} v & \text{if} \quad \Phi^{-1}(u) - \Phi^{-1}(v) \geq \frac{2h}{\sqrt{t}} \\ W(u,v) + \Phi\left(\Phi^{-1}(M(u,1-v)) - \frac{2h}{\sqrt{t}}\right) & \text{if} \quad \Phi^{-1}(u) - \Phi^{-1}(v) < \frac{2h}{\sqrt{t}} \end{cases} \tag{1.2}$$

and $\left(C_t^{ref,h}\right)_{t\geq 0} \in \mathcal{C}_B$. We call this copula the Reflection Brownian Copula.

3.2 Extensions

In this section, we give methods to construct new admissible copulae for Brownian motions from the Reflection Brownian Copula.

Proposition 1.2 and its proof gives an approach to construct different admissible copulae for Brownian motions based on the Reflection Brownian Copula considering a correlated Brownian motion to the reflection of the Brownian motion. Thus, the copula of Proposition 1.2 is the copula of $\left(B, \rho\tilde{B}^h + \sqrt{1-\rho^2}Z\right)$, with $h > 0$, $\rho \in (0,1)$ and $Z = (Z_t)_{t\geq 0}$ a Brownian motion independent from B.

Proposition 1.2. *Let $h > 0$ and $\rho \in (0,1)$. The copula*

$$C_t(u,v) = \begin{cases} \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v) + \frac{2\rho h}{\sqrt{t}}\right) + v - \Phi\left(\Phi^{-1}(v) + \frac{2\rho h}{\sqrt{t}}\right) & \text{if} \quad u \geq \Phi\left(\frac{h}{\sqrt{t}}\right) \\ \Phi_{-\rho}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) + \Phi_\rho\left(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}, \Phi^{-1}(1-v) - \frac{2\rho h}{\sqrt{t}}\right) + & \\ \Phi_\rho\left(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}, \Phi^{-1}(v)\right) - \Phi\left(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}\right) & \text{if} \quad u < \Phi\left(\frac{h}{\sqrt{t}}\right) \end{cases}$$

is in $\mathcal{C}^B$.

Contrary to the Reflection Brownian Copula, this copula is non degenerated in the sense that we have two distinct sources of randomness. Indeed, in the Reflection Brownian Copula case, if we know the trajectory of the Brownian motion, we also know the one of its reflection.

Remark 1.1. *In the case $\rho = 0$, we still have a copula which is the independent copula and then that is in $\mathcal{C}_B$.*

An other way to construct admissible copulae is to consider a random barrier ξ. By enlarging the filtration, the copula of the two processes is an admissible copula and it can be computed by integrating the copula of the Reflection Brownian motion according to the law of the barrier. Proposition 1.3 gives the copula of $\left(B, \tilde{B}^\xi\right)$ which is clearly in $\mathcal{C}^B$ because $\left(B, \tilde{B}^\xi, \xi\right)$ is Markovian.

Proposition 1.3. *Let ξ be a positive random variable with law having a density and $\overline{F}^\xi$ its survival function. The copula*

$$C_t^\xi(u,v) = v - \int_{-\infty}^{\Phi^{-1}(M(1-u,v))} \frac{e^{\frac{-w^2}{2}}}{\sqrt{2\pi}} \overline{F}^\xi\left(\frac{\sqrt{t}}{2}\left(\Phi^{-1}(M(u,1-v)) - w\right)\right) dw$$

is in $\mathcal{C}^B$.

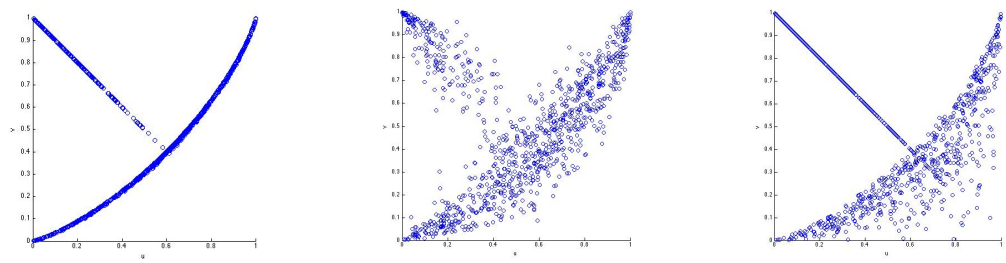
Example 1.1 below gives a copula with closed formula built with the method of Proposition 1.3.

Example 1.1. *Let $\xi \overset{d}{=} h + X$ with $h \in \mathbb{R}$ and X a random variable following an exponential law with parameter $\lambda > 0$. We have $\overline{F}^\xi(x) = \begin{cases} 1 & \text{if} \quad x \leq h \\ e^{-\lambda(x-h)} & \text{if} \quad x > h \end{cases}$ and the copula*

$$C_t^{exp,h,\lambda}(u,v) = W(u,v) + \min\left[\Phi\left(\Phi^{-1}(M(1-u,v)) - \frac{2h}{\sqrt{t}}\right), M(u,1-v)\right] \tag{1.3}$$
$$- \Phi\left(\min\left[\Phi^{-1}(M(1-u,v)) - \frac{2h}{\sqrt{t}}, \Phi^{-1}(M(u,1-v))\right] - \frac{\lambda\sqrt{t}}{2}\right) e^{\lambda h + \frac{\lambda^2 t}{4} + \frac{\lambda\sqrt{t}}{2}\Phi^{-1}(M(u,1-v))}$$

is in $\mathcal{C}_B$.

The methods of Proposition 1.2 and 1.3 could be used simultaneously to construct new classes of admissible copulae. Figure 1.1 represents the Reflection Brownian Copula and some of its extensions.



(a) (b) (c)

Figure 1.1: *The Reflection Brownian Copula $C^{ref,h}$ and some of its extensions at time $t = 1$ with $h = 2$. Figure 1.1a is the Reflection Brownian Copula. Figure 1.1b is the extension considering a Brownian motion correlated to the reflection of the first Brownian with a correlation $\rho = 0.95$, which is the copula of Proposition 1.2. Figure 1.1c is the extension in the case of a random barrier following an exponential law with parameter $\lambda = 2$, which is the copula of Example 1.1.*

Remark 1.2. *In Proposition 1.1, Proposition 1.2 and Proposition 1.3, the given copula is the copula between two Brownian motions B_t^1 and B_t^2 at time t. In Patton [21], the dynamic copula in a discrete time setting is the copula between B_t^1 and B_t^2 knowing the filtration generated by $\left(B_{t-1}^1, B_{t-1}^2, B_{t-2}^1, B_{t-2}^2, ..., B_0^1, B_0^2\right)$. In our case, the copula of $\left(B_t^1, B_t^2\right)$ is conditioned to the values of B_0^1 and B_0^2, which are equals almost surely to 0. One could compute the copula of $\left(B_t^1, B_t^2\right)$ knowing $\left(B_s^1, B_s^2\right)$, but this is not the subject of this paper. We want to study $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ that only requires the copula of $\left(B_t^1, B_t^2\right)$ and not $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta \mid B_s^1, B_s^2\right)$. Furthermore, the dependence between B_t^1, B_t^2, B_s^1 and B_s^2 is implicitely given in the dynamic of (B_1, B_2) that we know. For application, we want to price European spread options of the form $\left(f\left(B_t^1\right) - g\left(B_t^2\right)\right)^+$, that depends only of the law of $\left(B_t^1, B_t^2\right)$ and hence of the copula of $\left(B_t^1, B_t^2\right)$. For American spread options, we know the dynamic of $\left(B^1, B^2\right)$ and we work in a Markovian framework.*

Remark 1.3. *The study of the range of $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta \mid B_s^1, B_s^2\right)$ seems very close to our problem: $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta \mid B_s^1, B_s^2\right) = \mathbb{P}\left(B_t^1 - B_s^1 - \left(B_t^2 - B_s^2\right) \geq \eta - \left(B_s^1 - B_s^2\right) \mid B_1^s, B_2^s\right)$ and $B_1^t - B_1^s$ (resp. $B_2^t - B_2^s$) is a Brownian motion independent from B_1^s (resp. B_2^s). However, we let this problem for further studies.*

4 Control of the distribution of the difference between two Brownian motions

Let B^1 and B^2 be two standard Brownian motions defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_C)$ with $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual hypothesis and where $\mathbb{P}_C$ is the probability measure associated to $\left(B^1, B^2\right)$ and $C = (C_t)_{t\geq 0} \in \mathcal{C}_B$ is the copula of $\left(B^1, B^2\right)$. In this section, we are interested in the distribution of the difference between B^1 and B^2, i.e. the function $x \mapsto$

$\mathbb{P}_C\left(B_t^1 - B_t^2 \geq x\right)$ for $t > 0$ and in particular in the right tail of this distribution, i.e. when $x > 0$. Since the distributions of B^1 and B^2 are known, this function is entirely determined by the copula of $\left(B^1, B^2\right)$. Our goal is to find the range of values that can be achieved by this function at a given $x > 0$. Given $\eta > 0$ and $t \geq 0$, we define the function

$$\begin{array}{rccc} S_{\eta,t} & : & \mathcal{C}_B & \rightarrow & [0,1] \\ & & C & \mapsto & \mathbb{P}_C\left(B_t^1 - B_t^2 \geq \eta\right). \end{array} \tag{1.4}$$

Remark 1.4. *$\mathbb{P}_C$ is a probability measure that verifies*

$$\mathbb{P}_C\left(B_t^1 \leq x, B_t^2 \leq y\right) = C_t\left(\Phi\left(\frac{x}{\sqrt{t}}\right), \Phi\left(\frac{y}{\sqrt{t}}\right)\right)$$

for $x, y \in \mathbb{R}$. However, C does not describe entirely $\mathbb{P}_C$. Indeed, C describe the dependence between B_t^1 and B_t^2 at a given time t but not between B_s^1 and B_t^2 with $s \neq t$ for instance.

Our objective is to control the value of this function at a given time t by controlling the dependence between the two Brownian motions. For this, we first study the range of this function $Ran\left(S_{\eta,t}\right)$. We show that the Reflection Brownian Copula defined in Section 3 and its extensions allow us to control $S_{\eta,t}$ and to achieve all the values in $Ran\left(S_{\eta,t}\right)$. After showing the limitations of symmetric copulae for the control of $S_{\eta,t}$ in Section 4.1, we give a result about $Ran\left(S_{\eta,t}\right)$ in a static case in Section 4.2, i.e. in the case of two Gaussian random variables. Most of results of Section 4.1 and Section 4.2 are classic for the sum of random variables ; we adapt them to the difference case. Finally, we give the main result concerning the range of $S_{\eta,t}$ in Section 4.3.

4.1 Impact of symmetry on $S_{\eta,t}$

In this section, we show that modeling the dependence between two random variables with symmetric copulae limits the values that can be taken by the distribution of the difference between two random variables. It imposes some constraints on this distribution. Using asymmetric copulae is then necessary to control $S_{\eta,t}$. We also show that we can find asymmetric copulae suitable for Brownian motions.

Definition 1.2. *A copula C is symmetric if $C(u,v) = C(v,u)$, $u, v \in [0,1]$. We denote by $\mathcal{C}_s$ the set of symmetric copulae.*

Note that $\mathcal{C}_G \subset \mathcal{C}_s$ with $\mathcal{C}_G$ the set of Gaussian copulae.

If X and Y are two random variables with continuous cumulative distribution functions, we denote by $C^{X,Y}$ the copula of (X,Y). Sklar's Theorem [24] guarantees the existence and the unicity of $C^{X,Y}$. Proposition 1.4 gives properties on the distribution the difference of two random variables if their copula is symmetric.

Proposition 1.4. *Let X and Y be two real valued random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with copula $C^{X,Y}$ and with continuous marginal distribution functions F^X and F^Y. If $F^X = F^Y$ and $C^{X,Y} \in \mathcal{C}_s$ then $\mathbb{P}(X - Y \leq -x) = \mathbb{P}(X - Y \geq x)$.*

We can extend the definition of symmetry and asymmetry to Markov Diffusion Copulae: we denote by $\mathcal{C}_a^d = \{(C_t)_{t\geq 0} : \forall t \geq 0, C_t \in \mathcal{C}_s\}$ the set of symmetric Markov Diffusion Copulae and by $\mathcal{C}_s^d = \{(C_t)_{t\geq 0} : \forall t \geq 0, C_t \in \mathcal{C} \setminus \mathcal{C}_s\}$ the set of asymmetric Markov Diffusion Copulae.

Corollary 1.1. *For $\eta > 0$ and $t > 0$, we have:*

$$Ran\left(S_{\eta,t}|_{\mathcal{C}_s^d}\right) \subset \left[0, \frac{1}{2}\right]$$

with $S_{\eta,t}|_{\mathcal{C}_s^d}$ the restriction of $S_{\eta,t}$ to $\mathcal{C}_s^d$.

Proof If we consider two Brownian motions B^1 and B^2 with dynamic copula $C \in \mathcal{C}_s^d$, we have according to Proposition 1.4: $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \mathbb{P}\left(B_t^1 - B_t^2 \leq -x\right)$. However, $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) + \mathbb{P}\left(B_t^1 - B_t^2 \leq -x\right) \leq 1$ if $x \geq 0$. Then we have the constraint $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) \leq \frac{1}{2}$. □

In particular, since the Gaussian copula is symmetric, it is not possible to obtain asymmetry in the distribution of $B_t^1 - B_t^2$ at each time t when the dependence between two Brownian motions is given by a correlation structure. Limiting the modeling of the dependence to the Gaussian copula or to symmetric copulae makes the distribution of their difference symmetric and limits the value of $S_{\eta,t}$.

Modeling the dependence by an asymmetric copula is then necessary to have higher values than $\frac{1}{2}$ for $S_{\eta,t}$. We have

$$\mathcal{C}_B \cap \mathcal{C}_a^d \neq \emptyset.$$

Indeed, the Reflection Brownian Copula defined in Equation (1.2) is in $\mathcal{C}_B$ and is asymmetric. The set of admissible copulae for Brownian motion is not reduced to the set of Gaussian copulae and furthermore it contains an asymmetric copula which is the Reflection Brownian Copula. Jaworski and Krzywda [15] and Bosc [3] have proven the existence of symmetric suitable copulae for Brownian motions. However, they did not find asymmetric copulae suitable for Brownian motions. We can also show that extensions of the Brownian Reflection Copula defined in Section 3.2 are asymmetric. To our knowledge, these copulae are the only asymmetric copulae suitable for Brownian motions in the literature.

Remark 1.5. *Copulae constructed in Section 3.2 can also be used as a method to construct asymmetric copulae, which is not always evident.*

4.2 The Gaussian Random Variables Case

Let us consider two standard normal random variables X and Y defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P}_C)$ where $\mathbb{P}_C$ is the probability measure associated to the copula C of (X, Y). Since the laws of the marginals of X and Y are fixed, the probability measure only depends on the copula of (X, Y), which justifies the notation $\mathbb{P}_C$. In this section, we study the control of the distribution of the difference $\mathbb{P}_C\left(X - Y \geq \eta\right)$ for a given η. We need to adapt the definition of $S_{\eta,t}$ for the static case, i.e. when the copula are not dynamic. We define the function

$$\begin{array}{rccc} \tilde{S}_\eta & : & \mathcal{C} & \to & [0,1] \\ & & C & \mapsto & \mathbb{P}_C\left(X - Y \geq \eta\right) \end{array}$$

for a given $\eta > 0$.

Remark 1.6. *$\mathbb{P}_C$ is defined by $\mathbb{P}_C\left(X \leq x, Y \leq y\right) = C\left(\Phi\left(x\right), \Phi\left(y\right)\right)$ for $x, y \in \mathbb{R}$.*

In particular, we look for an upper bound of $\tilde{S}_\eta$. Lower bound is trivial and is achieved by the copula $M(u,v) = \min(u,v)$. Note that this copula is equivalent to having correlation 1 between the two random variables and corresponds to a case of comonotonicity. The problem is similar to the one consisting in finding bounds on the distribution of the sum. Makarov [19] finds bounds on the cumulative distribution function of the sum of two random variables at a given point given the marginals. Rüschendorf [22] proves this result using optimal transport theory. Frank et al. [13] prove the same result using copulae and find a copula that achieves the bound. Furthermore, the results are extended to dimensions greater than 2 and to the cumulative distribution function of $L(X,Y)$ where L is a non decreasing continuous function in X and Y with X and Y two random variables. Finding these bounds have several applications in finance and insurance such as finding bounds on value-at-risk [11].

In Proposition 1.5, we study the range of values taken by $\tilde{S}_\eta$. In particular, we look for an upper bound when the copula is taken among the set of Gaussian copulae then among all the copulae. We also find the range of $\tilde{S}_\eta$. In order to maximize $\tilde{S}_\eta(C)$ over all the copulae, we use the approach of Frank et al. [13] with copulae.

Proposition 1.5. *Let $\eta > 0$.*

Let

$$C^r(u,v) = \begin{cases} M(u-1+r,v) & \text{if } (u,v) \in [1-r,1] \times [0,r], \\ W(u,v) & \text{if } (u,v) \in [0,1]^2 \setminus ([1-r,1] \times [0,r]) \end{cases}$$

with $r = 2\Phi\left(\frac{-\eta}{2}\right)$.

We have:

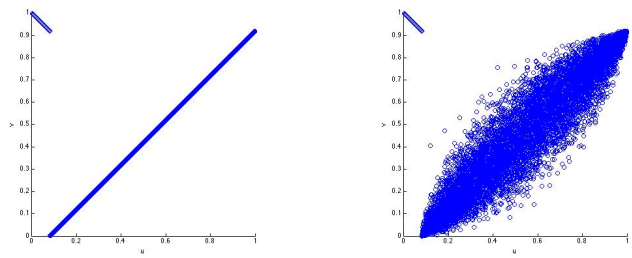
(i) Ran $\left(\tilde{S}_\eta|_{\mathcal{C}_G}\right) = \left[0, \Phi\left(\frac{-\eta}{2}\right)\right]$ with $\tilde{S}_\eta|_{\mathcal{C}_G}$ the restriction of $\tilde{S}_\eta$ to $\mathcal{C}_G$,

(ii) $\sup_{C \in \mathcal{C}} \tilde{S}_\eta(C) = 2\Phi\left(\frac{-\eta}{2}\right)$ and the supremum is achieved with C^r,

(iii) Ran $\left(\tilde{S}_\eta\right) = \left[0, 2\Phi\left(\frac{-\eta}{2}\right)\right]$.

If we only consider the set of Gaussian copulae, $\tilde{S}_\eta$ can only achieve the values in $\left[0, \Phi\left(\frac{-\eta}{2}\right)\right]$. If we consider all the copulae, values in $\left[\Phi\left(\frac{-\eta}{2}\right), 2\Phi\left(\frac{-\eta}{2}\right)\right]$ can also be achieved. Indeed, we can use the family of copulae constructed in Proposition 1.5 to achieve these values. It has a particular structure: it is divided in two parts according to the value of the first random variable. One state corresponds to a positive correlation and the upper bound is achieved in the comonotonic case. The other state corresponds to the countermonotonic case.

The family of copulae constructed in Proposition 1.5 are patchwork copulae [10]. Given a copula C, a patchwork copula is constructed by changing the value of C in a subrectangle of the unit square and replacing it with an other copula. In our case, we consider the countermonotonic copula and we change its values in the rectangle $[1-r,1] \times [0,r]$, replacing it by a Gaussian copula with parameter ρ. The copula achieving the bound corresponds to $\rho = 1$ and in this particular case, the copula is called a shuffle of M copula [20]. Figure 1.2 shows illustration of the copulae family constructed in Proposition 1.5 with a correlation of 1 and a correlation of -0.95.



(a) $\rho = 1$ (b) $\rho = 0.95$

Figure 1.2: *Patchwork copula $C^r(u,v)$ presenting two states depending on the value of u: the first copula is in the upper left part of the plan and is W ; the second one is the Gaussian copula with correlation equal to ρ, with $\rho = 1$ that is the degenerated copula M or $\rho = 0.95$. r is equal to $2\Phi\left(\frac{-\eta}{2}\right)$ with $\eta = 0.2$.*

If we consider two Brownian motions B^1 and B^2, B_t^1 and B_t^2 at a given time t are Gaussian random variables with variance t. Proposition 1.5 can be applied with $\eta^{'} = \frac{\eta}{\sqrt{t}}$. Modeling the dependence of Brownian motions with a Gaussian copula then limits us in terms of values taken by $S_{\eta,t}$. In particular, it is not possible to have probabilities greater than $\frac{1}{2}$ which was already proven with the symmetry property of Gaussian copulae.

In this section, we showed the limits of the Gaussian copulae and that it was possible to achieve new values for $\tilde{S}_\eta$ or to put asymmetry in the distribution of the difference with the use of different types of copulae. However, the copulae were used to model the dependence between the two Gaussian variables, i.e. two Brownian motions at given time t. We do not know if the copulae are suitable to model the dependence between $B^1 = (B_t^1)_{t\geq 0}$ and $B^2 = (B_t^2)_{t\geq 0}$, that is in a dynamical framework.

4.3 The Brownian Motion Case

Proposition 1.6 gives a time dynamical version of Proposition 1.5.

Proposition 1.6. *Let $\eta > 0$ and $t > 0$. We have:*

(i) $Ran\left(S_{\eta,t}|_{\mathcal{C}_G^d}\right) = \left[0, \Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$ *with $S_{\eta,t}|_{\mathcal{C}_G^d}$ the restriction of $S_{\eta,t}$ to $\mathcal{C}_G^d$,*

(ii) $\sup\limits_{C\in\mathcal{C}_B} S_{\eta,t}(C) = 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)$ *and the supremum is achieved with $C^{ref,\frac{\eta}{2}}$ which is the Reflection Brownian Copula defined by Equation (1.2),*

(iii) $Ran\left(S_{\eta,t}\right) = \left[0, 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$.

We have found a copula which maximizes $S_{\eta,t}$ at each time t which is admissible for Brownian motions. This copula is also a solution to the problem $\sup\limits_{C\in\mathcal{C}} \tilde{S}_{\frac{\eta}{\sqrt{t}}}(C)$ and gives an alternative solution of the supremum copula of Proposition 1.5. We also notice than $Ran\left(S_{\eta,t}\right) = Ran\left(\tilde{S}_{\frac{\eta}{\sqrt{t}}}\right)$. The constraint to be in $\mathcal{C}_B$ does not change the solution of our problem, values that can be achieved

are the same ; only copulae differ. Furthermore, copulae used to achieved the range of $S_{\eta,t}$ gives Markovian pairs of Brownian motions: the range of possible values for $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ is the same for Markovian pairs and all pairs of Brownian motions. As in the copula of Proposition 1.5, the copula that achieves the supremum has two states: one of comonotonicity and one of countermonotonicity, depending here on the value of the $B_t^1 - B_t^2$. Figure 1.3 represents the Reflection Brownian Copula at time $t = 1$ with a reflection at $\frac{\eta}{2} = 0.1$. We can see that the structure is the same than the copula of Figure 1.2a. However, in Figure 1.2a the two lines are in separated parts of the square and in Figure 1.3, there is a part of the square where they are both present. This is due to the fact that $\tilde{B}_t$ is not a deterministic function of B_t but a deterministic function of B_t and $\sup_{s \leq t} B_s$.

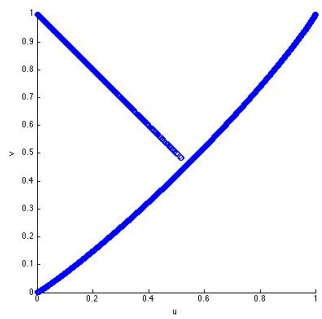


Figure 1.3: *Reflection Brownian Copula* $C_t^{ref,\frac{\eta}{2}}$ *at time* $t = 1$ *with* $\eta = 0.2$.

Part (iii) of Proposition 1.6 gives us a way to control $S_{\eta,t}$. Furthermore, when the copula is the Reflection Brownian one, the probability for $B_t^1 - B_t^2$ to have strictly higher value than η is equal to 0 and there is a discontinuity at η ; copulae of part (iii) allow us to solve this issue. The copulae become suboptimal but still achieves higher values than in the Gaussian copula case.

Result of Proposition 1.6 (ii) can be interpreted with coupling. Let X be a stochastic process. Let X^a and X^b be processes with the dynamic of X such that $X_0^a = a$ and $X_0^b = b$. A coupling is said successful if $T = \inf\{t \geq 0 : X_t^a = X_t^b\} < \infty$ almost surely. T is called the coupling time. In our situation, the two Brownian motions start at 0 and are coupled when $B_t = \tilde{B}_t^{\frac{\eta}{2}} + \eta$ which is equivalent to consider one Brownian starting at 0 and the other starting at η. We have the coupling inequality:

$$\|\mathbb{Q}_a(t) - \mathbb{Q}_b(t)\| \leq 2\mathbb{P}(T > t) \tag{1.5}$$

with $\|\|$ the total variation norm and $\mathbb{Q}_a(t)$ the distribution of X_t^a (same for $\mathbb{Q}_b(t)$ and X_b^t). In case of equality for (1.5), the coupling is said to be optimal [14]. The coupling by reflection [18], consisting of taking the reflection of the Brownian motion according to the hyperplane $x = \frac{a+b}{2}$, is optimal for Brownian motion. Hsu and Sturm [14] prove that in the case of Brownian motions, it is the only optimal Markovian coupling (definition 1.3).

Definition 1.3. *[14] Let $X = (X_1, X_2)$ be a coupling of Brownian motions. Let $\mathcal{F}^X$ be the filtration generated by X. We say that X is a Markovian coupling if for each $s \geq 0$, conditional on $\mathcal{F}_s^X$, the shifted process $\{(X_1(t+s), X_2(t+s)), t \geq 0\}$ is still a coupling of Brownian motions (now starting from $(X_1(s), X_2(s))$).*

In the optimal case, $\mathbb{P}(T > t)$ is minimal. The coupling by reflection can then be interpreted as the fastest way for the two processes to be equal. In our case, it is the fastest way for the $B^1 - B^2$ to be greater than η.

We found an admissible copula for Brownian motions which has the property to be asymmetric and to achieve upper bound for $S_{\eta,t}$. We have also constructed new families of asymmetric copulae allowing us to control the value of $S_{\eta,t}$.

5 Proofs

5.1 Preliminary results

We start with well known results that will be useful for the proofs of propositions.

Lemma 1.1. *Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. We have, for $y \geq 0$,*

$$\mathbb{P}\Big(B_t \leq x, \sup_{s\leq t} B_s \leq y\Big) = \begin{cases} \Phi\left(\frac{x}{\sqrt{t}}\right) - \Phi\left(\frac{x-2y}{\sqrt{t}}\right) & \text{if} \quad x < y \\ 2\Phi\left(\frac{y}{\sqrt{t}}\right) - 1 & \text{if} \quad x \geq y \end{cases}.$$

Proof The reader is referred to [16, Theorem 3.1.1.2, p. 137]. □

5.2 Proof of Proposition 1.1

We have:

$$\mathbb{P}\Big(B_t \leq x, \tilde{B}_t^h \leq y\Big) = \mathbb{P}\Big(B_t \leq x, \tilde{B}_t^h \leq y, \sup_{s\leq t} B_s \leq h\Big) + \mathbb{P}\Big(B_t \leq x, \tilde{B}_t^h \leq y, \sup_{s\leq t} B_s \geq h\Big). \quad (1.6)$$

We compute the first term of Equation (1.6):

$$\begin{aligned}
\mathbb{P}\Big(B_t \leq x, \tilde{B}_t^h \leq y, \sup_{s\leq t} B_s \leq h\Big) &= \mathbb{P}\Big(B_t \leq x, -B_t \leq y, \sup_{s\leq t} B_s \leq h\Big) \\
&= \mathbb{P}\Big(-y \leq B_t \leq x, \sup_{s\leq t} B_s \leq h\Big) \\
&= \Big(\mathbb{P}\Big(B_t \leq x, \sup_{s\leq t} B_s \leq h\Big) - \mathbb{P}\Big(B_t \leq -y, \sup_{s\leq t} B_s \leq h\Big)\Big)\mathbf{1}_{x+y>0} \\
&= \begin{cases} \Phi\left(\frac{x}{\sqrt{t}}\right) - \Phi\left(\frac{x-2h}{\sqrt{t}}\right) + \Phi\left(\frac{y}{\sqrt{t}}\right) - \Phi\left(\frac{y+2h}{\sqrt{t}}\right) & \text{if} \quad \begin{array}{c} x \leq h, \\ y \geq -h, \\ x+y>0 \end{array} \\ 2\Phi\left(\frac{h}{\sqrt{t}}\right) - 1 + \Phi\left(\frac{y}{\sqrt{t}}\right) - \Phi\left(\frac{y+2h}{\sqrt{t}}\right) & \text{if} \quad \begin{array}{c} x > h, \\ y \geq -h \end{array} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

by application of Lemma 1.1. In the same way, we compute the second term of Equation (1.6):

$$\begin{aligned}\mathbb{P}\Big(B_t \leq x, \tilde{B}_t^h \leq y, \sup_{s\leq t} B_s \geq h\Big) &= \mathbb{P}\Big(B_t \leq x, B_t \leq y+2h, \sup_{s\leq t} B_s \geq h\Big)\\ &= \mathbb{P}\Big(B_t \leq \min\left(x, y+2h\right), \sup_{s\leq t} B_s \geq h\Big)\\ &= \begin{cases} \Phi\Big(\frac{\min(x,y+2h)-2h}{\sqrt{t}}\Big) & \text{if} \quad \min\left(x,y+2h\right) < h \\ -2\Phi\Big(\frac{h}{\sqrt{t}}\Big) + 1 + \Phi\Big(\frac{\min(x,y+2h)}{\sqrt{t}}\Big) & \text{if} \quad \min\left(x,y+2h\right) \geq h \end{cases}.\end{aligned}$$

Combining the last two equations, we obtain

$$\begin{aligned}\mathbb{P}\left(B_t \leq x, \tilde{B}_t^h \leq y\right) &= \begin{cases} \Phi\left(\frac{\min(x,y+2h)-2h}{\sqrt{t}}\right) \text{ if } x+y \leq 0 \text{ or } (y \leq -h, x+y>0) \\ \Phi\left(\frac{\min(x,y+2h)}{\sqrt{t}}\right) - \Phi\left(\frac{y+2h}{\sqrt{t}}\right) + \Phi\left(\frac{y}{\sqrt{t}}\right) \text{ if } y > -h,\ x+y>0 \end{cases}\\ &= \begin{cases} \Phi\Big(\frac{y}{\sqrt{t}}\Big) \text{ if } x-y \geq 2h \\ \Phi\Big(\frac{x-2h}{\sqrt{t}}\Big) \text{ if } x-y < 2h, x+y \leq 0 \\ \Phi\Big(\frac{x}{\sqrt{t}}\Big) - \Phi\Big(\frac{y+2h}{\sqrt{t}}\Big) + \Phi\Big(\frac{y}{\sqrt{t}}\Big) \text{ if } x-y < 2h, x+y > 0 \end{cases} \qquad (1.7)\\ &= \begin{cases} \Phi\Big(\frac{y}{\sqrt{t}}\Big) \text{ if } x-y \geq 2h \\ W\Big(\Phi\left(\frac{y}{\sqrt{t}}\right) + \Phi\Big(\frac{x}{\sqrt{t}}\Big)\Big) + M\Big(\Phi\Big(\frac{x-2h}{\sqrt{t}}\Big), \Phi\Big(\frac{-y-2h}{\sqrt{t}}\Big)\Big) \text{ if } x-y < 2h \end{cases}.\end{aligned}$$

We conclude using $C_t^{ref,h}(u,v) = \mathbb{P}\Big(B_t \leq \sqrt{t}\Phi^{-1}(u), \tilde{B}_t^h \leq \sqrt{t}\Phi^{-1}(v)\Big)$.

5.3 Proof of Proposition 1.2

Recall that Φ_ρ denotes the bivariate cumulative distribution function of two standard normal variables correlated with $\rho \in [-1,1]$. We start with a technical lemma.

Lemma 1.2. *Let a, b and $x \in \mathbb{R}$. We have:*

(i)

$$\int_{-\infty}^{x} \Phi\left(au+b\right) \frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}} du = \Phi_{\frac{-a}{\sqrt{a^2+1}}}\Big(\frac{b}{\sqrt{a^2+1}}, x\Big).$$

(ii)

$$\Phi_{\sqrt{1-\rho^2}}(x,y) = \Phi(y)\,\Phi\Big(\frac{x-\sqrt{1-\rho^2}y}{\rho}\Big) + \Phi(x) - \Phi_\rho\Big(x, \frac{x-\sqrt{1-\rho^2}y}{\rho}\Big),\ x,y \in \mathbb{R},\ \rho > 0$$

(iii)

$$\Phi_\rho(x,y) = \Phi(y) - \Phi_{-\rho}(-x,y),\ x,y \in \mathbb{R}$$

Proof

(ii) Let $a < 0$, $b, z \in \mathbb{R}$. We have:

$$\begin{aligned}
\Phi_{\frac{-a}{\sqrt{a^2+1}}}\Big(\frac{b}{\sqrt{a^2+1}}, z\Big) &= \int_{-\infty}^{z} \Phi\left(au+b\right) \frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}} du \\
&= \Phi\left(az+b\right)\Phi\left(z\right) - a\int_{-\infty}^{z} \frac{e^{\frac{-(au+b)^2}{2}}}{\sqrt{2\pi}} \Phi\left(u\right) du \\
&= \Phi\left(az+b\right)\Phi\left(z\right) + \int_{az+b}^{+\infty} \frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}} \Phi\Big(\frac{u-b}{a}\Big) du \\
&= \Phi\left(az+b\right)\Phi\left(z\right) + \Phi\Big(\frac{b}{\sqrt{1+a^2}}\Big) - \Phi_{\frac{1}{\sqrt{a^2+1}}}\Big(\frac{b}{\sqrt{a^2+1}}, ax+b\Big).
\end{aligned}$$

We conclude by taking $a = \frac{-\rho}{\sqrt{1-\rho^2}}$, $b = x\sqrt{1+a^2}$, $z = \frac{y-b}{a}$.

(iii) Let (X, Y) be a Gaussian vector with X and Y having correlation ρ. We have:

$$\mathbb{P}\left(X \le x, Y \le y\right) = \mathbb{P}\left(-X \ge -x, Y \le y\right) = \mathbb{P}\left(Y \le y\right) - \mathbb{P}\left(-X \le -x, Y \le y\right).$$

□

We can now prove Proposition 1.2. Let $X = B$ and $Y = \rho\tilde{B}^h + \sqrt{1-\rho^2}Z$ where B and Z are two independent Brownian motions. X and Y are Brownian motions and we have

$$\begin{aligned}
\mathbb{P}\left(X_t \le x, Y_t \le y\right) &= \mathbb{P}\Big(B_t \le x, \tilde{B}_t^h \le \frac{y-\sqrt{1-\rho^2}Z_t}{\rho}\Big) \\
&= \mathbb{E}\Big[\mathbb{P}\Big(B_t \le x, \tilde{B}_t^h \le \frac{y-\sqrt{1-\rho^2}Z_t}{\rho} \mid Z_t\Big)\Big].
\end{aligned}$$

Since B is independent from Z, using Equation (1.7), we find that $\mathbb{P}\left(X_t \le x, Y_t \le y\right)$ is the sum of the three following terms:

(i)

$$\mathbb{E}\Big[\Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\mathbf{1}_{Z_t \ge \frac{\rho(2h-x)+y}{\sqrt{1-\rho^2}}}\Big],$$

(ii)

$$\mathbb{E}\Big[\Phi\Big(\frac{x-2h}{\sqrt{t}}\Big)\mathbf{1}_{Z_t \le \frac{\rho(2h-x)+y}{\sqrt{1-\rho^2}}}\mathbf{1}_{Z_t \ge \frac{\rho x+y}{\sqrt{1-\rho^2}}}\Big],$$

(iii)

$$\mathbb{E}\Big[\Big(\Phi\Big(\frac{x}{\sqrt{t}}\Big) - \Phi\Big(\frac{y+2h\rho-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big) + \Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\Big)\mathbf{1}_{Z_t \le \frac{\rho(2h-x)+y}{\sqrt{1-\rho^2}}}\mathbf{1}_{Z_t \le \frac{\rho x+y}{\sqrt{1-\rho^2}}}\Big].$$

The first term (i) is equal to:

$$\mathbb{E}\Big[\Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\Big] - \mathbb{E}\Big[\Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\mathbf{1}_{Z_t \le \frac{\rho(2h-x)+y}{\sqrt{1-\rho^2}}}\Big].$$

Furthermore, using Lemma 1.2, we have

$$\mathbb{E}\Big[\Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\mathbf{1}_{Z_t\leq\frac{\rho(2h-x)+y}{\sqrt{1-\rho^2}}}\Big]=\int_{-\infty}^{\frac{\rho(2h-x)+y}{\sqrt{(1-\rho^2)t}}}\Phi\Big(\frac{y-\sqrt{1-\rho^2}\sqrt{t}u}{\rho\sqrt{t}}\Big)\frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}}du$$
$$=\Phi_{\sqrt{1-\rho^2}}\Big(\frac{y}{\sqrt{t}},\frac{\rho(2h-x)+y}{\sqrt{(1-\rho^2)\,t}}\Big)$$

and

$$\mathbb{E}\Big[\Phi\Big(\frac{y-\sqrt{1-\rho^2}Z_t}{\rho\sqrt{t}}\Big)\Big]=\Phi\Big(\frac{y}{\sqrt{t}}\Big).$$

We compute terms (ii) and (iii) using the same method and we find that $\mathbb{P}\Big(X_t\leq x,Y_t\leq y\Big)$ is equal to

$$\Phi\Big(\frac{y}{\sqrt{t}}\Big)-\Phi_{\sqrt{1-\rho^2}}\Big(\frac{y}{\sqrt{t}},\frac{\rho\,(2h-x)+y}{\sqrt{(1-\rho^2)\,t}}\Big)+\Phi_{\sqrt{1-\rho^2}}\Big(\frac{y}{\sqrt{t}},\frac{\rho\min(2h-x,x)+y}{\sqrt{(1-\rho^2)\,t}}\Big)$$
$$-\Phi_{\sqrt{1-\rho^2}}\Big(\frac{y+2h\rho}{\sqrt{t}},\frac{\rho\min(2h-x,x)+y}{\sqrt{(1-\rho^2)\,t}}\Big)+\Phi\Big(\frac{x}{\sqrt{t}}\Big)\Phi\Big(\frac{\rho\min(2h-x,x)+y}{\sqrt{(1-\rho^2)\,t}}\Big)$$
$$+\Phi\Big(\frac{x-2h}{\sqrt{t}}\Big)\Big(\Phi\Big(\frac{\rho\,(\max(2h-x,x))+y}{\sqrt{(1-\rho^2)\,t}}\Big)-\Phi\Big(\frac{\rho x+y}{\sqrt{(1-\rho^2)\,t}}\Big)\Big).$$

After some algebra, we find using Lemma 1.2:

$$\mathbb{P}\Big(X_t\leq x,Y_t\leq y\Big)=\begin{cases}\Phi_\rho\Big(\frac{y+2\rho h}{\sqrt{t}},\frac{x}{\sqrt{t}}\Big)+\Phi\Big(\frac{y}{\sqrt{t}}\Big)-\Phi\Big(\frac{y+2\rho h}{\sqrt{t}}\Big) & \text{if } x\geq h\\ \Phi_{-\rho}\Big(\frac{y}{\sqrt{t}},\frac{x}{\sqrt{t}}\Big)+\Phi_\rho\Big(\frac{-y-2\rho h}{\sqrt{t}},\frac{x-2h}{\sqrt{t}}\Big)+\Phi_\rho\Big(\frac{y}{\sqrt{t}},\frac{x-2h}{\sqrt{t}}\Big)-\Phi\Big(\frac{x-2h}{\sqrt{t}}\Big) & \text{if } x<h\end{cases}$$

and the copula is equal to $\mathbb{P}\left(X_t\leq\sqrt{t}\Phi^{-1}(u),Y_t\leq\sqrt{t}\Phi^{-1}(v)\right)$.

5.4 Proof of Proposition 1.3

Let f^ξ be the density of ξ. Let B be a Brownian motion independent from ξ. We enlarge the filtration of B to take into account ξ. We consider the reflection of the Brownian motion $\tilde{B}^\xi$. We have:

$$\mathbb{P}\left(B_t\leq\sqrt{t}\Phi^{-1}(u),\tilde{B}_t^\xi\leq\sqrt{t}\Phi^{-1}(v)\right)=\mathbb{E}\left[\mathbb{P}\left(B_t\leq\sqrt{t}\Phi^{-1}(u),\tilde{B}_t^\xi\leq\sqrt{t}\Phi^{-1}(v)\mid\xi\right)\right].\quad(1.8)$$

Since B is independent from ξ, we have according to Proposition 1.1:

$$\mathbb{P}\left(B_t\leq\sqrt{t}\Phi^{-1}(u),\tilde{B}_t^\xi\leq\sqrt{t}\Phi^{-1}(v)\mid\xi\right)=v\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)\geq\frac{2\xi}{\sqrt{t}}}+W(u,v)\,\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)<\frac{2\xi}{\sqrt{t}}}$$
$$+\Phi\Big(\Phi^{-1}(M(u,1-v))-\frac{2\xi}{\sqrt{t}}\Big)\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)<\frac{2\xi}{\sqrt{t}}}$$

Thus, the right hand side of Equation (1.8) is the sum of the three following terms:

(i)

$$\mathbb{E}\Big[v\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)\geq\frac{2\xi}{\sqrt{t}}}\Big]=vF^\xi\Big(\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}\Big),$$

(ii)

$$\mathbb{E}\Big[W(u,v)\,\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)<\frac{2\xi}{\sqrt{t}}}\Big] = W(u,v)\,\overline{F}^{\xi}\Big(\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}\Big),$$

(iii)

$$\mathbb{E}\Big[\Phi\Big(\Phi^{-1}(M(u,1-v)) - \frac{2\xi}{\sqrt{t}}\Big)\mathbf{1}_{\Phi^{-1}(u)-\Phi^{-1}(v)<\frac{2\xi}{\sqrt{t}}}\Big],$$

that we denote by I.

We have:

$$\begin{aligned} I &= \int_{\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}}^{+\infty} \Phi\Big(\Phi^{-1}(M(u,1-v)) - \frac{2h}{\sqrt{t}}\Big) f^{\xi}(h)\,dh \\ &= M(1-u,v)\,\overline{F}^{\xi}\Big(\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}\Big) \\ &- \frac{2}{\sqrt{t}}\int_{\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}}^{+\infty} \Phi'\Big(\Phi^{-1}(M(u,1-v)) - \frac{2h}{\sqrt{t}}\Big)\overline{F}^{\xi}(h)\,dh. \end{aligned}$$

Adding the three terms of Equation (1.8), since $M(1-u,v)+W(u,v)=v$, we obtain:

$$\begin{aligned} C_t^{\xi}(u,v) &= v - \frac{2}{\sqrt{t}}\int_{\sqrt{t}\frac{\Phi^{-1}(u)-\Phi^{-1}(v)}{2}}^{+\infty} \Phi'\Big(\Phi^{-1}(M(u,1-v)) - \frac{2h}{\sqrt{t}}\Big)\overline{F}^{\xi}(h)\,dh \\ &= v - \int_{-\infty}^{\Phi^{-1}(M(1-u,v))} \Phi'(h)\,\overline{F}^{\xi}\Big(\frac{\sqrt{t}}{2}\big(\Phi^{-1}(M(u,1-v)) - h\big)\Big)dh \end{aligned}$$

with $\Phi'(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$, which achieves the proof.

5.5 Proof of Proposition 1.4

[8, Proposition 2.1] states that

$$\mathbb{P}(X+Y\le x) = \int_0^1 \partial_u C\Big(u, F^Y\Big(x - \big(F^X\big)^{-1}(u)\Big)\Big)\,du, x\in\mathbb{R}. \tag{1.9}$$

The existence of $\partial_u C\Big(u, F^Y\Big(x - \big(F^X\big)^{-1}(u)\Big)\Big)$ for $u\in[0,1]$ is assured by [8, Lemma 2.1].
We also have

$$F^{-Y}(y) = 1 - F^Y(-y),\ y\in\mathbb{R} \text{ and,} \tag{1.10}$$

$$C^{X,-Y}(u,v) = u - C(u,1-v),\ (u,v)\in[0,1], \tag{1.11}$$

with $C^{X,-Y}$ the copula of $(X,-Y)$.
Equation (1.9) is also valid for $(X,-Y)$. Using Equation (1.10) and Equation (1.11), we have

$$\mathbb{P}(X-Y\le x) = \int_0^1 \Big[1 - \partial_u C\Big(u, F^Y\Big(\big(F^X\big)^{-1}(u) - x\Big)\Big)\Big]\,du, x\in\mathbb{R} \tag{1.12}$$

and

$$\mathbb{P}(X-Y> x) = \int_0^1 \partial_u C\Big(u, F^Y\Big(\big(F^X\big)^{-1}(u) - x\Big)\Big)\,du, x\in\mathbb{R}. \tag{1.13}$$

Let us suppose that $C^{Y,X} \in \mathcal{C}_s$ and that X and Y have the same continous marginal distribution function F. Let $C^{Y,X}$ be the copula of (Y, X). We have $C^{Y,X}(u,v) = C^{X,Y}(v,u)$. However, $C^{X,Y}(v,u) = C^{X,Y}(u,v)$ then $C^{Y,X}(u,v) = C^{X,Y}(u,v)$ and

$$\begin{aligned}
\mathbb{P}(X - Y \geq x) &= \mathbb{P}(Y - X \leq -x) \\
&= \int_0^1 \left[1 - \partial_u C^{Y,X}\left(u, F\left((F)^{-1}(u) + x\right)\right)\right] du \\
&= \int_0^1 \left[1 - \partial_u C^{X,Y}\left(u, F\left((F)^{-1}(u) + x\right)\right)\right] du \\
&= \mathbb{P}(X - Y \leq -x)
\end{aligned}$$

using Equation (1.12).

5.6 Proof of Proposition 1.5

(i) Let $\rho \in [-1, 1]$. We have $\tilde{S}_\eta(C_{G,\rho}) = \Phi\left(\frac{-\eta}{\sqrt{2(1-\rho)}}\right)$. This function is decreasing in ρ and then the extremum are achieved for $\rho = 1$ and $\rho = -1$ and are equal to 0 and $\Phi\left(\frac{-\eta}{2}\right)$.
(ii) This is a direct application of the results of [13] where superior and inferior bounds on $\mathbb{P}(X + Y < \eta)$ are found and where X and Y are two random variables with known marginals. As

1) $\sup_{C \in \mathcal{C}} \mathbb{P}_C(X - Y \geq \eta) = 1 - \inf_{C \in \mathcal{C}} \mathbb{P}_C(X - Y < \eta)$,

2) $-Y$ and Y have the same law,

the bound is equal to

$$1 - \inf_{C \in \mathcal{C}} \mathbb{P}_C(X + Y < \eta).$$

The copula achieving the bound is defined by the transformation

$$C^{X,Y}(u,v) = u - C^{X,-Y}(u, 1 - v).$$

(iii) We want to prove that for all x in $\left[0, 2\Phi\left(\frac{-\eta}{2}\right)\right]$, there exists C in $\mathcal{C}$ such that $\tilde{S}_\eta(C) = x$.

If $x \in \left[0, \Phi\left(\frac{-\eta}{2}\right)\right]$, we use a Gaussian copula with $\rho = 1 - \frac{1}{2^2}\left(\frac{\eta}{\Phi^{-1}(x)}\right)^2$.

Let us suppose that $x \in \left[\Phi\left(\frac{-\eta}{2}\right), 2\Phi\left(\frac{-\eta}{2}\right)\right]$. We use the copula C^r to construct a new class of copulae. As for C^r, we separate the square $[0,1]^2$ in two parts and to define a copula in each part of the square. We use the concept of patchwork copula defined by Durante et al. [10]. Let $H = [1 - r, 1] \times [0, r]$, $H^c = [0,1]^2 \setminus H$ and $\rho \in (-1, 1)$. Let $C_\rho^p(u,v)$ the patchwork copula defined by C_ρ in H and W in H^c:

$$\begin{aligned}
C_\rho^p(u,v) &= \mu_W\left(([0,u] \times [0,v]) \cap H^c\right) + rC_{G,\rho}\left(\frac{1}{r}\max(u + r - 1, 0), \frac{1}{r}\min(v, r)\right) \\
&= \left(W(u,v) - W(u,r)\right)\mathbf{1}_{v \geq r} + rC_{G,\rho}\left(\frac{1}{r}\max(u + r - 1, 0), \min\left(\frac{v}{r}, 1\right)\right)
\end{aligned}$$

where μ_W is the measure induced by the copula W.

If we consider two standard normal random variables with copula C_ρ^p, the survival function of their difference at point x is equal, according to Equation (1.13), to

$$\begin{aligned}\int_0^1 \partial_u C_\rho^p\left(u, \Phi\left(\Phi^{-1}(u)-x\right)\right) du &= \int_0^1 \left(\mathbf{1}_{u\geq\Phi\left(\frac{x}{2}\right)} - \mathbf{1}_{u\geq 1-r}\right)\mathbf{1}_{u\geq\Phi(\Phi^{-1}(r)+x)} du \\ &+ \int_{1-r}^1 \Phi\Big(\frac{\Phi^{-1}\big(\min\big(\frac{\Phi(\Phi^{-1}(u)-x)}{r},1\big)\big) - \rho\Phi^{-1}\left(\frac{u+r-1}{r}\right)}{\sqrt{1-\rho^2}}\Big) du \\ &= \left(1-r-\Phi\left(\frac{x}{2}\right)\right)\mathbf{1}_{x\leq 2\Phi^{-1}(1-r)} \\ &+ \int_{1-r}^1 \Phi\Big(\frac{\Phi^{-1}\big(\min\big(\frac{\Phi(\Phi^{-1}(u)-x)}{r},1\big)\big) - \rho\Phi^{-1}\left(\frac{u+r-1}{r}\right)}{\sqrt{1-\rho^2}}\Big) du\end{aligned}$$

which is continuous at $x=\eta$. Thus, $\tilde{S}_\eta\left(C_\rho^p\right)$ is equal to the survival function of their difference at point η, which is:

$$\tilde{S}_\eta\left(C_\rho^p\right) = \int_{1-r}^1 \Phi\Big(\frac{\Phi^{-1}\big(\min\big(\frac{\Phi(\Phi^{-1}(u)-\eta)}{r},1\big)\big) - \rho\Phi^{-1}\left(\frac{u+r-1}{r}\right)}{\sqrt{1-\rho^2}}\Big) du.$$

Using the previous equation and dominated convergence theorem, we can prove that $\rho\mapsto\tilde{S}_\eta\left(C_\rho^p\right)$ is continuous on $(-1,1)$.
We have $C_1^p = C^r$ and $C_{-1}^p = W$. Furthermore, we can show after some algebra that

$$\tilde{S}_\eta\left(C_\rho^p\right) \underset{\rho\to 1}{\longrightarrow} 2\Phi\Big(\frac{-\eta}{2}\Big) = \tilde{S}_\eta\left(C_1^p\right)$$

and

$$\tilde{S}_\eta\left(C_\rho^p\right) \underset{\rho\to -1}{\longrightarrow} \Phi\Big(\frac{-\eta}{2}\Big) = \tilde{S}_\eta\left(C_{-1}^p\right).$$

Then $\rho\mapsto\tilde{S}_\eta\left(C_\rho^p\right)$ is continuous on $[-1,1]$, which achieves the proof.

5.7 Proof of Proposition 1.6

(i) As the copulae of $\mathcal{C}_G^d$ are of the form $\left(C_{G,\rho_t}\right)_{t\geq 0}$, the demonstration of this part of the proposition is similar to the one of the static framework.
(ii) Let (B^1, B^2) be two Brownian motion with copula $C^{ref,\frac{\eta}{2}}$. B^2 is then the reflection of B^1 according to the stopping time $\tau = \inf\{t\geq 0 : B_t^1 = \frac{\eta}{2}\} = \inf\{t\geq 0 : B_t^1 - B_t^2 = \eta\}$. For $t<\tau$, $B_t^1 - B_t^2 < \eta$ and for $t\geq\tau$, $B_t^1 - B_t^2 = \eta$. Thus, we have:

$$S_{\eta,t}\left(C^{ref,\frac{\eta}{2}}\right) = \mathbb{P}_{C^{ref,\frac{\eta}{2}}}\left(t\geq\tau\right) = \mathbb{P}_{C^{ref,\frac{\eta}{2}}}\Big(\sup_{s\leq t} B_s^1 \geq \frac{\eta}{2}\Big) = 2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big)$$

according to Lemma 1.1.
If $C\in\mathcal{C}_B$, the copula C_t is in $\mathcal{C}$ and then according to Proposition 1.5

$$\sup_{C\in\mathcal{C}_B} \mathbb{P}_C\left(B_t^1 - B_t^2 \geq \eta\right) \leq \sup_{C\in\mathcal{C}} \mathbb{P}_C\left(B_t^1 - B_t^2 \geq \eta\right) = 2\Phi\Big(\frac{-\eta}{2\sqrt{t}}\Big),$$

which concludes this part of the proof.
(iii) We want to prove that for all x in $\left[0, 2\Phi\left(\frac{-\eta}{2}\right)\right]$, there exists C in $\mathcal{C}$ such that $\tilde{S}_\eta(C) = x$. Let $x \in \left[0, 2\Phi\left(\frac{-\eta}{2\sqrt{T}}\right)\right]$.

If $x \in \left[0, \Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$, we consider the Gaussian dynamic copula with $(\rho_s)_{s\geq 0} = 1 - \frac{1}{2t}\left(\frac{\eta}{\Phi^{-1}(x)}\right)^2$ which is in $[-1, 1]$ and we have $S_{\eta,t}(C_{G,\rho}) = x$.

If $x \in \left[\Phi\left(\frac{-\eta}{2\sqrt{t}}\right), 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$, we consider the copula $C^{ref,\frac{\eta}{2}+\lambda}$ defined by Equation (1.2) with $\lambda \geq 0$. With the use of Lemma 1.1, we find that

$$\mathbb{P}_{C^{ref,\frac{\eta}{2}+\lambda}}\left(B_t^1 - B_t^2 \geq x\right) = \begin{cases} \Phi\left(\frac{-x}{2\sqrt{t}}\right) + \Phi\left(\frac{x-2\eta-4\lambda}{2\sqrt{t}}\right) & \text{if} \quad x \leq \eta + 2\lambda \\ 0 & \text{if} \quad x > \eta + 2\lambda \end{cases}$$

Then,

$$S_{\eta,t}\left(C^{ref,\frac{\eta}{2}+\lambda}\right) = \Phi\left(\frac{-\eta}{2\sqrt{t}}\right) + \Phi\left(\frac{-\eta-4\lambda}{2\sqrt{t}}\right)$$

As we have:

1) $\lambda \mapsto S_{\eta,t}\left(C^{ref,\frac{\eta}{2}+\lambda}\right)$ is continuous on $[0, \infty)$,

2) $S_{\eta,t}\left(C^{ref,\frac{\eta}{2}+\lambda}\right) \underset{\lambda\to 0}{\longrightarrow} 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)$,

3) $S_{\eta,t}\left(C^{ref,\frac{\eta}{2}+\lambda}\right) \underset{\lambda\to\infty}{\longrightarrow} \Phi\left(\frac{-\eta}{2\sqrt{t}}\right)$,

we can conclude.

Bibliography

[1] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

[2] Tomasz R Bielecki, Jacek Jakubowski, Andrea Vidozzi, and Luca Vidozzi. Study of dependence for some stochastic processes. *Stochastic analysis and applications*, 26(4):903–924, 2008.

[3] Damien Bosc. *Three essays on modeling the dependence between financial assets*. PhD thesis, Ecole Polytechnique X, 2012.

[4] René Carmona and Valdo Durrleman. Pricing and hedging spread options. *Siam Review*, 45(4):627–685, 2003.

[5] René Carmona and Valdo Durrleman. Generalizing the black-scholes formula to multivariate contingent claims. *Journal of computational finance*, 9(2):43, 2005.

[6] Mu-Fa Chen and Shao-Fu Li. Coupling methods for multidimensional diffusion processes. *The Annals of Probability*, pages 151–177, 1989.

[7] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004.

[8] Umberto Cherubini, Sabrina Mulinacci, and Silvia Romagnoli. On the distribution of the (un) bounded sum of random variables. *Insurance: Mathematics and Economics*, 48(1):56–63, 2011.

[9] William F Darsow, Bao Nguyen, and Elwood T Olsen. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

[10] Fabrizio Durante, Juan Fernández Sánchez, and Carlo Sempi. Multivariate patchwork copulas: a unified approach with applications to partial comonotonicity. *Insurance: Mathematics and Economics*, 53(3):897–905, 2013.

[11] Paul Embrechts, Andrea Höing, and Alessandro Juri. Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, 7(2):145–167, 2003.

[12] Jean-David Fermanian and Marten Wegkamp. Time dependent copulas. *Preprint INSEE, Paris, France*, 2004.

[13] Maurice J Frank, Roger B Nelsen, and Berthold Schweizer. Best-possible bounds for the distribution of a sum - a problem of kolmogorov. *Probability Theory and Related Fields*, 74(2):199–211, 1987.

[14] Elton P Hsu and Karl-Theodor Sturm. Maximal coupling of euclidean brownian motions. *Communications in Mathematics and Statistics*, 1(1):93–104, 2013.

[15] Piotr Jaworski and Marcin Krzywda. Coupling of wiener processes by using copulas. *Statistics & Probability Letters*, 83(9):2027–2033, 2013.

[16] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical methods for financial markets*. Springer, 2009.

[17] David X Li. On default correlation: A copula function approach. *Available at SSRN 187289*, 1999.

[18] Torgny Lindvall and L C G Rogers. Coupling of multidimensional diffusions by reflection. *The Annals of Probability*, pages 860–872, 1986.

[19] GD Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4):803–806, 1982.

[20] Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2006.

[21] Andrew J Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556, 2006.

[22] Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, pages 623–632, 1982.

[23] Carlo Sempi. Coupled brownian motion. In *Combining Soft Computing and Statistical Methods in Data Analysis*, pages 569–574. Springer, 2010.

[24] M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

Chapter 2

On the control of the difference between two Brownian motions: an application to energy markets modeling

Abstract

We derive a model based on the structure of dependence between a Brownian motion and its reflection according to a barrier. The structure of dependence presents two states of correlation: one of comonotonicity with a positive correlation and one of countermonotonicity with a negative correlation. This model of dependence between two Brownian motions B^1 and B^2 allows for the value of $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ to be higher than $\frac{1}{2}$ when x is close to 0, which is not the case when the dependence is modeled by a constant correlation. It can be used for risk management and option pricing in commodity energy markets. In particular, it allows to capture the asymmetry in the distribution of the difference between electricity prices and its combustible prices.

1 Introduction

1.1 Motivation

One of the major issues in commodity energy markets is the pricing and hedging of multi-assets options, in particular the spread options. For instance, if we denote by X_t the price of electricity at time t, by Y_t the price of coal at time t, and by H the heat rate (conversion factor) between the two, the income of the coal plant can be modeled by $(X_t - HY_t - K)^+$, with $x^+ = \max(x, 0)$ and K representing a fixed cost. To evaluate the value of this coal plant, one needs to model jointly the price of electricity and the price of coal. Because the coal is a fuel for electricity, the two prices can not be considered independent and dependence between the two needs to be modeled. For more information on spread options, the reader can refer to [6].

To model energy commodities forward prices, two different approaches exist: the first one consists in the modeling of the spot price and is equivalent to the Vasicek modeling of interest rates [20], the second one consists in the modeling of the forward curve and is similar to the a Heath-Jarrow-Morton approach [12]. In the first approach, one way to model dependence is the use of structural models [1; 2; 5]. In structural models, electricity is a function of the residual demand and of the fuels used to produce it. Some constraints are imposed in order for the electricity price to be higher than the minimum price of its combustibles with a high probability. An other way to model dependence is the use of co-integration between the different commodities spot prices [17]. However, structural models and co-integration models are very computational costly and are not adapted for practitioners. They prefer to use the second type of models, using a forward curve. We denote by $f^i(t, T)$ the forward price of commodity i at time t with maturity T, that is of the delivery of commodity i during one unit of time. The most common model for $f^i(t, T)$ is the two-factor model, see [3] for instance. The forward price of commodity i, $i = 1, .., n$ is modeled by the following stochastic differential equation:

$$df^i(t, T) = f^i(t, T)\left(\sigma_s^i e^{-\alpha^i(T-t)} dB_t^{s,i} + \sigma_l^i dB^{l,i}\right) \tag{2.1}$$

with $B^{s,i}$ and $B^{l,i}$, $i = 1, .., n$, $2n$ brownian motions. The dependence between the Brownian motions is usually modeled by a constant correlation matrix. In the following, we are interested only in factorial models with two commodities, electricity and one of its fuel. Marginals model (if we consider only one commodity) are really efficient and allow us to price efficiently options based on one underlying. However, dependence modeling is not satisfying because it does not capture the asymmetry in the distribution of the difference between the forward price of electricity and the one of its fuel. Furthermore, the probability for the price of electricity to be lower than the price of its fuel is closed to $\frac{1}{2}$ which is not consistent with the reality. Indeed, the fuel is used to produce the electricity.

Modeling the dependence between the forward prices is equivalent to the modeling of the dependence between the Brownian motions. We consider only two Brownian motions. To capture asymmetry, it is needed to consider an other approach than the constant correlation model. A common approach to construct a pair of Brownian motions is the use of stochastic correlation. Stochastic correlation models are a generalization in a multivariate framework of stochastic volatility models, such as the Heston model [13] where the volatility is modeled by a Cox-Ingersoll-Ross process. The matrix of volatility-correlation is stochastic and can be modeled for instance by a Wishart processes [11]. In a stochastic correlation framework, as the difference between the two Brownian motions does not follow a normal law, it is possible to capture asymmetry. However, if the

stochastic correlation $(\rho_s)_{s\geq 0}$ is independent from the two Brownian motions, we have for $x \geq 0$:

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \mathbb{E}\Big(\Phi\Big(\frac{-x}{\sqrt{2\int_0^t (1-\rho_s)\,ds}}\Big)\Big) \leq \frac{1}{2}$$

with Φ the normal cumulative distribution function. Stochastic correlation does not allow to have higher value then $\frac{1}{2}$ for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$. An other way to construct a pair of Brownian motions is the use of a local correlation. The concept of local correlation is directly derived from the one of local volatility. In a Black and Scholes framework, the volatility is constant with the maturity and strikes which is not coherent with the implied volatilities from call and put option prices. Dupire introduces the local volatility in order to have a price model which is compatible with the volatility smiles and which is a complete market model [9]. Langnau introduces local correlation model which is the generalization of local volatility for a multi-dimensional framework [16]. A less common approach is the use of copulae. Copulae are used to model the dependence between random variables and have many applications in finance [7]. Indeed, Sklar's theorem [19] states that modeling the distribution of a couple of random variables (X, Y) is equivalent than modeling the law of X, the law of Y and a copula function C corresponding to the dependence between the two. However, use of copulae is more complicated in a continuous time framework, that is when processes are involved. In [4] and [14], a partial derivative equation is derived linking the copula between two Brownian motions and their local correlation function based on the Kolmogorov forward equation. Constraints on the copula to be admissible for Brownian motions are very restrictive, especially if one want to find asymmetric copulae admissible for Brownian motions. Deschatre [8] derives families of copula that are admissible for Brownian motions and asymmetric. Furthermore, he studies the range of the function $C \mapsto \mathbb{P}_C\left(B_t^1 - B_t^2 \geq \eta\right)$ for $\eta > 0$ and $t > 0$ where B^1 and B^2 are two Brownian motions and $\mathbb{P}_C$ denotes the measure of probability when C is the copula of $\left(B^1, B^2\right)$. Some Markovian constraints are imposed on the copula C. The range of this function is equal to $\left[0, 2\Phi\left(\frac{-\eta}{2\sqrt{t}}\right)\right]$ and the supremum is achieved with the copula between the Brownian motion and its reflection according to the barrier $\frac{\eta}{2}$. However, those results are not adapted to a modeling framework because of the degenerescence of the model: the Brownian motions are either correlated to 1 or to -1 depending on the value of B^1.

1.2 Objectives and results

The main objective of this paper is to construct a model of dependence for solutions $f^1(t,T)$ and $f^2(t,T)$ of the stochastic differential equations (2.1). This model of dependence must create asymmetry in the difference between the two processes. In particular, we want to have high value for $\mathbb{P}\left(f^1(t,T) - f^2(t,T) \geq x\right)$ with x close to 0. The dependence between the two processes is determined by the dependence between the Brownian motions. We reduce our problem to the construction of two Brownian motions $B^1 = \left(B_t^1\right)_{t\geq 0}$ and $B^2 = \left(B_t^2\right)_{t\geq 0}$ presenting asymmetry in their dependence and with values for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ higher than $\frac{1}{2}$ when x is close to 0. Our model is based on the work of Deschatre [8]. The value of $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ for $\eta > 0$ is maximized when B^2 is the reflection of B^1 according to the barrier $\frac{\eta}{2}$. The copula between those two Brownian motions presents two states of dependence: one of comonotonicity corresponding to a correlation of 1 and one of countermonotonicity corresponding to a correlation of -1. We release these two states of dependence by allowing lower correlations in absolute value. This gives the copula of Proposition 2.1. This copula is asymmetric and Proposition 2.2 gives the survival function of difference between

the two Brownian motions coupled with this copula:

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \Phi\Big(\frac{-x+2\rho h}{\sqrt{2\left(1-\rho\right)t}}\Big)\Phi\Big(\frac{x-2h\left(1+\rho\right)}{\sqrt{2\left(1+\rho\right)t}}\Big) + \Phi\Big(\frac{2h-x}{\sqrt{2\left(1-\rho\right)t}}\Big)\Phi\Big(\frac{-x}{\sqrt{2\left(1+\rho\right)t}}\Big).$$

This model of dependence gives higher values for $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right)$ than the constant correlation case and than $\frac{1}{2}$ when x close to 0 and for ρ high enough.

We generalize this model by allowing several reflections: it is the multi-barrier correlation model. We define two barriers ν and η with $\nu < \eta$. We consider two independent Brownian motions X and B^Y, and we construct the Brownian motion Y^n that is correlated to $\tilde{X}^n$:

$$Y^n = \rho\tilde{X}^n + \sqrt{1-\rho^2}B^Y,$$

with $\tilde{X}^n$ the Brownian motion equal to $-X$ at the beginning and reflecting when $X - Y^n$ hits a two-state barrier equal to η before the first reflection and switching from η to ν or from ν to η at each reflection. For a given $x \in [\eta, \nu]$ and $t > 0$, Corollary 2.1 states that the sequence $\mathbb{P}\left(X_t - Y_t^n \geq x\right)$ is increasing with n. Furthermore, the number of reflections in $[0, t]$ N_t is finite almost surely, see Proposition 2.3 (iii). We then consider the process $Y_t = Y_t^{N_t}$ which is a Brownian motion, see Proposition 2.3 (iv), that corresponds to the case $n \to \infty$. Proposition 2.4 gives the survival function of $X_t - Y_t$, which is higher than in the constant correlation case and higher than $\frac{1}{2}$ when $x \in [\eta, \nu]$ and ρ is high enough. This model can be transposed to a local correlation model:

$$\begin{cases} dX_t = dB_t^X \\ dY_t = \tilde{\rho}\left(X_t - Y_t\right)dB_t^X + \sqrt{1-\tilde{\rho}\left(X_t - Y_t\right)^2}dB_t^Y \end{cases}$$

with $\tilde{\rho}$ a Lipschitz function such that $\sup_{x\in\mathbb{R}}|\tilde{\rho}\left(x\right)| < 1$, $\tilde{\rho}\left(x\right) = \rho_1$ if $x \leq \nu$ and $\tilde{\rho}\left(x\right) = \rho_2$ if $x \geq \eta$. This system of stochastic differential has a strong solution (X, Y), see Proposition 2.5. This model seems to be equivalent to the multi-barrier model when the two barriers have close values and $\rho_2 = -\rho_1 = \rho$. The solution has the advantage to be Markovian.

The multi-barrier correlation model is applied to the factorial model (2.1) in order to model jointly forward prices of electricity and forward prices of coal. Empirical results show that the model works well for products with a long delivery maturity (3 Month Ahead and 6 Month Ahead): the difference between the two products has an asymmetric distribution and the probability for the electricity product to be higher than the coal one is high. However, it is not the case for products with a short delivery maturity, such as the spot. This can be explained by a difference of volatility too high between the electricity spot price and the coal spot price. Indeed, the electricity and coal volatilities of the long term factors that drives the prices of long maturity products are close to each other whereas they are very different for the short term factors. An other limitations of our model is that it is highly sensitive to initial conditions, that is the initial prices of electricity and coal products. We also estimate prices of European spread options in our model with Monte Carlo. Results are the same in the local correlation model.

1.3 Structure of the paper

In Section 2, we provide a first model to construct two Brownian motions with a two-state correlation structure based on the dependence between a Brownian motion and its reflection. We give

a closed formula for the survival function of the difference between the two Brownian motions: the distribution of the difference is asymmetric and can take higher values than in the constant correlation case. In Section 3, we improve the model of Section 2 by allowing several reflections to construct a multi-barrier correlation model. We give results about the survival function between the two Brownian motions and show that it takes higher values than the one in the model of Section 2. We also derive a local correlation model which gives the same results than the multi-barrier correlation model. Section 3 is our major contributions. In Section 4, we apply our results to the modeling of the forward prices of two commodities which are electricity and coal and to the pricing of spread options. Proofs are given in Section 5.

2 A two-state correlation copula

In this section, we derive a copula based on the Brownian motion and its reflection according to a barrier. As seen in [8], this copula contains two states depending on the value of the difference between the two Brownian motions: one of comonotonicity and one of countermonotonicity, that is correlation equal to 1 and -1. This copula maximizes $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ when the barrier is equal to $\frac{\eta}{2}$, see [8, Proposition 3]. However, the dependence between the two Brownian motions when it is modeled by these copulae is degenerated in the sense that the difference between the two Brownian motions becomes constant in an infinite horizon. In this section, we construct a copula which does not present this degeneracy but which allows higher values for $\mathbb{P}\left(B_t^1 - B_t^2 \geq \eta\right)$ than in the Gaussian copula case. The idea is to relax the correlation: instead of having states of correlation with correlations equals to 1 and -1, we have states of correlation with correlations equals to ρ and $-\rho$, $|\rho| < 1$.

2.1 Model

Let us consider a filtered probability space $(\Omega,\ \mathcal{F},\ (\mathcal{F}_t)_{t\geq 0},\ \mathbb{P})$ with $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual hypothesis (right continuity and completion) and $B^1 = \left(B_t^1\right)_{t\geq 0}$ a Brownian motion adapted to $(\mathcal{F}_t)_{t\geq 0}$. We denote by $\tilde{B}^h$ the Brownian motion reflection of B on $x = h$ with $h \in \mathbb{R}$, i.e. $\tilde{B}_t^h = -B_t^1 + 2(B_t^1 - B_{\tau^h}^1)\mathbf{1}_{t\geq \tau^h}$ with $\tau^h = \inf\{t \geq 0 : B_t^1 = h\}$. Thus, $\tilde{B}^k$ is a $\mathcal{F}$ Brownian motion according to the reflection principle (see [15, Theorem 3.1.1.2, p. 137]). Let $\rho \in (0,1)$ and Z a Brownian motion independent from B^1. We consider the stochastic process $B^2 = \rho\tilde{B}_t^h + \sqrt{1-\rho^2}Z$, which is a Brownian motion by Lévy characterisation.

2.2 The copula

Let us recall that a function $C : [0,1]^2 \mapsto [0,1]$ is a copula if:

(i) C is 2-increasing, i.e. $C\left(u_2, v_2\right) - C\left(u_1, v_2\right) + C\left(u_1, v_1\right) - C\left(u_2, v_1\right) \geq 0$ for $u_2 \geq u_1, v_2 \geq v_1$ and $u_1, u_2, v_1, v_2 \in [0,1]$,

(ii) $C\left(u, 0\right) = C\left(0, v\right) = 0,\ u, v \in [0,1]$,

(iii) $C\left(u, 1\right) = u, C\left(1, u\right) = u,\ u \in [0,1]$.

According to Skar's theorem [19], if X and Y are two random variables with continuous distribution function F^X and F^Y, there exists an unique copula C such that $\mathbb{P}\left(X \leq x, Y \leq y\right) = C\left(F^X\left(x\right), F^Y\left(y\right)\right)$. We will call C the copula of (X, Y).

In the following, we will denote by Φ the cumulative distribution function of a standard normal random variable and by Φ_ρ the cumulative distribution function of a bivariate gaussian vector of two standard normal random variables correlated with correlation ρ, $\rho \in (-1,1)$.

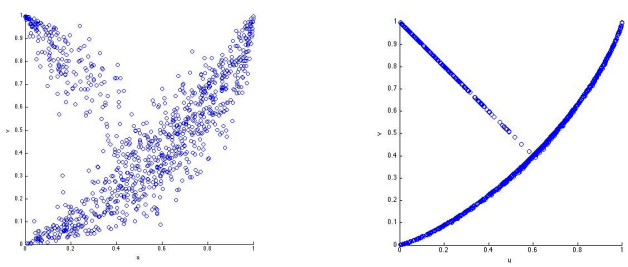
Proposition 2.1 gives the copula between B^1 and B^2.

Proposition 2.1 (Proposition 3 of [8])**.** *Let $h>0$, $t>0$ and $\rho \in (0,1)$. The copula*

$$C_t(u,v) = \begin{cases} \Phi_\rho\Big(\Phi^{-1}(u), \Phi^{-1}(v) + \frac{2\rho h}{\sqrt{t}}\Big) + v - \Phi\Big(\Phi^{-1}(v) + \frac{2\rho h}{\sqrt{t}}\Big) & \text{if } u \geq \Phi\Big(\frac{h}{\sqrt{t}}\Big) \\ \Phi_{-\rho}\Big(\Phi^{-1}(u), \Phi^{-1}(v)\Big) + \Phi_\rho\Big(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}, \Phi^{-1}(1-v) - \frac{2\rho h}{\sqrt{t}}\Big) + & \\ \Phi_\rho\Big(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}, \Phi^{-1}(v)\Big) - \Phi\Big(\Phi^{-1}(u) - \frac{2h}{\sqrt{t}}\Big) & \text{if } u < \Phi\Big(\frac{h}{\sqrt{t}}\Big), \end{cases}$$

is the copula between B_t^1 and B_t^2 at time t which are defined in the model of Section 2.1.

This copula is clearly asymmetric in the sense that $C_t(u,v) \neq C_t(v,u)$ for $u,v \in [0,1]$, $t>0$ which is a necessary condition if we want to have for $x>0$, $\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) \geq \frac{1}{2}$, see [8, Proposition 2]. The copula contains two states of correlation: one of positive dependence ($\rho > 0$) and one of negative dependence ($\rho < 0$). Figure 2.1 gives the copula of Proposition 2.1 with $\rho = 0.95$ and in the degenerated case $\rho = 1$, $h = 2$ and $t = 1$.



(a) $\rho = 0.95$. (b) $\rho = 1$.

Figure 2.1: *Copula between a Brownian motion and Brownian motion correlated to the refection of the first one with a correlation $\rho = 0.95$ and in the degenerated case $\rho = 1$ at time $t = 1$ and a barrier $h = 2$, which is the copula of Proposition 2.1.*

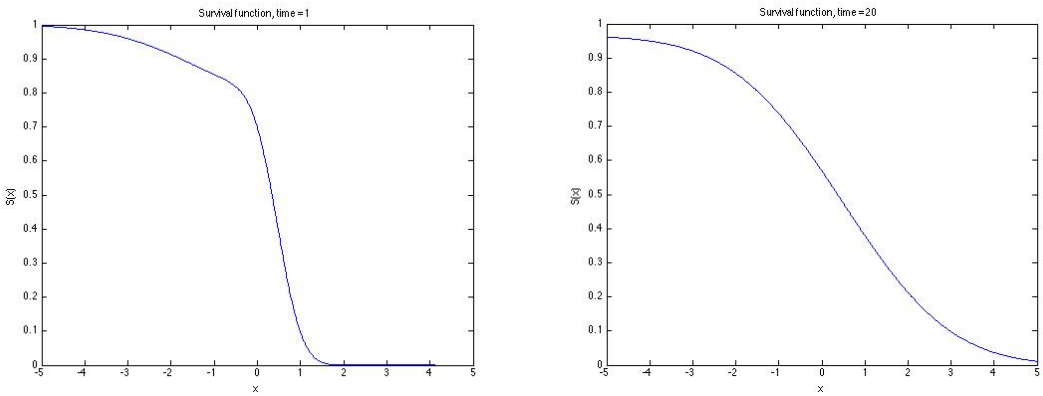
2.3 Distribution of the difference between the two Brownian motions

Proposition 2.2 gives the survival function of $B_t^1 - B_t^2$.

Proposition 2.2. *Let $t>0$, $h>0$, $\rho \in (0,1)$ and $x \in \mathbb{R}$. Let B^1 and B^2 the two Brownian motions defined in the model of Section 2.1. We have:*

$$\mathbb{P}\left(B_t^1 - B_t^2 \geq x\right) = \Phi\Big(\frac{-x+2\rho h}{\sqrt{2(1-\rho)t}}\Big)\Phi\Big(\frac{x-2h(1+\rho)}{\sqrt{2(1+\rho)t}}\Big) + \Phi\Big(\frac{2h-x}{\sqrt{2(1-\rho)t}}\Big)\Phi\Big(\frac{-x}{\sqrt{2(1+\rho)t}}\Big).$$

Figure 2.2 represents the survival function of $B_t^1 - B_t^2$ at time $t = 1$ and $t = 20$ with $h = 0.25$ and $\rho = 0.9$. The value of this function is close to 0.7 when x is close to 0 at time $t = 1$. However, when $t = 20$, it becomes close to $\frac{1}{2}$ and the asymmetry disappears.



(a) $t = 1$. (b) $t = 20$.

Figure 2.2: *Survival function of* $B_t^1 - B_t^2$ *in the model of Section 2.1 at time* $t = 1$ *and* $t = 20$ *with parameters* $h = 0.25$ *and* $\rho = 0.9$.

This model allows us to have higher values than in the Gaussian copula case for $\mathbb{P}\left(B_t^1 - B_t^2 \geq z\right)$ when z is close to 0. However, it presents some limitations in terms of modeling:

(i) $\left(B^1, \tilde{B}^h, B^2\right)$ is Markovian but the couple $\left(B, B^2\right)$ is not.

(ii) The asymmetry disappears in the distribution of $B_t^1 - B_t^2$ when t becomes large.

(iii) Let us consider the probability $\mathbb{P}\left(B_t^1 - B_t^2 \geq z \mid \mathcal{G}_s\right)$ with $\mathcal{G}$ the filtration generated by $\left(B^1, \tilde{B}^h, B^2\right)$. Let us suppose that the barrier has already been crossed at time s, i.e. $\tilde{B}_s^h = B_s^1 - 2h$. Thus, the correlation between B^1 and B^2 at times $t \geq s$ is equal to ρ and does not change. We are in the same case than in the Gaussian copula case after time s, and then we do not optimize $\mathbb{P}\left(B_t^1 - B_t^2 \geq z \mid \mathcal{G}_s\right)$.

3 Multi-barrier correlation model

In this section, we improve the model of Section 2 by allowing several reflections. In the model of Section 2, once the reflection has happened, the two Brownian motions stay correlated with correlation ρ even if the difference between the two becomes low. We want to have two Brownian motions X and Y with the following correlation structure: if the value of $X - Y$ is under a certain level that we denote by ν, X and Y have a negative correlation $-\rho$ and if it is over an other level denoted by η, their correlation is positive and equal to ρ. One way to obtain this structure is to start with two Brownian motions having a negative correlation. When the difference between them reaches the barrier η, Y reflects and the correlation becomes positive. If the correlation is positive (resp. negative) and $X - Y$ reaches ν (resp. η), Y reflects and the correlation becomes negative (resp. positive). The number of reflection that can happen is a parameter of our model denoted by n. Y is then correlated to a reflection of X reflecting each time the difference between the two reaches one of the two barriers. Figure 2.3 gives an illustration of our model. In Section 3.3, we develop a local correlation model based on the same principle. The local correlation model seems to

be equivalent to the multi-barrier correlation model when the two barriers are close. Furthermore, in the local correlation model, the couple (X, Y) is Markovian.

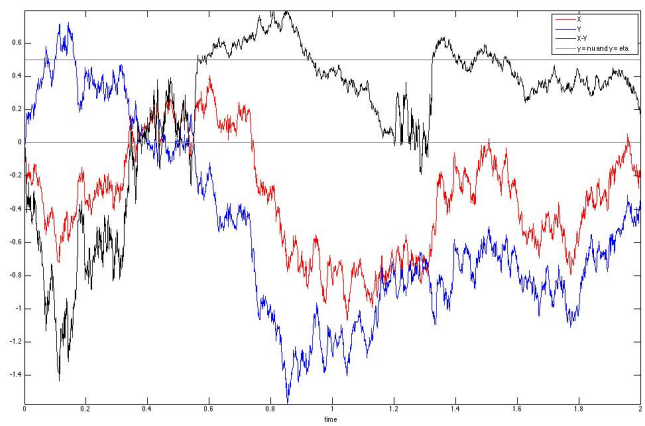


Figure 2.3: *One trajectory of X, Y and $X - Y$ in the multi-barrier correlation model with $\nu = 0$, $\eta = 0.5$, $\rho = 0.9$.*

3.1 Model

Let B^X and B^Y be two independent Brownian motions defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual properties. We will denote indifferently B^X by X.

Let $\eta > 0$, $\nu < \eta$ and $\rho \in [0, 1]$.

Let $\alpha_k = \begin{cases} 0 \text{ if } k = 0 \\ \eta \text{ if } k \text{ odd} \\ \nu \text{ if } k \text{ even, } k \neq 0 \end{cases}$.

Let $\left(\tilde{B}^k\right)_{k\geq 0}$, $\left(Y^k\right)_{k\geq 0}$ and $(\tau_k)_{k\geq 0}$ be defined by

$$\begin{cases} \tau_0 = 0 \\ \tilde{B}^0 = -B^X \\ Y_t^0 = \rho\tilde{B}^0 + \sqrt{1-\rho^2}B^Y \end{cases},$$

$$\begin{cases} \tau_k = \inf\{t \geq \tau_{k-1} : B_t^X - Y_t^{k-1} = \alpha_k\} \quad k \geq 1 \\ \tilde{B}^k = \mathcal{R}(\tilde{B}^{k-1}, \tau_k) \quad k \geq 1 \\ Y^k = \rho\tilde{B}^k + \sqrt{1-\rho^2}B^Y \quad k \geq 1, \end{cases}$$

where $\mathcal{R}(B, \tau)$ is the reflection Brownian motion of B with the reflection happening at time τ and τ a stopping time, i.e. $\mathcal{R}(B, \tau)_t = -B_t + 2(B_t - B_\tau)\mathbf{1}_{t\geq\tau}$.

Let $N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\tau_n \leq t}$ be the number of reflections that happened before time t and $Y_t = Y_t^{N_t}$. Y^N is well defined because $N_t < \infty$ almost surely according to Proposition 2.3 (iii). Proposition 2.3 gives results about the model.

Proposition 2.3. *(i)* $\left(Y^k\right)_{k\geq 0}$ *is a sequence of* $(\mathcal{F}_t)_{t\geq 0}$ *Brownian motions and* $(\tau_k)_{k\geq 0}$ *is a sequence of* $(\mathcal{F}_t)_{t\geq 0}$ *stopping times.*

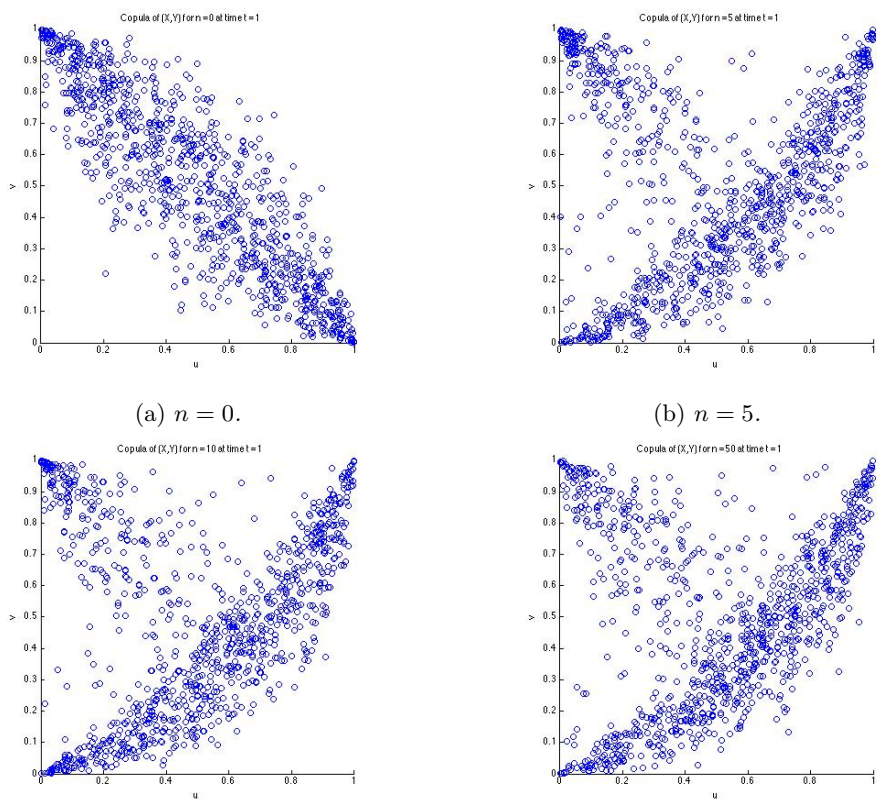
(ii) For $t > 0$,

$$X_t - Y_t^n = \begin{cases} \left(1+(-1)^k\rho\right)\left(B_t^X - B_{\tau_k}^X\right) - \sqrt{1-\rho^2}\left(B_t^Y - B_{\tau_k}^Y\right) + \alpha_k,\ \tau_k \leq t \leq \tau_{k+1},\ 0 \leq k \leq n \\ \left(1+(-1)^n\rho\right)\left(B_t^X - B_{\tau_{n+1}}^X\right) - \sqrt{1-\rho^2}\left(B_t^Y - B_{\tau_{n+1}}^Y\right) + \alpha_{n+1},\ \tau_{n+1} \leq t \end{cases}. \tag{2.2}$$

(iii) $N_t < \infty$ *almost surely.*

(iv) Y *is a Brownian motion.*

Figure 2.4 is the empirical copula of (X_t, Y_t^n) for different n at time $t = 1$. The copula is asymmetric and we observe two states of correlation, as for the model of Section 2.



(a) $n = 0$. (b) $n = 5$.

(c) $n = 10$. (d) $n = 50$.

Figure 2.4: *Empirical copula of* (X, Y^n) *in the multi-barrier correlation model at time* $t = 1$ *with parameters* $\nu = 0$, $\eta = 0.5$ *and* $\rho = 0.9$ *and a time step of 0.001 for different values of* n *done with 1000 simulations.*

3.2 Results on the distribution of the difference between the two Brownian motions

Proposition 2.4 gives an analytic formula for the survival function for $X_t - Y_t^n$ and $X_t - Y_t$.

Proposition 2.4. *Let $t > 0$ and $x \in \mathbb{R}$. Let $(p_n(t,x))_{n\geq 0}$ the sequence defined by:*

$$p_0(t,x) = \Phi\Big(\frac{-x}{\sqrt{2(1+\rho)t}}\Big), \tag{2.3}$$

$$p_n(t,x) = \begin{cases} \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2(1+(-1)^n\rho)t}} - \frac{u_{n+1}}{\sqrt{t}}\Big) - \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2\left(1+(-1)^{n+1}\rho\right)t}} - \frac{u_{n+1}}{\sqrt{t}}\Big) & \text{if} \quad x < \alpha_{n+1} \\ \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2(1+(-1)^n\rho)t}} + \frac{u_{n+1}}{\sqrt{t}}\Big) - \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2\left(1+(-1)^{n+1}\rho\right)t}} + \frac{u_{n+1}}{\sqrt{t}}\Big) & \text{if} \quad x \geq \alpha_{n+1} \end{cases} \tag{2.4}$$

where $(u_n)_{n\geq 0}$ is the sequence defined by:

$$\begin{cases} u_0 = 0 \\ u_n = \frac{\eta}{\sqrt{2(1+\rho)}} + \frac{(\eta-\nu)}{\sqrt{2}}\Big(\frac{\lfloor \frac{n}{2} \rfloor}{\sqrt{1-\rho}} + \frac{\lfloor \frac{n-1}{2} \rfloor}{\sqrt{1+\rho}}\Big) \quad k \geq 1 \end{cases}$$

and $\lfloor . \rfloor$ is the floor function.

We have:

$$\mathbb{P}\left(X_t - Y_t^n \geq x\right) = \sum_{k=0}^{n} p_k\left(t,x\right)$$

and

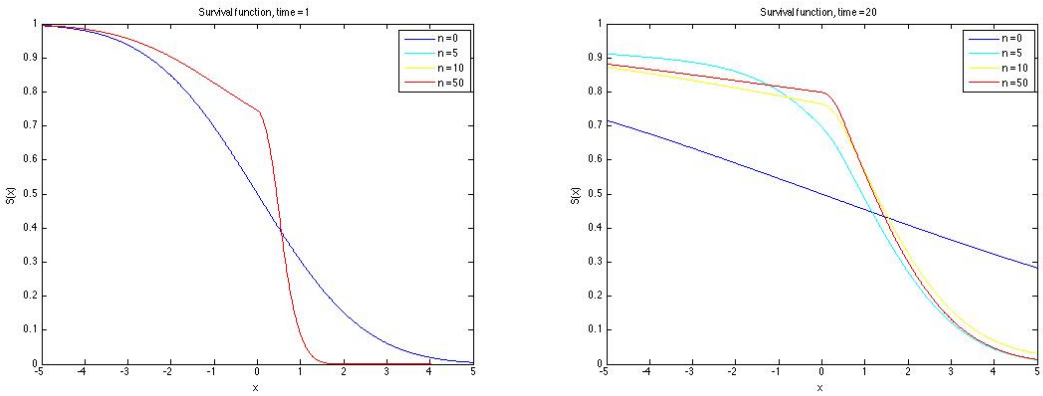
$$\mathbb{P}\left(X_t - Y_t \geq x\right) = \sum_{k=0}^{\infty} p_k\left(t,x\right).$$

Corollary 2.1. *Let $t > 0$. For $x \in [\nu,\eta]$, the sequence $\mathbb{P}\left(X_t - Y_t^n \geq x\right)$ is increasing with n when $\rho > 0$.*

For $x \in [\nu,\eta]$, the survival function takes higher values than in the constant correlation case and than $\frac{1}{2}$. Furthermore, it is possible to increase the value of $\mathbb{P}\left(X_t - Y_t^n \geq x\right)$ by increasing the number of reflections with this model, which is why the case $n = \infty$ is considered.

Results of Proposition 2.1 are illustrated in Figure 2.5a. The case $n = 0$ corresponds to the Gaussian case. We can see that in $[\nu,\eta]$, the survival function is increasing with n. In Figure 2.5a, the curves for $n = 5$, $n = 10$ and $n = 50$ are the same. At time $t = 1$, the probability to cross more than 5 barrier is very weak then the Brownian reflection reflects less than 5 times with a high probability. The convergence in n at small time is fast. In Figure 2.5b, we can observe the difference between the cases $n = 5$, $n = 10$ and $n = 50$ at time $t = 20$. The survival function continues to grow. The survival function does not present the problem of the one the model of Section 2: its value stays high when $t = 20$ which is caused by the several reflections.

The results are confirmed with Figure 2.6. The higher the number of reflections is, more $X - Y^n$ is concentrated in the region $[\nu,\eta]$. However, in the positive part of the plan, $X - Y^n$ take lower values than in the Gaussian case $n = 0$. One explanation comes from the martingality of $X - Y^n$. As $X - Y^n$ is a martingal, we have $\mathbb{E}\left(X_t - Y_t^n\right) = \mathbb{E}\left(X_0 - Y_0^n\right) = 0$. Furthermore, $\mathbb{P}\left(X_t - Y_t^n \geq 0\right) > \frac{1}{2}$ and is



(a) $t = 1$. (b) $t = 20$.

Figure 2.5: *Survival function of $X - Y^n$ in the multi-barrier correlation model at time t with parameters $\nu = 0$, $\eta = 0.5$ and $\rho = 0.9$ for different values of n.*

higher than in the case of a constant correlation between the two Brownian motions. The probability mass in the positive part of the real line increases, but the expectation on all the real line stay the same: values taken by the random variables become lower in the positive part of the real line and becomes higher in the negative.We also remark that the symmetry present in the case $n = 0$ disappears when n is higher.

3.3 A local correlation model

As in Section 3.1, we develop a model based on a two-states structure of correlation. However, we use a totally different approach where the reflection of the Brownian motion does not appear. Our model is a local correlation model and the correlation depends on the value of the difference between the two Brownian motions. The local correlation function presents two states of correlation: one of negative correlation if the difference of the two Brownian motion is under a certain barrier, one of positive correlation if the difference if over an other barrier and between the two barriers the function is chosen with sufficient regularity.

Let B^X and B^Y be two independent Brownian motions defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Let $\eta, \nu, \rho_{\min}$ and $\rho_{\max}$ be real numbers with $\eta > \nu$, $|\rho_{\min}| < 1$, $|\rho_{\max}| < 1$.

Let $\tilde{\rho}(x)$ be a function such that $\tilde{\rho}(x) = \rho_{\min}$ for $x \leq \nu$, $\tilde{\rho}(x) = \rho_{\max}$ for $x \geq \eta$ and $\sup_{x\in\mathbb{R}} |\tilde{\rho}(x)| < 1$.

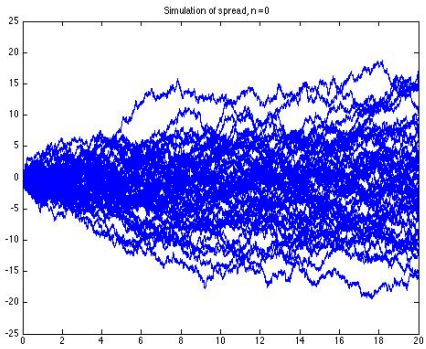
Let us assume that $\tilde{\rho}$ is Lipschitz.

Let us consider the following system of stochastic differential equations:

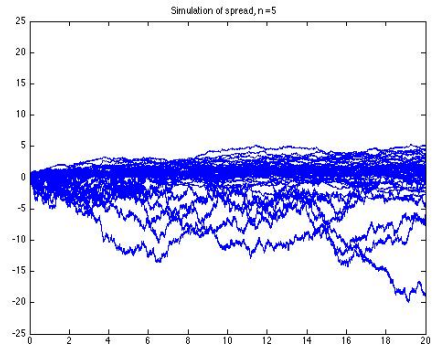
$$\begin{cases} dX_t = dB_t^X \\ dY_t = \tilde{\rho}(X_t - Y_t)\, dB_t^X + \sqrt{1 - \tilde{\rho}(X_t - Y_t)^2}\, dB_t^Y \end{cases} \tag{2.5}$$

with $X_0 = 0$ and $Y_0 = 0$.

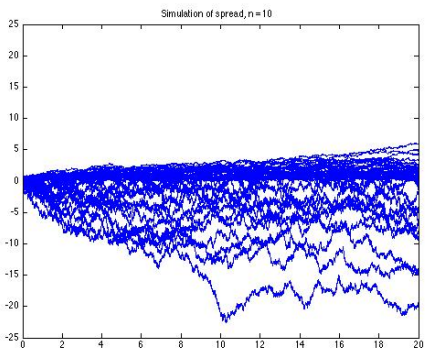
Proposition 2.5 gives results about the solution of (2.5).



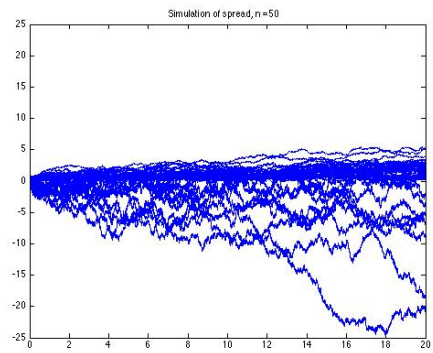
(a) $n = 0$.



(b) $n = 5$.



(c) $n = 10$.



(d) $n = 50$.

Figure 2.6: *50 simulations of $X - Y^n$ in the multi-barrier correlation model between time 0 and 20 with parameters $\nu = 0$, $\eta = 0.5$ and $\rho = 0.9$ and a time step of 0.001 for different values of n.*

Proposition 2.5. *The system of stochastic differential equations (2.5) has an unique strong solution (X, Y) with X and Y two Brownian motions. Furthermore, (X, Y) is Markovian.*

Contrary to the multi-barrier correlation model, the local correlation model has the advantage to give a Markovian solution, which has some importance in practice. However, less analytical results are available for this model. In the following, we gives empirical results about it. Results are close to the ones of the multi-barrier correlation model.

As the local correlation function is asymmetric, i.e. $\rho(x, y) \neq \rho(x, y)$, $x, y \in \mathbb{R}$, the copula of the solution of (2.5) is expected to be asymmetric. Figure 2.7 represents the copula of (X_t, Y_t) at time $t = 1$. It is similar to the one of the multi-barrier correlation model.

Figure 2.8 represents the survival function of the $X_t - Y_t$ in the local correlation model at time $t = 1$ and $t = 20$ with parameters $\nu = 0$, $\eta = 0.5$, $\rho_{\min} = -0.9$ and $\rho_{\max} = 0.9$. The local correlation function is chosen linear between ν and η. As for the multi-barrier correlation model, the distribution of $X_t - Y_t$ is asymmetric. The survival function seems equivalent to the one of the multi-barrier correlation model. Between ν and η, the survival function is over $\frac{1}{2}$ (Gaussian copula case). The survival function increases at the right of ν between time $t = 1$ and $t = 20$.

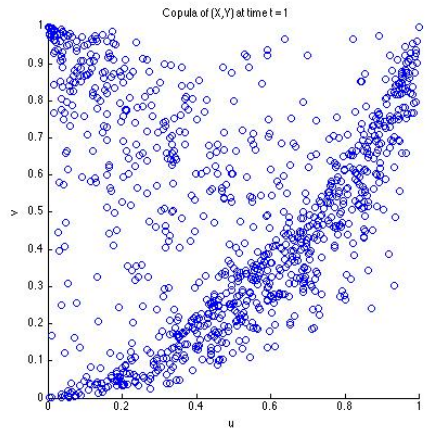
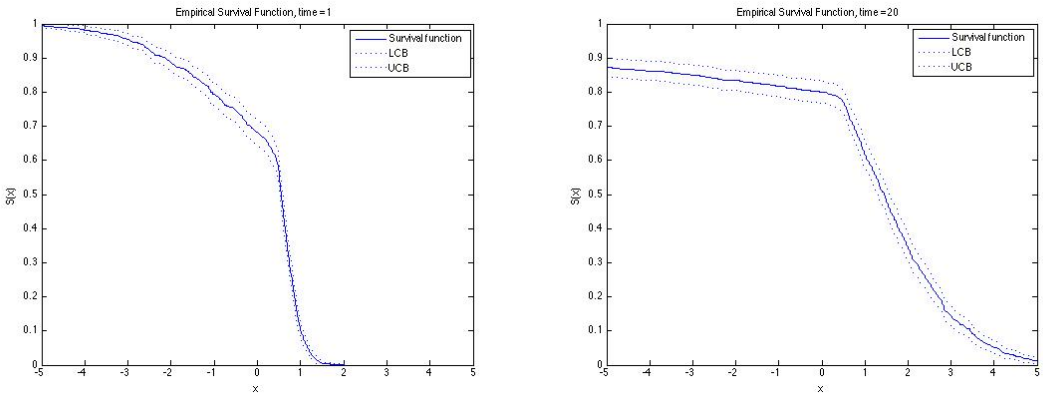


Figure 2.7: *Empirical copula of* (X_t, Y_t) *in the local correlation model at time* $t = 1$ *with parameters* $\nu = 0$, $\eta = 0.5$, $\rho_1 = -0.9$ *and* $\rho_2 = 0.9$ *and a time step of 0.001 with 1000 simulations.*



(a) $t = 1$. (b) $t = 20$.

Figure 2.8: *Empirical survival function of* $X_t - Y_t$ *in the local correlation model at time* t *with parameters* $\nu = 0$, $\eta = 0.5$, $\rho_{\min} = -0.9$ *and* $\rho_{\max} = 0.9$ *with interval confidence bounds at* 99% *and estimated with 1000 simulations and a step time of 0.001.*

Figure 2.9 represents 50 simulations of $X - Y$ in the correlation local model with parameters $\nu = 0$, $\eta = 0.5$, $\rho_{\min} = -0.9$ and $\rho_{\max} = 0.9$. As for the multi-barrier correlation model, the trajectories are concentrated in the positive part of the plan.

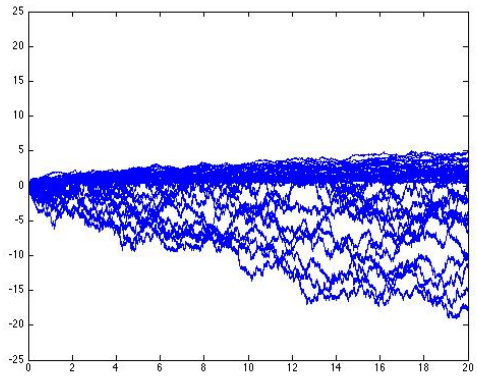


Figure 2.9: *50 simulations of $X-Y$ in the correlation local model with parameters $\nu = 0$, $\eta = 0.5$, $\rho_{\min} = -0.9$ and $\rho_{\max} = 0.9$ between time $t = 0$ and $t = 20$ and a time step of 0.001.*

4 An application for joint modeling of commodity prices on energy market

In this section, we use the multi-barrier correlation model for the joint modeling of the forward prices of two commodities, electricity and coal. Coal is a fuel used to produce electricity which implies an asymmetry in the distribution of the difference between the two prices ; it is more likely that price of coal is lower than price of electricity (in the same unit). Modeling the dependence with a Gaussian copula is then not adapted. An advantage of our model is that it contains asymmetry in the distribution of the difference between the two prices. Furthermore, it allows not to change the marginal models.

4.1 Model

Let us consider a two-factor model for both electricity and coal. For more information on the two-factor model, we refer to the study of Benth and Koekebakker [3].

Let $f^E(t,T)$ (resp. $f^C(t,T)$) the forward price of the electricity (resp. coal) at time t with maturity T, that is of the delivery of electricity (resp. coal) at maturity T during one day. Stochastic differential equation (2.6) gives dynamic of these products.

$$\begin{cases} df^E(t,T) = f^E(t,T)\left(\sigma_s^E e^{-\alpha_s^E(T-t)} dB_t^{E,s} + \sigma_l^E dB_t^{E,l}\right) \\ df^C(t,T) = f^C(t,T)\left(\sigma_s^C e^{-\alpha_s^C(T-t)} dB_t^{C,s} + \sigma_l^C dB_t^{C,l}\right) \end{cases} \tag{2.6}$$

where $B^{E,s}$, $B^{E,l}$, $B^{C,s}$, $B^{C,l}$ are standard Brownian motions defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

In the dynamic of each commodity, there is one factor corresponding to the short term factor with a volatility $\sigma_s^i e^{-\alpha_s^i(T-t)}, i = E, C$. This short term factor is used to model the Samuelson effect [18], that is the decrease of volatility with time to maturity. The other factor is the long term factor with a constant volatility $\sigma_l^i, i = E, C$.

Products traded on the market have a delivery period, except for the spot. We denote by $f^i(t,T,\theta), i = E, C$ the price of the product at time t that delivers i at time T during a period θ. By absence of arbitrage opportunities, we have

$$f^i(t,T,\theta) = \frac{1}{\theta}\int_T^{T+\theta} f^i(t,u)\,du.$$

In the following, we will only consider n Month Ahead (nMAH), $n \geq 1$, which are products with a delivery period of one month and a delivery date which is the 1st of the nth following month from today.

Equation (2.7) gives the solutions of (2.6).

$$\begin{cases} f^E(t,T) = f^E(0,T)\, e^{\int_0^t \sigma_s^E e^{-\alpha_s^E(T-u)} dB_u^{E,s} - \frac{1}{2}\int_0^t \left(\sigma_s^E\right)^2 e^{-2\alpha_s^E(T-u)} du + \sigma_l^E B_t^{E,l} - \frac{1}{2}\left(\sigma_l^E\right)^2 t} \\ f^C(t,T) = f^C(0,T)\, e^{\int_0^t \sigma_s^C e^{-\alpha_s^C(T-u)} dB_u^{C,s} - \frac{1}{2}\int_0^t \left(\sigma_s^C\right)^2 e^{-2\alpha_s^C(T-u)} du + \sigma_l^C B_t^{C,l} - \frac{1}{2}\left(\sigma_l^C\right)^2 t} \end{cases} \tag{2.7}$$

The spot price of electricity is given by $S_t^E = f^E(t,t)$ and the one of coal by $S_t^C = f^C(t,t)$. Then we have

$$\begin{cases} S_t^E = f^E(0,t)\, e^{\int_0^t \sigma_s^E e^{-\alpha_s^E(t-u)} dB_u^{E,s} - \frac{1}{2}\int_0^t \left(\sigma_s^E\right)^2 e^{-2\alpha_s^E(t-u)} du + \sigma_l^E B_t^{E,l} - \frac{1}{2}\left(\sigma_l^E\right)^2 t} \\ S_t^C = f^C(0,t)\, e^{\int_0^t \sigma_s^C e^{-\alpha_s^C(t-s)} dB_u^{C,s} - \frac{1}{2}\int_0^t \left(\sigma_s^C\right)^2 e^{-2\alpha_s^C(t-u)} du + \sigma_l^C B_t^{C,l} - \frac{1}{2}\left(\sigma_l^C\right)^2 t} \end{cases} \tag{2.8}$$

We model the dependence as follow:

- $B^{E,s}$ and $B^{E,l}$ are independent,
- $B^{C,s}$ and $B^{C,l}$ are independent,
- $B^{E,s}$ and $B^{C,s}$ are independent,
- $B^{E,l}$ and $B^{C,l}$ are constructed following the multi-barrier correlation model defined in Section 3.

Usually, a constant correlation matrix is used to model the dependence between the 4 Brownian motions.

4.2 Parameters

We consider the parameters of the marginal laws given in Table 2.1. Units are taken according to the year. We use the forward prices of electricity and of coal during 2014 in France to estimate these parameters. The method used for estimation is the first one of [10].

Parameters	Electricity	Coal
σ_l	10.2555%	9.2602%
σ_s	97.2925%	11.2134%
α_s	17.0363	2.07832

Table 2.1: *Parameters of the two-factor model for electricity and coal.*

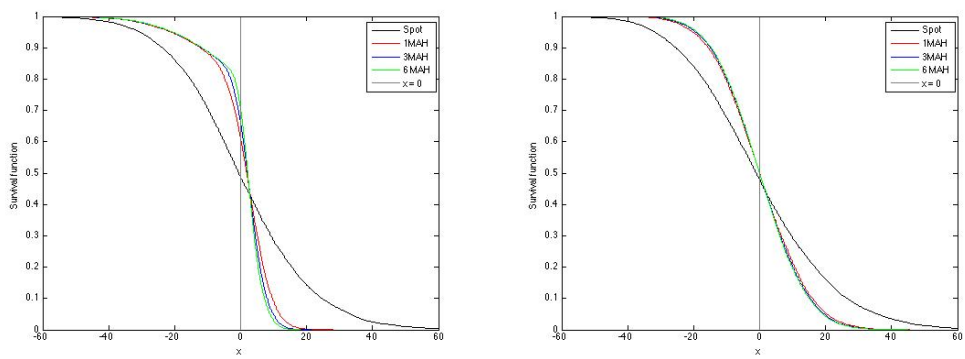
Parameters for the multi-barrier correlation model used to model the dependence between $B^{E,l}$ and $B^{C,l}$ are chosen arbitrarily ; we choose $\nu = 0$, $\eta = 0.5$, $\rho = 0.9$, $n = \infty$.

In the benchmark model where dependence between $B^{E,l}$ and $B^{C,l}$ is modeled by a constant correlation, the correlation is equal to 0.275. The other correlation are equals to 0.

We assume that $f^E(0,T) - Hf^C(0,T) = 0$ and $f^E(0,T) = 100$ for all T (which does not represent the reality because we do not take into account the seasonality of the prices of electricity and coal). H is a conversion factor between the unit of electricity prices and the unit of coal prices and is called the heat rate.

4.3 Numerical results

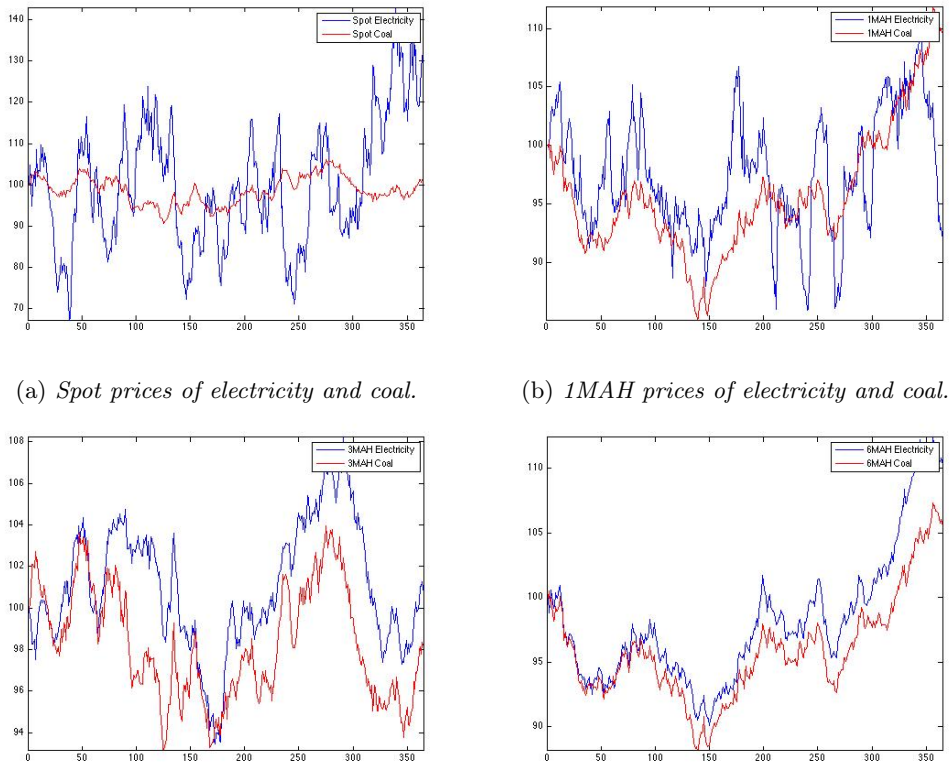
We are interested in the difference between $f^E(t,T)$ and $Hf^C(t,T)$. We only are interested in the multi-barrier correlation model ; results are the same for the local correlation model.



(a) *Multi-barrier correlation model.* (b) *Benchmark model.*

Figure 2.10: *Empirical survival function of the difference between the price of electricity and the price of coal at time $t = 365$ days estimated with 10000 simulations with a time step of $\frac{1}{24}$ days for different products (Spot, 1MAH, 3MAH, 6MAH) in the multi-barrier correlation model and in the benchmark model.*

Figure 2.10 represents the survival function of the difference between spot, 1MAH, 3MAH, and 6MAH prices. In the multi-barrier correlation model, the probability for the difference between the two spot prices to be non negative is close to 50%, which is the same value than in the benchmark model. However, we have good results if we consider long term products as 1MAH, 3MAH and 6MAH: we have probabilities closed to 60% for the 1MAH, and 70% for the 3MAH and 6MAH in the multi-barrier correlation model whereas we have probabilities closed to 50% in the benchmark model. The probability increases with the time to maturity. In the case of spot prices, the volatilities of the prices of the commodities is dominated by the short term factor, which we do not control ; in the other cases, these volatilities are small and the long term factor which we control dominates. This explains that we do not increase a lot the probability for the difference between the spot prices to be non negative. We also observed that in the multi-barrier correlation model, the survival function decreases faster than in the benchmark model and probability of being superior to 20 is closed to 0, which is not the case in the benchmark model.



(a) *Spot prices of electricity and coal.*

(b) *1MAH prices of electricity and coal.*

(c) *3MAH prices of electricity and coal.*

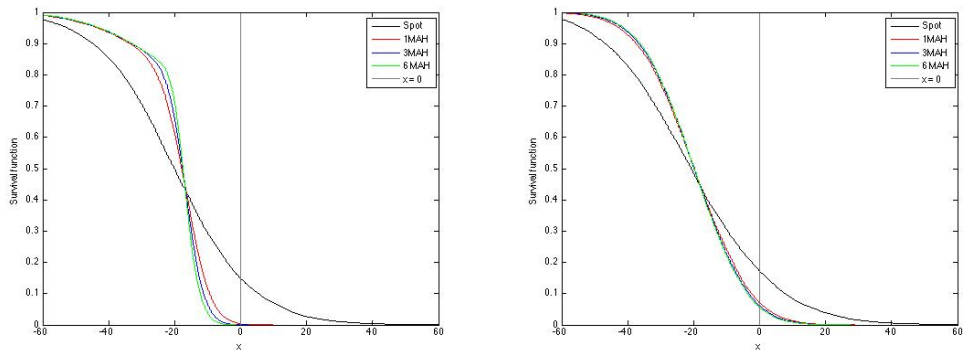
(d) *6MAH prices of electricity and coal.*

Figure 2.11: *One year trajectory of electricity and coal products in the multi-barrier correlation model with a time step of* $\frac{1}{24}$ *days.*

Figure 2.11 represents one trajectory of the different products. In the case of the spot prices, since electricity has a high volatility, it is difficult to control the difference between the two processes. For the other products, as the short term volatility decreases, we see that there is a control between the two processes.

Remark 2.1. *Using a multi-barrier correlation model to model the dependence between* $B^{E,s}$ *and* $B^{C,s}$ *does not improve the results for the different survival functions. That is why we consider them independent.*

Results are sensitive to initial conditions. If we choose $f^E(0,T) = 100$ and $Hf^C(0,T) = 120$ for instance, $f^E(0,T) - Hf^C(0,T) = -20$ and we will have a distribution that is concentrated around -20, because the difference between the price is a martingale. The probability to be greater than -20 is higher in the multi-barrier correlation model than in the benchmark model but the probability to be positive is lower than in the benchmark model: it is closed to 0 in the multi-barrier correlation model whereas it is closed to 10% in the benchmark model. Figure 2.12 represents the survival function of the difference between prices of electricity and coal for different products with $\nu = 0$ and $\eta = 0.5$. As we choose a barrier near 0, the survival function will be maximized around -20.



(a) *Multi-barrier correlation model.* (b) *Benchmark model.*

Figure 2.12: *Empirical survival function of the difference between the price of electricity and the price of coal at time $t = 335$ days estimated with 10000 simulations with a time step of $\frac{1}{24}$ days for different products (Spot, 1MAH, 3MAH, 6MAH) in the multi-barrier correlation model and in the benchmark model if the difference is equal to -20 at time $t = 0$.*

One way to improve the value of the survival function around 0 is to choose a higher η. The idea in our model is that we want $B^{E,l}$ to go over $B^{C,l} + \eta$, using correlation of -1 when the two prices are equals at time $t = 0$. We want for the price of the electricity to go over the price of coal, that happens when $f^E(t,T) = Hf^C(t,T)$, i.e. when $\sigma_l^E B_t^{E,l} - \sigma_l^C B_t^{C,l} = \log\left(\frac{Hf^C(0,T)}{f^E(0,T)}\right)$ if we neglect the short term factors. We have $\sigma_l^E \approx \sigma_l^C \approx \sigma = 0.1$ year^{-1}. Then, we want $B_t^{E,l} - B_t^{C,l} \approx \frac{1}{\sigma}\log\left(\frac{Hf^C(0,T)}{f^E(0,T)}\right)$. In the case with the same initial conditions, the right hand side term is equal to 0 and we choose a barrier of η. Heuristically, we then choose a barrier of $\eta' = \eta + \frac{1}{\sigma}\log\left(\frac{Hf^C(0,T)}{f^E(0,T)}\right) \approx 170.5$ and $\nu = 170$. Figure 2.13 gives the survival function of the different products in the multi-barrier correlation model with barriers $\nu = 170$ and $\eta = 170.5$.

We can see that around 0, the values of the survival function are much better than in the benchmark model: around 20% in the multi-barrier correlation model and around 10% in the benchmark model. However, the values are still low. Indeed, even in the maximal case where the second Brownian motion is the reflection of the first one and the volatilities are equals, the probability for the difference between the Brownian motions to be positive knowing that one starts at $-x$, $x > 0$ and the other at 0 is equal to $2\Phi\left(\frac{-x}{2\sqrt{t}}\right)$ which decreases with x.

4.4 Pricing of European spread options

In this Section, we compare prices of European spread options in the factorial model (2.6) with different structures of dependence: the multi-barrier correlation model (m-b) with correlation equals to 0.3, 0.6, 0.9 and the benchmark model (constant correlation) with correlation equals to 0 and 0.275. Benchmark model with correlation equal to 0 is the same model than multi-barrier correlation model with correlation equal to 0. We price options with payoff $(X_t - HY_t)^+$, where X_t is an electricity product, Y_t a coal product and H is the conversion factor between electricity and coal. X_t and Y_t are Spot, 1MAH, 3MAH and 6MAH. Parameters used are those of Table 2.1. The price

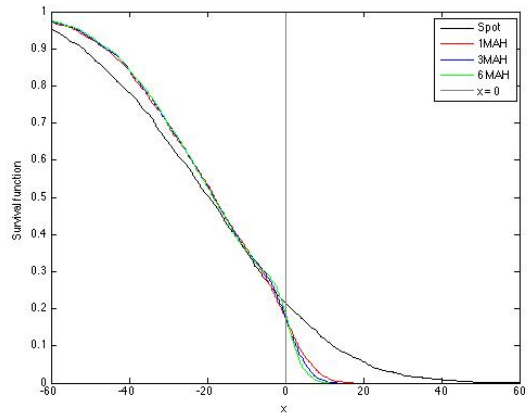


Figure 2.13: *Empirical survival function of the difference between the price of electricity and the price of coal at time $t = 335$ days estimated with 10000 simulations with a time step of $\frac{1}{24}$ days for different products (Spot, 1MAH, 3MAH, 6MAH) in the multi-barrier correlation model if the difference is equal to -20 at time $t = 0$ and with barriers $\nu = 170$ and $\eta = 170.5$.*

of the option is equal to $\mathbb{E}\left((X_t - Y_t)^+\right)$. We use Monte Carlo to estimate this expectation with a number of simulations equal to 10000. To simulate the processes, we use a step time of 1 hour.

Table 2.2 gives 95% confidence intervals for the price of spread options with maturity 1 year when $X_0 = HY_0 = 100$. For the multi-barrier correlation model, we choose $\nu = 0$ and $\eta = 0.5$. In the multi-barrier correlation model, the value of the option decreases with the correlation parameters. Indeed, when the correlation parameters increases, the probability to be over 0 is higher, but the values taken by the difference $X_t - HY_t$ are smaller and smaller. The increase in the probability do not compensate the decrease in the values that can be taken and the expectation, i.e. the value of the option decreases. Value of the option in the benchmark model with correlation equal to 0.275 is close to the one in the multi-barrier correlation model with correlation equal to 0.6. We also observe that the value of the option decreases with the product maturity, in all the models. Table 2.3 gives

Products / Parameters	$\rho = 0$	$\rho = 0.3$, m-b	$\rho = 0.6$, m-b	$\rho = 0.9$, m-b	$\rho = 0.275$, benchmark
Spot	$[8.39, 8.92]$	$[8.44, 8.96]$	$[7.87, 8.37]$	$[7.29, 7.75]$	$[7.69, 8.19]$
1MAH	$[6.54, 6.94]$	$[6.56, 6.94]$	$[5.96, 6.30]$	$[5.00, 5.29]$	$[5.80, 6.16]$
3MAH	$[5.45, 5.78]$	$[5.41, 5.70]$	$[4.79, 5.03]$	$[3.27, 3.41]$	$[4.72, 5.00]$
6MAH	$[5.33, 5.69]$	$[5.26, 5.55]$	$[4.65, 4.87]$	$[3.02, 3.15]$	$[4.60, 4.88]$

Table 2.2: *Values of European Spread options $(X_t - HY_t)^+$ between electricity and coal products in the benchmark model and in the multi-barrier correlation model with parameters $\nu = 0$, $\eta = 0.5$ with $X_0 = HY_0 = 100$.*

95% confidence intervals for the price of spread options with maturity 1 year when $X_0 = 100$ and $HY_0 = 120$. For the multi-barrier correlation model, we choose $\nu = 170$ and $\eta = 170.5$. Contrarily to results of Table 2.2, the value of the option increases with the correlation parameter in the multi-barrier correlation model. Furthermore, the value of the option in the multi-barrier case is greater than the one of the benchmark model, for all the given correlations. In the constant

correlation case, the probability to be greater than 0 is very low. The increase of probability in the multi-barrier correlation model is enough for the option value to be higher.

Products / Parameters	$\rho = 0$	$\rho = 0.3$, m-b	$\rho = 0.6$, m-b	$\rho = 0.9$, m-b	$\rho = 0.275$, benchmark
Spot	$[2.52, 2.83]$	$[2.92, 3.25]$	$[3.03, 3.36]$	$[3.13, 3.48]$	$[2.09, 2.37]$
1MAH	$[1.24, 1.42]$	$[1.57, 1.77]$	$[1.72, 1.92]$	$[1.74, 1.98]$	$[0.88, 1.02]$
3MAH	$[0.67, 0.79]$	$[0.9, 1.02]$	$[1.03, 1.15]$	$[0.81, 0.90]$	$[0.37, 0.45]$
6MAH	$[0.63, 0.74]$	$[0.82, 0.94]$	$[0.92, 1.03]$	$[0.67, 0.74]$	$[0.33, 0.41]$

Table 2.3: *Values of European Spread options* $(X_t - HY_t)^+$ *between electricity and coal products in the benchmark model and in the multi-barrier correlation model with parameters* $\nu = 170$, $\eta = 170.5$ *with* $X_0 = 100$ *and* $HY_0 = 120$.

5 Proofs

5.1 Preliminary results

We start with well known results that will be useful for the proofs of propositions.

Lemma 2.1. *Let* $B = (B_t)_{t\geq 0}$ *be a standard Brownian motion on a filtered probability space* $\Big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\Big)$. *We have:*

(i) for $y \geq 0$,

$$\mathbb{P}\Big(B_t \leq x, \sup_{s\leq t} B_s \leq y\Big) = \begin{cases} \Phi\Big(\frac{x}{\sqrt{t}}\Big) - \Phi\Big(\frac{x-2y}{\sqrt{t}}\Big) & \text{if} \quad x < y \\ 2\Phi\Big(\frac{y}{\sqrt{t}}\Big) - 1 & \text{if} \quad x \geq y \end{cases},$$

(ii) for $y \leq 0$,

$$\mathbb{P}\Big(B_t \leq x, \inf_{s\leq t} B_s \leq y\Big) = \begin{cases} \Phi\Big(\frac{x}{\sqrt{t}}\Big) & \text{if} \quad x \leq y \\ 2\Phi\Big(\frac{y}{\sqrt{t}}\Big) - \Phi\Big(\frac{-x+2y}{\sqrt{t}}\Big) & \text{if} \quad x > y \end{cases}.$$

Proof The reader is referred to [15, Theorem 3.1.1.2, p. 137] for the proof of (i) and to [15, Section 3.1.5, p. 142] for the proof of (ii). □

Lemma 2.2. *Let* $B^1 = \big(B^1_t\big)_{t\geq 0}$ *and* $B^2 = \big(B^2_t\big)_{t\geq 0}$ *be two independent standard Brownian motion defined on a common filtered probability space* $\Big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\Big)$ *with* $(\mathcal{F}_t)_{t\geq 0}$ *having all the good properties. Let* $h \geq 0$ *and* $\tau^h = \inf\{t \geq 0 : B^2_t = h\}$. *We have:*

$$\mathbb{P}\Big(B^1_t - B^1_{\tau^h} \leq x, \tau^h \leq t\Big) = \Phi\Big(\frac{x-h}{\sqrt{t}}\Big)\mathbf{1}_{x<0} + \Big(\Phi\Big(\frac{x+h}{\sqrt{t}}\Big) - 2\Phi\Big(\frac{h}{\sqrt{t}}\Big) + 1\Big)\mathbf{1}_{x\geq 0}.$$

Proof Conditional on $\{t \geq \tau^h\}$, $B^1_t - B^1_{\tau^h}$ is a Brownian motion independent to $\mathcal{F}_{\tau^h}$. Then

$$\mathbb{P}\Big(B^1_t - B^1_{\tau^h} \leq x, \tau^h \leq t\Big) = \mathbb{E}\Big(\Phi\Big(\frac{x}{\sqrt{t-\tau^h}}\Big)\mathbf{1}_{t\geq\tau^h}\Big).$$

The same argument can be used to prove that

$$\mathbb{P}\Big(B^2_t - B^2_{\tau^h} \leq x, \tau^h \leq t\Big) = \mathbb{E}\Big(\Phi\left(\frac{x}{\sqrt{t-\tau^h}}\right)\mathbf{1}_{t\geq\tau^h}\Big).$$

Then we have

$$\begin{aligned}\mathbb{P}\Big(B^1_t - B^1_{\tau^h} \leq x, \tau^h \leq t\Big) &= \mathbb{P}\Big(B^2_t - B^2_{\tau^h} \leq x, \tau \leq t\Big)\\ &= \mathbb{P}\Big(B^2_t \leq x+h, \sup_{s\leq t} B^2_s \geq h\Big).\end{aligned}$$

We can conclude using Lemma 2.1. □

Lemma 2.3. *Let a, b and $x \in \mathbb{R}$. We have:*

(i)

$$\int_{-\infty}^{x} \Phi\left(au+b\right)\frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}}du = \Phi_{\frac{-a}{\sqrt{a^2+1}}}\Big(\frac{b}{\sqrt{a^2+1}}, x\Big).$$

(ii)

$$\Phi_{\sqrt{1-\rho^2}}\left(x,y\right) = \Phi\left(y\right)\Phi\Big(\frac{x-\sqrt{1-\rho^2}y}{\rho}\Big) + \Phi\left(x\right) - \Phi_\rho\Big(x, \frac{x-\sqrt{1-\rho^2}y}{\rho}\Big),\ x,y\in\mathbb{R},\ \rho>0$$

(iii)

$$\Phi_\rho\left(x,y\right) = \Phi\left(y\right) - \Phi_{-\rho}\left(-x,y\right),\ x,y\in\mathbb{R}$$

Proof The reader is referred to [8, Proof of Lemma 19, Section 5.3]. □

5.2 Proof of Proposition 2.2

Let B^1 and Z two independent Brownian motion. We consider $B^2 = \rho\tilde{B}^h + \sqrt{1-\rho^2}Z$ with $\tilde{B}^h$ the reflection of B according to the barrier h. We have:

$$\mathbb{P}\left(B^1_t - B^2_t \geq x\right) = \mathbb{P}\Big(B^1_t - B^2_t \geq x, \sup_{s\leq t} B^1_s \leq h\Big) + \mathbb{P}\Big(B^1_t - B^2_t \geq x, \sup_{s\leq t} B^1_s \geq h\Big)$$

When $\sup_{s\leq t} B^1_s \leq h$, $B^2_t = -\rho B^1_t + \sqrt{1-\rho^2}Z_t$ and when $\sup_{s\leq t} B^1_s \geq h$, $B^2_t = \rho B^1_t - 2h\rho + \sqrt{1-\rho^2}Z_t$. Thus, $\mathbb{P}\left(B^1_t - B^2_t \geq x\right)$ is the sum of the three following terms:

(i) $\mathbb{P}\Big(B^1_t \leq \frac{x-2\rho h+\sqrt{1-\rho^2}Z_t}{(1-\rho)}, \sup_{s\leq t} B^1_s \leq h\Big)$,

(ii) $-\mathbb{P}\Big(B^1_t \leq \frac{x+\sqrt{1-\rho^2}Z_t}{(1+\rho)}, \sup_{s\leq t} B^1_s \leq h\Big)$,

(iii) $\mathbb{P}\Big((1-\rho)\,B_t^1 - \sqrt{1-\rho^2}Z_t \geq x - 2\rho h\Big)$.

Since B^1 and Z are independent, (i) is equal to the sum of the three following terms:

$$\mathbb{E}\Big(\Phi\Big(\frac{x-2\rho h+\sqrt{1-\rho^2}Z_t}{(1-\rho)\sqrt{t}}\Big)\mathbf{1}_{Z_t\leq\frac{h(1+\rho)-x}{\sqrt{1-\rho^2}}}\Big), \tag{2.9}$$

$$-\,\mathbb{E}\Big(\Phi\Big(\frac{x-2h+\sqrt{1-\rho^2}Z_t}{(1-\rho)\sqrt{t}}\Big)\mathbf{1}_{Z_t\leq\frac{h(1+\rho)-x}{\sqrt{1-\rho^2}}}\Big) \tag{2.10}$$

and

$$\mathbb{E}\Big(\Big(2\Phi\Big(\frac{h}{\sqrt{t}}\Big)-1\Big)\mathbf{1}_{Z_t\geq\frac{h(1+\rho)-x}{\sqrt{1-\rho^2}}}\Big) \tag{2.11}$$

with the use of Lemma 2.1. According to Lemma 2.3 (i), (2.9) is equal to

$$\Phi_{-\sqrt{\frac{1+\rho}{2}}}\Big(\frac{x-2\rho h}{\sqrt{2\,(1-\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.12}$$

Using Lemma 2.3 (ii), we find that the first term of (i) (2.12) is equal to

$$-\,\Phi_{\sqrt{\frac{1+\rho}{2}}}\Big(\frac{-x+2\rho h}{\sqrt{2\,(1-\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big)+\Phi\Big(\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.13}$$

In the same way, the second term of (i) (2.10) is equal to:

$$\Phi_{\sqrt{\frac{1+\rho}{2}}}\Big(\frac{-x+2h}{\sqrt{2\,(1-\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big)-\Phi\Big(\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.14}$$

The last one (2.11) is equal to

$$\Big(2\Phi\Big(\frac{h}{\sqrt{t}}\Big)-1\Big)\Phi\Big(\frac{x-h\,(1+\rho)}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.15}$$

Using the same scheme of proof that for (i), we find that (ii) is equal to the sum of the three following terms:

$$\Phi_{\sqrt{\frac{1-\rho}{2}}}\Big(\frac{-x}{\sqrt{2\,(1+\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big)-\Phi\Big(\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big), \tag{2.16}$$

$$-\,\Phi_{\sqrt{\frac{1-\rho}{2}}}\Big(\frac{-x+2h\,(1+\rho)}{\sqrt{2\,(1+\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big)+\Phi\Big(\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big) \tag{2.17}$$

and

$$-\,\Big(2\Phi\Big(\frac{h}{\sqrt{t}}\Big)-1\Big)\Phi\Big(\frac{x-h\,(1+\rho)}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.18}$$

Using Lemma 2.3 (iii), we find that (2.16) is equal to

$$\Phi\Big(\frac{-x}{\sqrt{2\,(1+\rho)\,t}}\Big)\Phi\Big(\frac{-x+2h}{\sqrt{2\,(1-\rho)\,t}}\Big)-\Phi_{\sqrt{\frac{1+\rho}{2}}}\Big(\frac{-x+2h}{\sqrt{2\,(1-\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.19}$$

and that (2.17) to

$$-\Phi\Big(\frac{-x+2h\,(1+\rho)}{\sqrt{2\,(1+\rho)\,t}}\Big)\Phi\Big(\frac{-x+2\rho h}{\sqrt{2\,(1+\rho)\,t}}\Big)+\Phi_{\sqrt{\frac{1+\rho}{2}}}\Big(\frac{-x+2\rho h}{\sqrt{2\,(1-\rho)\,t}},\frac{h\,(1+\rho)-x}{\sqrt{(1-\rho^2)\,t}}\Big). \tag{2.20}$$

Finally, we have (iii) equal to

$$\Phi\Big(\frac{-x+2\rho h}{\sqrt{2\,(1-\rho)\,t}}\Big). \tag{2.21}$$

$\mathbb{P}\left(B_t^1-B_t^2\geq x\right)$ is the sum of (2.13), (2.14), (2.15), (2.19), (2.20), (2.18) and (2.21).

5.3 Proof of Proposition 2.3

(i) This part of the proof can be done by induction.

(ii) For $\tau_0=0\leq t\leq\tau_1$, $X_t-Y_t^n=(1+\rho)\,B_t^X-\sqrt{1-\rho^2}B_t^Y$. The equality holds for k $=0$.

Let us suppose that the property true at rank $k<n+1$, that is

$$X_t-Y_t^n=\Big(1+(-1)^k\,\rho\Big)\left(B_t^X-B_{\tau_k}^X\right)-\sqrt{1-\rho^2}\left(B_t^Y-B_{\tau_k}^Y\right)+\alpha_k,\quad \tau_k\leq t\leq\tau_{k+1}.$$

If $\tau_k\leq t\leq\tau_{k+1}$, $Y_t^n=\rho\tilde{B}_t^k+\sqrt{1-\rho^2}B_t^Y$ then

$$X_t-\rho\tilde{B}_t^k-\sqrt{1-\rho^2}B_t^Y=\Big(1+(-1)^k\,\rho\Big)\left(B_t^X-B_{\tau_k}^X\right)-\sqrt{1-\rho^2}\left(B_t^Y-B_{\tau_k}^Y\right)+\alpha_k. \tag{2.22}$$

As $\tilde{B}_t^k$ does not change after time τ_{k+1}, this relationship remains true for all time greater than τ_k. At time τ_{k+1}, we have the equation

$$\alpha_{k+1}=\Big(1+(-1)^k\,\rho\Big)\left(B_{\tau_{k+1}}^X-B_{\tau_k}^X\right)-\sqrt{1-\rho^2}\left(B_{\tau_{k+1}}^Y-B_{\tau_k}^Y\right)+\alpha_k. \tag{2.23}$$

Taking the difference between Equation (2.22) and Equation (2.23), we have

$$X_t-\rho\tilde{B}_t^k-\sqrt{1-\rho^2}B_t^Y=\Big(1+(-1)^k\,\rho\Big)\left(B_t^X-B_{\tau_{k+1}}^X\right)-\sqrt{1-\rho^2}\left(B_t^Y-B_{\tau_{k+1}}^Y\right)+\alpha_{k+1}.$$

Let $\tau_{k+1}\leq t\leq\tau_{k+2}$. If $k=n$, the proof is over because $Y_t^n=\rho\tilde{B}_t^n+\sqrt{1-\rho^2}B_t^Y$ for τ_{n+1}. Otherwise, $Y_t^n=\rho\tilde{B}_t^{k+1}+\sqrt{1-\rho^2}B_t^Y$ with $\tilde{B}_t^{k+1}=\mathcal{R}\left(\tilde{B}_t^k,\tau_{k+1}\right)=2\tilde{B}_{\tau_{k+1}}^k-\tilde{B}_t^k$ and

$$\begin{aligned}
X_t-Y_t^n&=X_t-\rho\tilde{B}_t^{k+1}-\sqrt{1-\rho^2}B_t^Y\\
&=X_t-\rho\tilde{B}_t^k-\sqrt{1-\rho^2}B_t^Y+\rho(\tilde{B}_t^k-\tilde{B}_t^{k+1})\\
&=X_t-\rho\tilde{B}_t^k-\sqrt{1-\rho^2}B_t^Y+2\rho(\tilde{B}_t^k-\tilde{B}_{\tau_{k+1}}^k)\\
&=\Big(1+(-1)^k\,\rho\Big)\left(B_t^X-B_{\tau_{k+1}}^X\right)-\sqrt{1-\rho^2}\left(B_t^Y-B_{\tau_{k+1}}^Y\right)+\alpha_{k+1}+2\rho\left(\tilde{B}_t^k-\tilde{B}_{\tau_{k+1}}^k\right).
\end{aligned}$$

Let $s,t>\tau_k$, we have

$$\begin{aligned}
\tilde{B}_t^k-\tilde{B}_s^k&=-\tilde{B}_t^{k-1}+2\tilde{B}_{\tau_k}^{k-1}+\tilde{B}_s^{k-1}-2\tilde{B}_{\tau_k}^{k-1}\\
&=-\left(\tilde{B}_t^{k-1}-\tilde{B}_s^{k-1}\right)=(-1)^k\,(\tilde{B}_t^0-\tilde{B}_s^0)\\
&=(-1)^{k+1}\left(B_t^X-B_s^X\right).
\end{aligned}$$

Then $2\rho\left(\tilde{B}_t^k - \tilde{B}_{\tau_{k+1}}^k\right) = 2\rho\,(-1)^{k+1}\left(B_t^X - B_{\tau_{k+1}}^X\right)$ and we find that the property holds at rank $k+1$, which achieves the proof.

(iii) We first need Lemma 2.4.

Lemma 2.4. *We have*

$$\tau_k \overset{d}{=} \inf\{t \geq 0 : B_t = u_k\} \tag{2.24}$$

where

$$\begin{cases} u_0 = 0 \\ u_k = \frac{\eta}{\sqrt{2(1+\rho)}} + \frac{(\eta-\nu)}{\sqrt{2}}\left(\frac{\lfloor \frac{k}{2} \rfloor}{\sqrt{1-\rho}} + \frac{\lfloor \frac{k-1}{2} \rfloor}{\sqrt{1+\rho}}\right) \quad k \geq 1 \end{cases} \tag{2.25}$$

with B a standard Brownian motion and $\lfloor . \rfloor$ the floor function.

Proof. The property holds for $k = 1$. Let us suppose that the property holds for $k = 2p+1$. $X_{\tau_k} - Y_{\tau_k}^n = \eta$ and τ_{k+1} is the first time greater than τ_k when $X_t - Y_t^n$ goes to ν. According to Equation (2.2),

$$\mathbb{P}\left(\tau_{k+1} \leq t\right) = \mathbb{P}\Big(\inf_{\tau_k \leq s \leq t} (1-\rho)\left(B_s^X - B_{\tau_k}^X\right) - \sqrt{1-\rho^2}\left(B_s^Y - B_{\tau_k}^Y\right) + \eta \leq \nu, t \geq \tau_k\Big).$$

If $t \geq \tau_k$, $(B_t^X - B_{\tau_k}^X)$ and $(B_t^Y - B_{\tau_k}^Y)$ are Brownian motions independent of $\mathcal{F}_{\tau_k}$. Then using Lemma 2.1 and Lemma 2.2, we have

$$\begin{aligned} \mathbb{P}\left(\tau_{k+1} \leq t\right) &= \mathbb{E}\Big(2\Phi\Big(\frac{\nu-\eta}{\sqrt{2\,(1-\rho)\,(t-\tau_k)}}\Big)\mathbf{1}_{t \geq \tau_k}\Big) \\ &= 2\mathbb{P}\left((1-\rho)\left(B_t^X - B_{\tau_k}^X\right) - \sqrt{1-\rho^2}\left(B_t^Y - B_{\tau_k}^Y\right) \leq \nu - \eta, t \geq \tau_k\right) \\ &= 2\Phi\Big(\frac{\nu-\eta}{\sqrt{2\,(1-\rho)\,t}} - u_k\Big). \end{aligned}$$

This is the law of the stopping time $\tau = \inf\{t \geq 0 : B_t = u_k + \frac{\eta-\nu}{\sqrt{2(1-\rho)}}\}$ and the property holds for $k+1$. The proof is similar for $k = 2p$. □

The proof of (iii) can be done. $\{N_t = n\} = \{\tau_n \leq t, \tau_{n+1} > t\}$ and then we have

$$\mathbb{E}\left(N_t\right) = \sum_{n=1}^{\infty} n\mathbb{P}\left(\tau_n \leq t, \tau_{n+1} > t\right) \leq \sum_{n=1}^{\infty} n\mathbb{P}\left(\tau_n \leq t\right)$$

According to Lemma 2.4, $\mathbb{P}\left(\tau_n \leq t\right) = 2\int_{\frac{u_n}{\sqrt{t}}}^{\infty} \frac{e^{\frac{-y^2}{2}}}{\sqrt{2\pi}} dy = 2\Phi\left(\frac{-u_n}{\sqrt{t}}\right)$. Since $\lim_{n\to\infty} u_n = \infty$ and $n = \underset{n\to\infty}{O}(u_n)$,

$$\mathbb{P}\left(\tau_n \leq t\right) = \underset{n\to\infty}{o}\left(e^{\frac{-u_n^2}{2t}}\right) = \underset{n\to\infty}{o}\left(\frac{1}{u_n^3}\right) = \underset{n\to\infty}{O}\left(\frac{1}{n^3}\right).$$

Then $n\mathbb{P}\left(\tau_n \leq t\right) = \underset{n\to\infty}{O}\left(\frac{1}{n^2}\right)$ and $\mathbb{E}\left(N_t\right) < \infty$ by comparison theorem of positive series, implying $N_t < \infty$ almost surely.

(iv) If $n \geq N_t$, the number of reflections of $X - Y^n$ between time 0 and time t is equal to N_t and $Y_t^{N_t} = Y_t^n =$ almost surely.

5.4 Proof of Proposition 2.4 and Corollary 2.1

We start with Lemma 2.5.

Lemma 2.5. *For $t > 0$, $x \in \mathbb{R}$,*

$$\mathbb{P}\left(X_t - Y_t^n \leq x, t \geq \tau_n\right) = \begin{cases} \Phi\Big(\frac{x-\alpha_n}{\sqrt{2(1+(-1)^n\rho)t}} - \frac{u_n}{\sqrt{t}}\Big) & \text{if} \quad x < \alpha_n \\ \Phi\Big(\frac{x-\alpha_n}{\sqrt{2(1+(-1)^n\rho)t}} + \frac{u_n}{\sqrt{t}}\Big) - 2\Phi\left(\frac{u_n}{\sqrt{t}}\right) + 1 & \text{if} \quad x \geq \alpha_n \end{cases}$$

and

$$\mathbb{P}\left(X_t - Y_t^n \leq x, t \geq \tau_{n+1}\right) = \begin{cases} \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2(1+(-1)^n\rho)t}} - \frac{u_{n+1}}{\sqrt{t}}\Big) & \text{if} \quad x < \alpha_{n+1} \\ \Phi\Big(\frac{x-\alpha_{n+1}}{\sqrt{2(1+(-1)^n\rho)t}} + \frac{u_{n+1}}{\sqrt{t}}\Big) - 2\Phi\Big(\frac{u_{n+1}}{\sqrt{t}}\Big) + 1 & \text{if} \quad x \geq \alpha_{n+1} \end{cases}.$$

Proof We have $\mathbb{P}\left(X_t - Y_t^n \leq x, t \geq \tau_n\right)$ equal to

$$\begin{aligned} &\mathbb{P}\Big((1+(-1)^n\rho)\left(B_t^X - B_{\tau_n}^X\right) - \sqrt{1-\rho^2}\left(B_t^Y - B_{\tau_n}^Y\right) + \alpha_n \leq x, t \geq \tau_n\Big) \\ &= \mathbb{E}\Big(\Phi\Big(\frac{x-\alpha_n}{\sqrt{2\left(1+(-1)^n\rho\right)(t-\tau_n)}}\Big)\mathbf{1}_{t\geq\tau_n}\Big). \end{aligned}$$

However, according to Equation (2.24), $\tau_n \sim \tau' = \inf\{t \geq 0 : B_t = u_n\}$ where B_t is a standard Brownian motion. Then we have, using Lemma 2.2,

$$\begin{aligned} \mathbb{P}\left(X_t - Y_t^n \leq x, t \geq \tau_n\right) &= \mathbb{E}\Big(\Phi\Big(\frac{x-\alpha_n}{\sqrt{2\left(1+(-1)^n\rho\right)(t-\tau')}}\Big)\mathbf{1}_{t\geq\tau'}\Big) \\ &= \Phi\Big(\frac{x-\alpha_n}{\sqrt{2\left(1+(-1)^n\rho\right)t}} - \frac{u_n}{\sqrt{t}}\Big)\mathbf{1}_{x<\alpha_n} \\ &+ \Big(\Phi\Big(\frac{x-\alpha_n}{\sqrt{2\left(1+(-1)^n\rho\right)t}} - \frac{u_n}{\sqrt{t}}\Big) - 2\Phi\left(\frac{u_n}{\sqrt{t}}\right) + 1\Big)\mathbf{1}_{x\geq\alpha_n}. \end{aligned}$$

The proof is the same for $\mathbb{P}\left(X_t - Y_t^n \leq x, t \geq \tau_{n+1}\right)$.

□

We can now prove Proposition 2.4. We have:

$$\mathbb{P}\left(X_t - Y_t^{n+1} \geq x\right) - \mathbb{P}\left(X_t - Y_t^n \geq x\right) = \mathbb{P}\left(X_t - Y_t^n \leq x\right) - \mathbb{P}\left(X_t - Y_t^{n+1} \leq x\right)$$

which is equal to

$$\begin{aligned} &\mathbb{P}\left(X_t - Y_t^n \leq x, \tau_{n+1} \leq t\right) - \mathbb{P}\left(X_t - Y_t^{n+1} \leq x, \tau_{n+1} \leq t\right) \\ &+ \mathbb{P}\left(X_t - Y_t^n \leq x, \tau_{n+1} \geq t\right) - \mathbb{P}\left(X_t - Y_t^{n+1} \leq x, \tau_{n+1} \leq t\right). \end{aligned}$$

For $\tau_{n+1} \geq t$, $X_t - Y_t^n$ and $X_t - Y_t^{n+1}$ are equals then

$$\mathbb{P}\left(X_t - Y_t^n \leq x, \tau_{n+1} \geq t\right) = \mathbb{P}\left(X_t - Y_t^{n+1} \leq x, \tau_{n+1} \geq t\right).$$

We then have

$$\mathbb{P}\left(X_t - Y_t^{n+1} \geq x\right) - \mathbb{P}\left(X_t - Y_t^n \geq x\right) = \mathbb{P}\left(X_t - Y_t^n \leq x, \tau_{n+1} \leq t\right) - \mathbb{P}\left(X_t - Y_t^{n+1} \leq x, \tau_{n+1} \leq t\right)$$

and we can conclude using Lemma 2.5.

Since for $n \geq N_t$ $X_t - Y_t^n = X_t - Y_t^{N_t}$, $X_t - Y_t^{N_t}$ is the limit in law of $X_t - Y_t^n$, and $\mathbb{P}\left(X_t - Y_t \geq x\right) = \lim_{n\to\infty} \mathbb{P}\left(X_t - Y_t^n \geq x\right)$.

The proof for Corollary 2.1 follows.

Let $x \in [\nu, \eta]$ and let assume $\rho > 0$. We have:

$$\mathbb{P}\left(X_t - Y_t^{n+1} \geq x\right) - \mathbb{P}\left(X_t - Y_t^n \geq x\right) = p_{n+1}\left(t, x\right).$$

If n is even,

$$p_{n+1}\left(t, x\right) = \Phi\Big(\frac{x-\eta}{\sqrt{2\left(1+\rho\right)t}} - \frac{u_{n+1}}{\sqrt{t}}\Big) - \Phi\Big(\frac{x-\eta}{\sqrt{2\left(1-\rho\right)t}} - \frac{u_{n+1}}{\sqrt{t}}\Big) > 0.$$

If n is odd,

$$p_{n+1}\left(t, x\right) = \Phi\Big(\frac{x-\nu}{\sqrt{2\left(1-\rho\right)t}} + \frac{u_{n+1}}{\sqrt{t}}\Big) - \Phi\Big(\frac{x-\nu}{\sqrt{2\left(1+\rho\right)t}} + \frac{u_{n+1}}{\sqrt{t}}\Big) > 0,$$

which achieves the proof.

5.5 Proof of Proposition 2.5

As $\tilde{\rho}$ is Lipschitz and $\sup_{x\in\mathbb{R}} |\tilde{\rho}\left(x\right)| < 1$, $\sqrt{1-\tilde{\rho}^2}$ is Lipschitz on $\mathbb{R}$ and $(x, y) \mapsto \begin{pmatrix} 1 & 0 \\ \tilde{\rho}\left(x-y\right) & \sqrt{1-\tilde{\rho}\left(x-y\right)^2} \end{pmatrix}$ is Lipschitz on $\mathbb{R}^2$, which is a sufficient condition for the system to have a strong solution. This solution is Markovian.

X is clearly a Brownian motion. By the Lévy characterization of the Brownian motion, Y is also a Brownian motion.

Bibliography

[1] René Aid, Luciano Campi, Adrien Nguyen Huu, and Nizar Touzi. A structural risk-neutral model of electricity prices. *International Journal of Theoretical and Applied Finance*, 12(07):925–947, 2009.

[2] René Aid, Luciano Campi, and Nicolas Langrené. A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*, 23(3):387–438, 2013.

[3] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

[4] Damien Bosc. *Three essays on modeling the dependence between financial assets*. PhD thesis, Ecole Polytechnique X, 2012.

[5] René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer, 2014.

[6] René Carmona and Valdo Durrleman. Pricing and hedging spread options. *Siam Review*, 45(4):627–685, 2003.

[7] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004.

[8] Thomas Deschatre. On the control of the difference between two brownian motions: a dynamic copula approach. *Dependence Modeling*, 4:141–160, 2016.

[9] Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[10] Olivier Féron and Elias Daboussi. *Commodities, Energy and Environmental Finance*, chapter Calibration of electricity price models, pages 183–207. Springer, 2015.

[11] Christian Gouriéroux, Joann Jasiak, and Razvan Sufana. The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181, 2009.

[12] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105, 1992.

[13] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[14] Piotr Jaworski and Marcin Krzywda. Coupling of wiener processes by using copulas. *Statistics & Probability Letters*, 83(9):2027–2033, 2013.

[15] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical methods for financial markets*. Springer, 2009.

[16] Alex Langnau. A dynamic model for correlation. *Risk*, 23(4):74, 2010.

[17] Katsushi Nakajima and Kazuhiko Ohashi. A cointegrated commodity pricing model. *Journal of Futures markets*, 32(11):995–1033, 2012.

[18] Paul A Samuelson. Proof that properly anticipated prices fluctuate randomly. *IMR; Industrial Management Review (pre-1986)*, 6(2):41, 1965.

[19] M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[20] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.

Chapter 3

Estimation of a fast mean reverting jump process with application to spike modeling in electricity prices

Abstract

Let us consider the model $X_t = X_t^c + Z_t^\beta$ where X_t^c is a continuous semimartingale and $Z_t^\beta = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}(dt, dx)$ is a mean reverting compound Poisson process. From discrete observations $X_{\frac{i}{n}}$, $i = 0, ..., nT$ with T fixed, an estimation procedure of the parameter β is proposed when $\Delta_n = \frac{T}{n} \to 0$ but also when $\beta = \beta_n \to \infty$, allowing for the identification of β which is a drift parameter. The asymptotic on β allows to model spikes in time series, that is a jump process reverting to 0 very fast. A method for detection of the jumps is also proposed in this new framework. We apply these results on the modeling of spikes in electricity spot prices. A model including spikes is derived and includes both forward and spot modeling ; this model can generalize most of the continuous models of the literature in a simply way. The estimation procedure is applied on French, German and Australian spot prices data.

1 Introduction

1.1 Motivation

A striking empirical feature of electricity spot prices is the presence of spikes, that can be described by a jump in the price process immediately followed by a fast mean reversion, see Figure 3.1 below that shows in particular the behaviour of the French and German EPEX electricity spot markets over one year of historical data. These spikes can be related to abnormal temperature oscillations, failures of power plant or any other dramatic physical event. The modeling of such exogenous events has some importance, especially for risk management. Since electricity is not a storable commodity, hedging options with spot electricity is meaningless, and forward contracts are to be used instead. If $(S_t)_{t\geq 0}$ denotes the spot price in a given electricity market, the value of the forward at time t with delivery period T can be defined as

$$f(t,T) = \mathbb{E}\big[S_T \,\big|\, \mathcal{F}_t\big], \quad t \geq 0 \tag{3.1}$$

where $\mathcal{F}_t$ is the available information up to time t. The quantity $f(t,T)$ specifies the price at time t of the contract that delivers one unit of electricity during 1 hour (in MWh) starting in T. Obtaining tractable formulas for $f\,(t,T)$ from a given stochastic model for the process $(S_t)_{t\geq 0}$ is thus of major importance, especially in the presence of fast mean-reverting spikes, as observed empirically. In this context, a popular and tractable model, developed for instance in Lucia *et al.* [18], consists in modeling $(S_t)_{t\geq 0}$ as

$$S_t = \rho(t) + Y_t, \quad dY_t = -\beta Y_t dt + \sigma dW_t, \quad t \geq 0$$

where $\rho(t)$ is a deterministic component and $(Y_t)_{t\geq 0}$ is an Ornstein-Uhlenbeck process driven by a Brownian motion $(\sigma W_t)_{t\geq 0}$ with variance σ^2 and mean-reversion parameter $\beta > 0$. This tractable model however leaves out the feature of spikes. Cartea and Figueroa [8] extend this approach by adding a jump component to the Ornstein-Uhlenbeck part (replacing S_t by $\log S_t$), resulting in

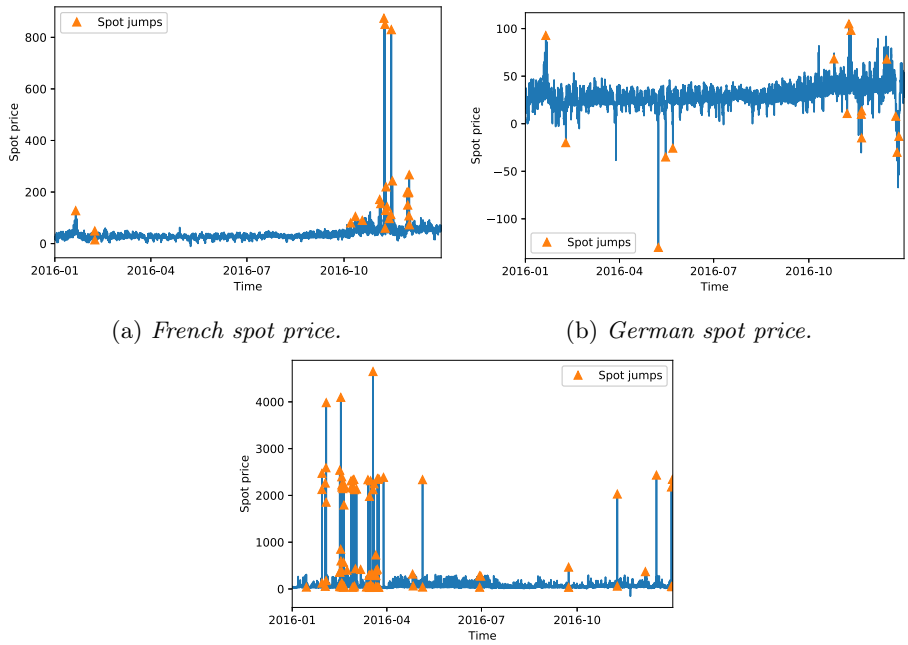
$$\log S_t = \rho(t) + Y_t, \quad dY_t = -\beta Y_t dt + \sigma(t) dW_t + \log J dN_t, \quad t \geq 0,$$

where $(N_t)_{t\geq 0}$ is a Poisson process and J is the jump size drawn proportional to a log-normal distribution. A similar model is proposed in Geman and Roncoroni [12], adding up a threshold parameter that determines the sign of the jumps.

In these two extensions, the mean reverting coefficient $\beta > 0$ is the same for the continuous component and for the spike component. However, statistical evidence shows that the mean reversion of the spike component is much stronger than the one of the Brownian component, see for instance Benth *et al.* [6]. The estimated parameter for the mean reversion underestimates the one of the spike mean reversion and overestimates the one of the continuous mean reversion, which can have dramatic consequences for the calibration of the forward prices $f(t,T)$ from historical data. Furthermore, the model of Geman and Roncoroni [12] does not provide with explicit formulae for deriving $f(t,T)$. Yet another approach is undertaken in Benth *et al.* [5; 6] with multi-factor models. The dynamic of the spot now takes the form

$$S_t = \sum_{i=1}^{m} w_i Y_t^i, \quad dY_t^i = -\beta_i Y_t^i dt + dL_t^i, \quad t \geq 0, \quad i = 1, \ldots, m,$$

for some weights w_i, and where $(L_t^i)_{t\geq 0}$ are independent time-inhomogeneous subordinators that ensure that $(S_t)_{t\geq 0}$ remains nonnegative. Benth and collaborators establish in [5; 6] that $m = 2$



(a) *French spot price.*

(b) *German spot price.*

(c) *Australian spot price.*

Figure 3.1: *Series of the spot price during the year 2016 for France, Germany and Australia. The frequency of the data is 1 hour for France and Germany and 30 minutes for Australia. Spot jumps are estimated using a threshold of* $5\hat{\sigma}\Delta_n^{0.49}$ *where* $\hat{\sigma}$ *is the multi-power variation of order 20.*

is sufficient for modeling purposes, each factor $(Y_t^i)_{t\geq 0}$ having its own mean reverting parameter, allowing for a fast mean reversion and a slower one in the case of two factors. However, the use of subordinators implies that the volatility of the process seems to be underestimated. Finally, multi-factor models with a Brownian component and a spike component are studied in Meyer and Tankov [21], Schmidt [25] and Gonzales *et al.* [13]. Meyer and Tankov estimate the mean-reverting parameters using spectral methods and the jumps are detected by filtering. In Schmidt [25], the parameters of the model are estimated using maximum likelihood with the EM algorithm, implying an approximation of the process with its Euler scheme. Gonzales and co-authors develope a Bayesian framework and recover the parameters of the model by MCMC. In a more general context than electricity price modeling , Moreno et al. [22] use a method of moments to estimate the parameter of a jump diffusion model when the log-price is the sum of an arithmetic Brownian motion and a mean reverting compound Poisson process.

The goal of the paper is threefold:

i) Introduce a spot electricity price model consistent with historical data that encompasses the previous approaches and overcomes their limitations.

ii) Develop within this model efficient and robust statistical procedures that estimates the characteristics of the presence of spikes.

iii) Derive an explicit correction formula for the value $f(t,T)$ of forward contracts (3.1) revealing the effect of spikes that can be implemented thanks to ii).

1.2 Main results

We consider an extended framework that encompasses [21], [25] and [13]. In particular, our approach does not require that the continuous part of the price process is an Ornstein-Uhlenbeck, a necessary condition in the aforementioned models.

A semimartingale models with fast mean-reverting jumps

On a rich enough filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ that will accommodate all the considered random quantities, we model the electricity spot price $X_t = S_t$ or $\log S_t$ (depending on the choice of modeling) by

$$X_t = X_t^c + Z_t^\beta, \quad t \geq 0, \tag{3.2}$$

where $(X_t^c)_{t\geq_0}$ is a continuous Itô semimartingale and $(Z_t^\beta)_{t\geq 0}$ is what we call the *spike process*, governed by a mean-reverting factor $\beta > 0$. More specifically, we assume that

$$X_t^c = X_0^c + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \quad t \geq 0 \tag{3.3}$$

where $(\sigma_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ are two adapted càdlàg processes, $(W_t)_{t\geq 0}$ a $(\mathcal{F}_t)$-standard Brownian motion and

$$Z_t^\beta = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}\,(dt, dx)\,, \quad t \geq 0, \tag{3.4}$$

with $\underline{p}$ a random Poisson measure on $[0,\infty) \times \mathbb{R}$ independent of $(W_t)_{t\geq 0}$, with intensity

$$\underline{q} = \lambda\, dt \otimes \nu(dx),$$

for some $\lambda > 0$ and a probability measure $\nu(dx)$ on $\mathbb{R}$. We thus model the electricity spot price as a classical continuous Itô semimartingale $(X_t^c)_{t\geq 0}$ allowing for the usual financial fluctuations and usual models (factor models, mean-reverting models and so on) to which we add an perturbation $(Z_t^\beta)_{t\geq 0}$ of "spikes" or "jumps", triggered by exogeneous physical hazard, at intensity λ and magnitude $\nu(dx)$, but with a relaxation period $1/\beta$ comparable to λ that accounts for the absorption of such events by the market toward resulting in stable prices at large scales. The term *comparable* is a bit vague at this stage, and will be assessed precisely in Section 2.1 below, enabling us to speak of *fast mean-reversion*. In this setting, model (3.2)-(3.3)-(3.4) is well posed and can reproduce, at least *visually*, the general shape of electricity spot markets, compare historical data from Figure 3.1 and sample paths simulations given in Figure 3.2 and detailed in the simulation Section 3.2.

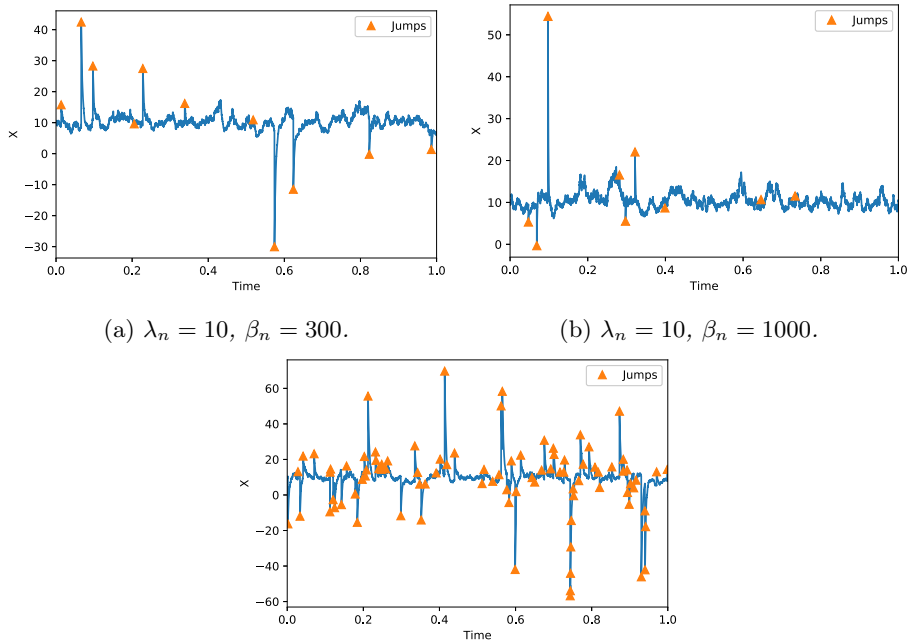
Statistical setting

We assume that we observe the process $(X_t)_{t\geq 0}$ given by (3.2)-(3.3)-(3.4) over the time interval $[0,T]$ on a regular grid

$$0 = t_{0,n} < t_1 < \ldots, t_n, n = T,\ t_{i,n} = i\Delta_n, \ \text{ for } \ 0 \leq i \leq n,$$

with mesh Δ_n. Thus we have n (or rather $n+1$) observations

$$X^n = (X_0, X_{\Delta_n}, \ldots, X_{n\Delta_n} = X_T).$$



(a) $\lambda_n = 10$, $\beta_n = 300$.

(b) $\lambda_n = 10$, $\beta_n = 1000$.

(c) $\lambda_n = 75$, $\beta_n = 300$.

Figure 3.2: *Simulation of the process X in the case of a model having continuous part defined in* (3.11) *and with jump sizes having law* $0.4\left(-\mathcal{E}\left(40\right)\right) + 0.6\mathcal{E}\left(30\right)$ *for different values of* λ_n *and* β_n.

Asymptotics are taken as $n \to \infty$. We assume that T is constant, and we take $T = 1$ with no loss of generality. Equivalently, $\Delta_n = 1/n \to 0$ as $n \to \infty$. This asymptotic setting is usually referred to as the "high-frequency" framework (for instance the classical textbook [3] by Aït-Sahalia and Jacod), but this terminology is a bit misleading: our framework certainly belongs to statistical finance, but it has no link to high-frequency finance or microstructure modeling of any sort. In practice, we apply our methodology to three markets: the French EPEX, the German EPEX and the Australian electricity spot in Queensland, see Section 3 below. We use data between 2015, Jan. 01 and 2016, Dec. 31. with hourly data (even less in the case of Australian data), so that $n = 17064$ is considered to be large. Equivalently, 30 minutes is considered to be small in front of 2 years. In our setting, the important fact about the assumption that T is fixed is that we leave out any stationarity or ergodicity of the underlying process. We thus make an implicit statistical robustness assumption, which we believe is of importance when considering recent energy markets over such time horizons.

The parameters of interest are $\lambda, \beta > 0$ that govern our correction formulas see the application to forward contracts prices $f(t, T)$ below. In particular we leave out the issue of identifying the continuous semimartingale part $(X_t^c)_{t\geq 0}$ *i.e.* the drift $(\mu_t)_{t\geq 0}$ and the volatility process $(\sigma_t)_{t\geq 0}$ as well as the jump distribution $\nu(dx)$.

The mean-reversion factor over the observation increment $[t_{i-1,n}, t_{i,n}]$ is of size $\beta\Delta_n$, and by requir-

ing $\beta\Delta_n$ to be large compared to the order of magnitude $\Delta_n^{1/2}$ of $X^c_{t_{i,n}} - X^c_{t_{i-1,n}}$, we may hope to recover β asymptotically. We thus introduce the asymptotic setting $\beta = \beta_n$ with the requirement

$$\beta = \beta_n \to \infty \text{ while } \beta_n\sqrt{\Delta_n} \to \infty \tag{3.5}$$

Condition (3.5) becomes

$$\beta_n\sqrt{\Delta_n\lambda_n} \to \infty \tag{3.6}$$

if we let $\lambda = \lambda_n \to \infty$, another necessary condition that enough jumps are available over the observation period $[0, T]$. A second crucial assumption is

$$\beta_n\Delta_n \lesssim 1, \tag{3.7}$$

since otherwise, the spikes caused by the jumps of $\underline{p}$ are absorbed by the Brownian fluctuations of X^c due to the fast relaxation period $1/\beta_n$ and therefore cannot be detected by X^n.

Statistical results

Heavily relying on classical techniques in high-frequency finance (for instance [3, Theorem 10.26, p.374]), we estimate in a first step the times and sizes of the jumps which are random quantities, allowing λ_n and β_n to be either bounded or to grow to ∞ with n, see Proposition 3.1. In a second step, we construct an estimator of β_n taking into account the interplay between β_n, λ_n and Δ_n dictated by the asymptotic regime (3.5)-(3.6)-(3.7). This estimator is based on an estimator $\widehat{s}_n$ of the right-derivative or *instantaneous slope* of $t \mapsto Z_t^\beta$ right after a jump is detected . The estimator $\widehat{s}_n$ is based on averaging of instantaneous slope proxies of the form $(X_{T_q} - X_{T_q^-})(1 - e^{-\beta_n\Delta_n})$ that govern the relaxation effect after a jump of size $X_{T_q} - X_{T_q^-}$ has occured at time T_q and it enables us to consider

$$\widehat{\beta}_n = -\frac{1}{\Delta_n}\log\left(1 - \widehat{s}_n\right)$$

as our estimator of β_n. Since β_n itself varies with n and grows to infinity, the notion of convergence has to be considered carefully. Under suitable assumptions, we prove in Theorem 3.1 that the relative error

$$\mathcal{E}_n = \frac{\widehat{\beta}_n - \beta_n}{\beta_n} \to 0 \tag{3.8}$$

in probability as $n \to \infty$. The error $\mathcal{E}_n$ has two components: a first term of order $1/(\beta_n\sqrt{\lambda_n\Delta_n})$ due to Brownian oscillations, and a second term of order $\min\{\lambda_n/\beta_n, 1/\sqrt{\beta_n}\} + \sqrt{\lambda_n}/\beta_n$ that accounts to the effect of jumps that are still present in the price process despite the relaxation effect. When $\beta_n\sqrt{\Delta_n\lambda_n} \to \infty$ and $\lambda_n/\beta_n \lesssim 1$, we have $\mathcal{E}_n$ converges to 0. If we assume further $\sqrt{\beta_n}/\lambda_n \to 0$, we obtain a central limit theorem for $\mathcal{E}_n$ with a Gaussian limit and an explicit rate of convergence that depends on the interplay between λ_n, β_n and Δ_n. We have an analogous result (although less demanding) for the estimation of the jump intensity λ_n detailed in Proposition 3.2.

Application to pricing forward contracts

We show in Theorem 3.2 that in the model defined by (3.2)-(3.3)-(3.4), the value of a forward contract at time t and delivery period T (over one hour) is given by

$$f(t,T) = f^c(t,T) + f^\beta(t,T),$$

where

$$f^c(t,T) = \mathbb{E}\left[X_T^c \mid \mathcal{F}_t\right]$$

and

$$f^\beta(t,T) = e^{-\beta(T-t)} Z_t^\beta + \frac{\lambda}{\beta} \int_{\mathbb{R}} x\nu(dx) \left(1 - e^{-\beta(T-t)}\right).$$

The term $f^c(t,T)$ corresponds to the price of the forward contract in a continuous case framework. The computation of this value has been extensively studied for different continuous models and it is known analytically for the most common models, see for instance [5; 6] among others. The term $f^\beta(t,T)$ is a correction that follows from our approach. It is of order λ/β and is usually small for the applications we have in mind, see the practical implementation Section 3 and 4. On balance, the presence of spikes does not impact significantly the price of forward contracts to within these order of magnitudes. This is consistent with our data, for which spikes are not observed on forward prices. By neglecting the term $f^\beta(t,T)$, we can calibrate the process X_t^c on $f(t,T)$. Our model can easily be used for risk management: we implement the prices of the forward with historical data on electricity prices in Section 4 and we show the impact of the spike modeling on the valuation of strip options with payoff of the form $\sum_{i=1}^p (S_{t_i} - K)^+$ for different times t_i. As expected, the value of this option increases if we add significant large spikes. In some particular, we show that some out-of-the-money options have a significant value if we incorporate spikes in our modeling, while having a value close to 0 otherwise.

1.3 Organisation of the paper

Section 2 develops a rigorous mathematical framework for the stochastic model (3.2)-(3.3)-(3.4) and gives the explicit construction of the estimators described in Section 1.2 above together with their asymptotic properties in Proposition 3.1, 3.2 and Theorem 3.1. Section 3 establishes the numerical feasibility and consistency of our statistical estimation results on simulated and real data, based over two years (2015 and 2016) of hourly sampled data over three different markets (French, German and Australian). Section 4 is devoted to the application of our model and statistical results to forward contracts. We establish in Theorem 3.2 a correction formula that takes into account the modeling of spikes and discuss its consistency with daily forward prices on the French market with historical data over two years (2015 and 2016). The proofs are given in Section 5.

2 Statistical results

2.1 Model assumptions

We consider the process $(X_t)_{t\geq 0}$ defined by (3.2)-(3.3)-(3.4) in Section 1.2. Following closely the standard notation of Aït-Sahalia and Jacod [3] we write $f(x) \star \nu$ for the integral $\int_{\mathbb{R}} f(x)\,\rho(dx)$ of a real-valued function f with respect to the positive measure ν on $\mathbb{R}$. Remember that we work over a finite time horizon $T = 1$.

Assumption 3.1. *We have $\mathbb{E}[(X_t^c)^2] < \infty$ for every $t \geq 0$. Moreover, $t \mapsto \sigma_t$ is continuous on $[0,1]$ and $\inf_t \sigma_t > 0$ and for some deterministic $\underline{\sigma}, \bar{\sigma}, c_0 > 0$, we have $0 < \underline{\sigma}^2 \leq \inf_t \sigma_t^2 \leq \sup_t \sigma_t^2 \leq \bar{\sigma}^2$ and $\sup_t |b_t| \leq c_0$. We also add the assumption that the process $\int_0^t \sigma_s dW_s$ is independent from Z^β. Finally the jump probability measure $\nu(dx)$ satisfies $\nu(\{0\}) = 0$ and $|x|^2 \star \nu < \infty$.*

Since our asymptotic results will be given in distribution (see Theorem 3.1 below), the conditions on the drift $(b_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$ can substantially be weakened (in order to accomodate for instance diffusion coefficients of the form $\sigma_t = X_t^c h(X_t^c)$ with a bounded h or even locally integrable) by standard localisation procedures, see for instance [14, Section 4.4.1].
We observe

$$X^n = (X_0, X_{\Delta_n}, \ldots, X_{n\Delta_n} = X_1)$$

and asymptotics are taken as $n \to \infty$ or equivalently $\Delta_n = n^{-1} \to 0$. We also allow $\beta = \beta_n$ and $\lambda = \lambda_n$ to either grow to ∞ with n or remain bounded.

Assumption 3.2. *We have*

$$\lambda_n \lesssim \beta_n, \quad \beta_n \Delta_n \lesssim 1, \quad \textit{and} \quad \lambda_n \Delta_n \to 0.$$

The condition $\lambda_n \lesssim \beta_n$ ensures the stability of X_t as $n \to \infty$ since $\mathrm{Var}(X_t) = \mathrm{Var}\,(X_t^c) + \mathrm{Var}(Z_t^{\beta_n}) = \mathrm{Var}\,(X_t^c) + |x|^2 \star \nu \frac{\lambda_n t}{2\beta_n}\left(1 - e^{-2\beta_n}\right) \to \infty$ if $\sup_n \lambda_n/\beta_n = \infty$. The condition $\beta_n \Delta_n \lesssim 1$ is necessary to identify spikes (or jumps): otherwise, a spike that occurs in the interval $((i-1)\,\Delta_n, i\Delta_n]$ will be absorbed by the relaxation effect before we observe $X_{i\Delta_n}$. Finally, the condition $\lambda_n \Delta_n \to 0$ controls the no accumulation of jumps within the rate of observation.

In order to estimate the time and sizes of the jumps, we need the following:

Assumption 3.3. *We have either* (I) *or* (II)*, where*

(I) For some $\varpi \in (0, 1/2)$*:*

$$\text{(i) } \lambda_n^2 \Delta_n \to 0, \quad \text{(ii) } \lambda_n \mathbf{1}_{\{|x| > \Delta_n^{1/2-\varpi}/(\beta_n \Delta_n)\}} \star \nu \to 0, \quad \textit{and} \quad \text{(iii) } \lambda_n \mathbf{1}_{\{|x| < \Delta_n^{1/2-\varpi}\}} \star \nu \to 0.$$

(II) For some $\varpi \in (0, 1/2)$ *and a sequence of integers* $k_n \geq 1$*:*

$$\text{(i) } \lambda_n^2 \Delta_n k_n^2 \to 0, \quad \text{(ii) } \lambda_n \mathbf{1}_{\{|x| > e^{\beta_n \Delta_n k_n} \Delta_n^{1/2-\varpi}\}} \star \nu \to 0, \quad \text{(iii) } \lambda_n \mathbf{1}_{\{|x| < \Delta_n^{\frac{1}{2}-\varpi}\}} \star \nu \to 0$$

and (iv) $\lambda_n e^{-(\beta_n \Delta_n^{1-\varpi})^2} \to 0$.

Assumption 3.3 (I) implies $\beta_n \Delta_n = o\,(1)$. Condition (i) ensures that the number of jumps in a interval of size Δ_n is essentially 1. In the case where β_n is bounded, (ii) is implied by (i) and we have the usual conditions for the detection of jumps (see Mancini [19]). Condition (ii) controls the size of the mean-reversion. Condition (iii) controls the size of the small jumps: $\left(\Delta X_{T_q}\right)_{q \in \{1,\ldots,N_1\}}$ that cannot converge too fast to 0 ; if $|\Delta X_{T_q}| > C$ If the jumps are bounded below by some constant as Mancini, [19], the condition is automatically satisfied.

Assumption 3.3 (II) (iv) implies that $\beta_n \Delta_n^{1-\varpi} \to \infty$ and in particular $\beta_n \Delta_n^{1/2} \to \infty$, implying that the mean reversion of order $\beta_n \Delta_n$ is stronger than the order of magnitude $\Delta_n^{1/2}$ of Brownian increments. It also allows for the case $\beta_n \Delta_n \approx 1$. In the setting of Assumption 3.3 (II), the mean reversion is more difficult to distinguish from the jumps and in the case $\beta_n \Delta_n \approx 1$, the jumps and the drift have the same size and are not distinguishable from their size solely. Condition (i) states that there is at most one jump in an interval of size k_n which is large enough for the spike to vanish. Assumption 3.3 (II) also implies that $\lambda_n^2/\beta_n \to 0$. Assumption 3.3 (II) allows for high values of β_n but the number of jumps needs then to be smaller than in case Assumption 3.3 (II) compared to β_n.

2.2 Estimation of the jumps times and λ_n

We first construct estimators of the sequence of the jumps and of their intensity λ_n. Assumption 3.3 is in force. Let

$$N_t = \sum_{s \leq t} \mathbf{1}_{\Delta X_s \neq 0}, \quad \Delta X_s = X_s - X_{s^-}, \quad t \geq 0,$$

denote the number of jumps of $(X_t)_{t\geq 0}$ up to time t and let $T_1 < T_2 < \ldots, T_q < \ldots$ denote the random times at which jumps occur. By construction, the sizes of jumps $\left(\Delta X_{T_q}\right)_{q\geq 1}$ form a sequence of IID random variables independent of $(N_t)_{t\geq 0}$. Let $i\,(n,q)$ be the random integer such that

$$(i\,(n,q) - 1)\,\Delta_n < T_q \leq i\,(n,q)\,\Delta_n$$

Define the increasing sequence

$$\mathcal{I}_n\,(1) < \ldots < \mathcal{I}_n(\widehat{\lambda}_n) \subset \{1, \ldots, n\}$$

of indices $i \in \{1, \ldots, n\}$ defined by the realisation of the following successive events:

$$\left\{ \frac{|\Delta_i^n X|}{\sqrt{\Delta_n}} > \Delta_n^{-\varpi} \right\} \text{ under Assumption 3.3 (I)}$$

and

$$\left\{ \frac{|\Delta_i^n X|}{\sqrt{\Delta_n}} > \Delta_n^{-\varpi},\ \Delta_i^n X \Delta_{i+1}^n X < 0 \right\} \text{ under Assumption 3.3 (II).}$$

Finally, define

$$\widehat{T}(n,q) = \mathcal{I}_n(q)\Delta_n \quad q = 1, \ldots, \widehat{\lambda}_n.$$

Under Assumption 3.3 (II), we need the supplementary condition $\Delta_i^n X \Delta_{i+1}^n X < 0$ for the following reason: whenever a jump occurs, the mean reverting is dominant in the next observation interval and has a direction opposite to the sign of the jump. Furthermore, it enables us to discard the increments caused by the mean reversion that are large enough to be detected as jumps. Indeed, if we detect a false jump due to the mean reversion effect, the next increments will follow the same dynamics and it will share the same sign with first increment. The proprerty that there no jump lies within the next observation interval is ensured by the existence of k_n. Let

$$\Omega_n = \left\{ \widehat{\lambda}_n = N_1, \forall q \in \{1, ..., N_1\} : T_q \in \left(\widehat{T}(n,q) - \Delta_n, \widehat{T}(n,q)\right] \right\}.$$

Proposition 3.1. *Work under Assumption 3.1, 3.2 and 3.3. We have* $\mathbb{P}\left(\Omega_n\right) \to 1$.

The proof of Proposition 3.1 relies on a result of Aït-Sahalia and Jacod [3, Theorem 10.26, p.374]. However, the presence of a drift term $-\beta_n \int_0^t Z_s^{\beta_n} ds$ that depends on n together with the fact that $\lambda_n \to \infty$ makes the extension not completely trivial. Proposition 3.1 also provides us with an estimator $\widehat{\lambda}_n$ of λ_n. In the case $\lambda_n \to \infty$, we have the following asymptotic property:

Proposition 3.2. *Work under Assumption 3.1, 3.2 and 3.3 and assume that* $\lambda_n \to \infty$. *We have*

$$\sqrt{\lambda_n} \frac{\widehat{\lambda}_n - \lambda_n}{\lambda_n} \to \mathcal{N}(0,1) \tag{3.9}$$

in distribution as $n \to \infty$.

This result is asymptotically optimal : consider indeed the seemingly richer experiment where one continuously observes a Poisson process $(\mathcal{P}_t)_{0\leq t\leq 1}$ with intensity $\lambda > 0$. The variable $\mathcal{P}_1$ is a sufficient statistic and the Cramer-Rao bound tells us that any unbiased estimator $\widehat{\lambda}$ necessarily satisfies $\mathbb{E}[(\widehat{\lambda}-\lambda)^2] \geq I(\lambda)^{-1}$, where $I(\lambda) = 1+\lambda^{-1}$ is the Fisher information associated to the observation of $\mathcal{P}_1$, *i.e.* a Poisson random variable with parameter λ. Equivalently $\mathbb{E}\big[\big(\frac{\widehat{\lambda}-\lambda}{\lambda}\big)^2\big] \geq (\lambda^{-1}+\lambda^{-2}) \sim \lambda^{-1}$ as $\lambda \to \infty$ which is consistent with the convergence (3.9).

A natural estimator of the jump sizes is $\Delta^n_{\mathcal{I}_n(q)}X$ for $q \in \{1, ..., \widehat{\lambda}_n\}$, see [3, Theorem 10.21, p.370]. In our case, $\Delta^n_{\mathcal{I}_n(q)}X$ is equal to $\Delta X_{T_q} e^{-\beta_n(T_q-\mathcal{I}_n(q)\Delta_n)}$ plus an negligible term. If $\beta_n\Delta_n \to 0$, $\Delta^n_{\mathcal{I}_n(q)}X$ is then equivalent to ΔX_{T_q} but if $\beta_n\Delta_n \asymp 1$, the bias $\Delta X_{T_q}\left(1-e^{-\beta_n(T_q-\mathcal{I}_n(q)\Delta_n)}\right)$ remains and it is not possible to identify the size of the jump. However, if $\lambda_n \to \infty$, one can infer some statistical properties of the jumps size. Indeed, $(T_q - \mathcal{I}_n(q)\Delta_n)_{1\leq q\leq \widehat{\lambda}_n}$ has a known distribution and this error can be averaged. We can easily prove the following result:

$$\frac{m\beta_n\Delta_n}{(1-e^{-m\beta_n\Delta_n})\,\widehat{\lambda}_n}\sum_{q=1}^{\widehat{\lambda}_n}\left(\Delta_{\mathcal{I}_n(q)}X\right)^m \to \int_{\mathbb{R}} x^m \nu(dx)$$

in probability for every integer $m > 0$ such that $\int_{\mathbb{R}} x^m \nu(dx) < \infty$. The proof for the case $m = 1$ appears in the proof of Theorem 3.1. Combining this result with an estimator of β_n provided in the following allows to have an estimator for the moments of ν.

2.3 Estimator of β_n

We are ready go construct an estimator of β_n. Define the sign function as $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and -1 otherwise. On the event $\{\widehat{\lambda}_n > 0\}$, define $\widehat{\beta}_n$ via

$$\exp(-\Delta_n\widehat{\beta}_n) = \max\left\{1 + \frac{\sum_{q=1}^{\widehat{\lambda}_n}\mathrm{sgn}(\Delta^n_{\mathcal{I}_n(q)}X)\big(\Delta^n_{\mathcal{I}_n(q)+1}X + 2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{\mathcal{I}_n(j)}X\big)}{\sum_{q=1}^{\widehat{\lambda}_n}|\Delta^n_{\mathcal{I}_n(q)}X|}, \Delta_n\right\} \quad (3.10)$$

and set $\widehat{\beta}_n = 0$ otherwise. Our main result describes precisely the behaviour of $\widehat{\beta}_n$ under the different asymptotic regimes of interest:

Theorem 3.1. *Work Assumption 3.1, 3.2 and 3.3. Let $\beta_n\sqrt{\lambda_n\Delta_n} \to \infty$.*

i) The following expansion holds on the set $\{\widehat{\lambda}_n > 0\}$:

$$\frac{\widehat{\beta}_n - \beta_n}{\beta_n} = \mathcal{M}_n + \mathcal{V}_n\mathcal{J}_n^T,$$

where

$$\mathcal{M}_n = e^{\beta_n\Delta_n}\frac{\lambda_n}{\beta_n}\frac{x\star\nu\,\mathrm{sgn}(x)\star\nu}{|x|\star\nu}\Big(\frac{e^{\beta_n\Delta_n}-1}{\beta_n\Delta_n}-1\Big),$$

$\mathcal{V}_n = (\mathcal{V}_n^{(i)})_{1\le i\le 4} \in \mathbb{R}^4$ *is such that*

$$\mathcal{V}_n^{(1)} = e^{\beta_n\Delta_n}\frac{\sqrt{\lambda_n}}{\sqrt{3}\beta_n|x|\star\nu}\sqrt{(\mathrm{sgn}(x)\star\nu)^2|x|^2\star\nu+(x\star\nu)^2-2\mathrm{sgn}(x)\star\nu|x|^2\star\nu},$$

$$\mathcal{V}_n^{(2)} = e^{\beta_n\Delta_n}\min\Big\{\sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(1-e^{-2\beta_n\Delta_n})}{2\beta_n\Delta_n}},\frac{\lambda_n}{\beta_n}\Big\},$$

$$\mathcal{V}_n^{(3)} = e^{\beta_n\Delta_n}\frac{\sqrt{\int_0^1\sigma_s^2ds}}{|x|\star\nu\sqrt{\lambda_n}\beta_n\sqrt{\Delta_n}},$$

$$\mathcal{V}_n^{(4)} = e^{\beta_n\Delta_n}\Big(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\Big)\frac{\sqrt{\int_0^1\sigma_s^2ds}\sqrt{\Delta_n}}{|x|\star\nu\sqrt{\lambda_n}},$$

and $\mathcal{J}_n = (\mathcal{J}_n^{(i)})_{1\le i\le 4} \in \mathbb{R}^4$ *is bounded in probability as* $n\to\infty$.

ii) If $\lambda_n\to\infty$, *then*

$$(\mathcal{J}_n^{(3)},\mathcal{J}_n^{(4)})\to\mathcal{N}(0,\mathrm{Id}_{\mathbb{R}^2})$$

in distribution as $n\to\infty$,

iii) If $\lambda_n\to\infty$, $|x|^3\star\nu<\infty$ *and* $(\mathrm{sgn}(x)\star\nu)^2|x|^2\star\nu+(x\star\nu)^2-2\mathrm{sgn}(x)\star\nu|x|^2\star\nu\neq 0$, *we have*

$$(\mathcal{J}_n^{(1)},\mathcal{J}_n^{(3)},\mathcal{J}_n^{(4)})\to\mathcal{N}(0,\mathrm{Id}_{\mathbb{R}^3})$$

in distribution as $n\to\infty$,

iv) If $\beta_n/\lambda_n^2\to 0$, *the conditions of* iii) *and* $|x|^4\star\nu<\infty$ *hold together, we finally obtain*

$$\mathcal{J}_n\to\mathcal{N}(0,\mathrm{Id}_{\mathbb{R}^4}).$$

in distribution as $n\to\infty$.

The asymptotic regime (3.5)-(3.6)-(3.7) and the need for β_n to grow to ∞ with n discussed in Section 1.2 appear naturally in Theorem 3.1. Some remarks are in order to analyse the different errors terms: **1)** the term of order $1/(\beta_n\sqrt{\lambda_n\Delta_n})$ accounts for the presence of a Brownian motion in the term $(X_t^c)_{t\ge 0}$. When λ_n is bounded, we need $\beta_n\sqrt{\Delta_n}\to\infty$ or equivalently $\sqrt{\Delta_n}=o(\beta_n\Delta_n)$: the size of the slope of $(Z_t^\beta)_{t\ge 0}$ after a jump needs to dominate the Brownian motion part which is of order $\sqrt{\Delta_n}$. In the case where $\lambda_n\to\infty$, we can average the error due to the Brownian martingale part and then diminish the order of the error. In that case, we do not need the restriction $\sqrt{\Delta_n}=o(\beta_n\Delta_n)$ anymore but rather $\sqrt{\Delta_n/\lambda_n}=o(\beta_n\Delta_n)$. **2)** The error terms of order $\min\{\frac{1}{\sqrt{\beta_n}},\frac{\lambda_n}{\beta_n}\}$, $\frac{\sqrt{\lambda_n}}{\beta_n}$ and $\lambda_n\Delta_n$ account the jumps that occur before the observation increment used to estimate the slope of the process. **3)** The term $2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{\mathcal{I}_n(j)}X$ introduced in the definition of $\widehat{\beta}_n$ in (3.10) is a bias correction that enables us to obtain a consistent estimator in the case $\lambda_n/\beta_n\approx 1$.

3 Practical implementation

3.1 Choice of the threshold

The method to detect the jumps is based on the use of the threshold which is proportional to $\Delta_n^{\frac{1}{2}-\varpi}$; this method is classic for jumps estimation in high frequency statistics. Concerning the choice of

this threshold and the choice of ϖ, no exact method is provided in the literature. However, the threshold is recommended to be chosen of the form $C\hat{\sigma}\Delta_n^{\frac{1}{2}-\varpi}$ in [2, Section 5.3] and [3, Section 6.2.2, p. 187] where C is a constant and $\hat{\sigma}$ is an estimator of the integrated volatility $\int_0^1 \sigma_s^2 ds$. They advise to choose $\frac{1}{2}-\varpi$ closed to 0.5, that is ϖ close to 0. Moreover, [2] suggests to choose C between 3 and 5. These choices are the one considered in the literature, see [16] ($C=3$) and [17] ($C=4$).

It remains to find an estimator of the integrated volatility. A natural choice is the multipower variation estimator, see [4] and [27] for more details. The order of the multipower variation estimator is set to 20 in the practical applications in this paper, which is high compared to the orders typically chosen in the literature. This choice is justified by the strong mean reversion of the spikes. As for the jumps, spikes have large increments that need to be compensated in the multipower variation estimator and can be present during two or three time steps. We compensate these large increments with a higher order of the multipower variation. Some simulations on simple models show that an order of 20 seems to be reasonable.

3.2 Numerical illustration

In this section, we use our estimation procedure to estimate the parameter β_n on simulated data of the process defined by (3.2)-(3.3)-(3.4) on simulated data in order to illustrate the results of Theorem 3.1. To be consistent with real data, the process is simulated with a step time $\Delta_n = 10^{-4}$, which is the order of the size of one year of data with a step time of one hour, see Section 3.3 for calibration on real data. We pick

$$dX_t^c = X_t^c\big((\tfrac{1}{2}2^2 - 100\log(X_t^c))dt + 2dW_t\big) \tag{3.11}$$

corresponding to the logarithm of an Ornstein-Uhlenbeck process with the mean reverting parameter equals to 100 and volatility parameter equal to 2. The sizes of the jumps follow the law $0.4\,(-\mathcal{E}\,(15)) + 0.6\mathcal{E}\,(10)$, with $\mathcal{E}(\rho)$ denoting the exponential distribution with parameter $\rho > 0$. Figure 3.2 illustrates a sample paths of the process for different parameters λ_n and β_n. We realize 10000 simulations. In Table 3.1, we consider the estimator of β in the case where the jump times and sizes are known. Concerning $\widehat{\beta}_n$, we first use jump detection under Assumption 3.3 (II): we only keep increments having successive opposite signs. We use three different thresholds for the detection of jumps of the form $C\hat{\sigma}\Delta_n^{0.49}$, corresponding to $\varpi = 0.01$, with $\hat{\sigma}$ the multi-power variation of order 20 of X : $C = 3$ in Table 3.2, $C = 4$ in Table 3.3 and $C = 5$ in Table 3.4. We only consider the increments verifying $\Delta_i^n X \Delta_{i+1}^n X < 0$. In Table 3.5, we only keep the increments greater bigger than the threshold that we fix to $5\hat{\sigma}\Delta_n^{0.49}$.

As expected, quality of convergence increases with the values of β_n and λ_n. In the case where $\beta_n = 300$ and λ_n, the convergence is not good due to the value of $\frac{1}{\sqrt{\lambda_n \Delta_n}\beta_n} = 0.1$. Concerning the method to detect the jump, the one filtering the strong mean reverting increments gives better results, except for $\beta_n = 300$ corresponding to a low value of β. We also observe that false jump detection impacts the quality of estimation of β_n but missing jumps does not impact it: we conclude that the second method with filtering is more promising than the first one.

3.3 Practical implementation on real data

In this section, we use our procedure estimation on real data. Electricity spot historical data do exhibit spikes with strong mean reversion, see Figure 3.1. We expect to obtain relatively high values

(λ_n, β_n)	Mean	Quantile interval
$(10, 300)$	295.01	$[173.89, 411.81]$
$(10, 1000)$	996.92	$[854.81, 1136.97]$
$(75, 300)$	295.88	$[249.91, 342.86]$
$(75, 1000)$	995.63	$[925.65, 1066.34]$
$(75, 5000)$	4993.75	$[4799.59.59, 5194.22]$

Table 3.1: *Results on the estimator $\widehat{\beta}_n^c$ for different parameters of (λ_n, β_n) in the case of a model having continuous part defined in (3.11) and with jump sizes having law $0.4\left(-\mathcal{E}\left(15\right)\right) + 0.6\mathcal{E}\left(10\right)$. The mean and the quantile interval are computed with 10000 simulations and a step time of $\Delta_n = 10^{-4}$.*

(λ_n, β_n)	Mean	Quantile interval	% of jumps of false detection	% not detected
$(10, 300)$	534.89	$[311.81, 983.54]$	79.27	20.72
$(10, 1000)$	1128.04	$[955.14, 1463.84]$	55.03	10.94
$(75, 300)$	356.92	$[309.97, 434.46]$	0.53	24.30
$(75, 1000)$	1031.02	$[973.08, 1111.91]$	0.56	15.82
$(75, 5000)$	4997.81	$[4856.85, 5147.49]$	3.02	10.57

Table 3.2: *Results on the estimator $\widehat{\beta}_n$ with threshold $3\hat{\sigma}\Delta_n^{0.49}$ and keeping the increments having the opposite sign of the following increments. The results are given for different parameters of (λ_n, β_n) in the case of a model having continuous part defined in (3.11) and with jump sizes having law $0.4\left(-\mathcal{E}\left(15\right)\right) + 0.6\mathcal{E}\left(10\right)$. The results are computed with 10000 simulations and a step time of $\Delta_n = 10^{-4}$.*

(λ_n, β_n)	Mean	Quantile interval	% of jumps of false detection	% not detected
$(10, 300)$	358.79	$[258.22, 505.61]$	5.15	21.82
$(10, 1000)$	1025.00	$[903.85, 1163.90]$	3.35	12.47
$(75, 300)$	353.54	$[[307.26, 429.79]]$	0.057	25.96
$(75, 1000)$	1027.88	$[970.38, 1107.22]$	0.27	18.49
$(75, 5000)$	5003.33	$[4862.48, 5152.72]$	0.76	13.42

Table 3.3: *Results on the estimator $\widehat{\beta}_n$ with threshold $4\hat{\sigma}\Delta_n^{0.49}$ and keeping the increments having the opposite sign of the following increments. The results are given for different parameters of (λ_n, β_n) in the case of a model having continuous part defined in (3.11) and with jump sizes having law $0.4\left(-\mathcal{E}\left(15\right)\right) + 0.6\mathcal{E}\left(10\right)$. The results are computed with 10000 simulations and a step time of $\Delta_n = 10^{-4}$.*

for β_n, a necessary condition in order to apply our procedure estimation. The goal is to estimate the parameters of the process Z^{β_n}, that is λ_n and β_n using spot prices time series, assuming that the spot price is the sum of a continuous semimartingale and a spike process. We dispose of the following data:

1. French electricity EPEX spot prices between the first of January of 2015 (included) and the first of January 2017 (not included) with data each hour [1],

[1] Source: https://www.epexspot.com/

(λ_n, β_n)	Mean	Quantile interval	% of jumps of false detection	% not detected
$(10, 300)$	341.33	$[249.55, 458.29]$	0.34	22.94
$(10, 1000)$	1014.96	$[-895.84, 1142.14]$	0.25	13.98
$(75, 300)$	350.98	$[305.06, 427.46]$	0.01	27.62
$(75, 1000)$	1025.25	$[968.22, 1104.04]$	0.14	21.17
$(75, 5000)$	5002.79	$[4862.25, 5148.03]$	0.52	16.23

Table 3.4: *Results on the estimator $\widehat{\beta}_n$ with threshold $5\hat{\sigma}\Delta_n^{0.49}$ and keeping the increments having the opposite sign of the following increments. The results are given for different parameters of (λ_n, β_n) in the case of a model having continuous part defined in (3.11) and with jump sizes having law $0.4\left(-\mathcal{E}\left(15\right)\right) + 0.6\mathcal{E}\left(10\right)$. The results are computed with 10000 simulations and a step time of $\Delta_n = 10^{-4}$.*

(λ_n, β_n)	Mean	Quantile interval	% of jumps of false detection	% not detected
$(10, 300)$	315.2	$[248.7, 391.1]$	26.96	10.37
$(10, 1000)$	-416.43	$[-1993.92, 669.86]$	171.87	11.21
$(75, 300)$	263.99	$[181.78, 328.40]$	4.34	15.09
$(75, 1000)$	434.26	$[150.92, 674.31]$	54.90	18.47
$(75, 5000)$	-2572.70	$[-2572.70, -2554.73]$	168.76	15.54

Table 3.5: *Results on the estimator $\widehat{\beta}_n$ with threshold $5\hat{\sigma}\Delta_n^{0.49}$. The results are given for different parameters of (λ_n, β_n) in the case of a model having continuous part defined in (3.11) and with jump sizes having law $0.4\left(-\mathcal{E}\left(15\right)\right) + 0.6\mathcal{E}\left(10\right)$. The results are computed with 10000 simulations and a step time of $\Delta_n = 10^{-4}$.*

2. German electricity EPEX spot prices between the first of January of 2015 (included) and the first of January 2017 (not included) with data each hour [1],

3. Australian electricity spot prices in Queensland between the first of January of 2015 (included) and the first of January 2017 (not included) with data each 30 minutes [2].

The Australian market is particularly interesting because of the number of spikes and because of the frequency of the data. We estimate those parameters using a threshold $v_n = C\hat{\sigma}\Delta_n^{0.49}$, corresponding to $\varpi = 0.01$ with $\hat{\sigma}$ the multi-power variation of order 20 and C a constant set to 3, 4 or 5 and filtering the mean reversion increments. Results are presented in Table 3.6. Figure 3.1 gives the time series of these three sets of data with jumps time estimated in the case $C = 5$.

Market	$C = 3$	$C = 4$	$C = 5$
French	$(100, 19169.57)$	$(51, 20259.01)$	$(35, 21042.533)$
German	$(145, 9848.07)$	$(62, 13438.35)$	$(34, 14531.32)$
Australian	$(337, 22897.06)$	$(227, 22883.81.6)$	$(177, 22884.36)$

Table 3.6: *Estimation of (λ_n, β_n) for different markets using a threshold of the form $C\hat{\sigma}\Delta_n^{0.49}$ where $\hat{\sigma}$ is the multi-power variation of order 20 and C takes different values.*

[2] Source: https://www.aemo.com.au/

4 Application to electricity spot price

Let us consider the forward product that delivers one MWh of electricity during one hour at time T and with price $f(t,T)$ at time t. For risk management purposes, it is a valuable property of a model to provide analytical formulas for $f(t,T)$. Two approaches generally exist for the modeling of electricity prices:

1. The spot price approach: we model the spot price $(S_t)_{t\geq 0}$ under a risk neutral probability $\mathbb{Q}$, which makes $f(t,T)$ martingale and $f(t,T)$ is defined by
$$f(t,T) = \mathbb{E}^{\mathbb{Q}}\big[S_T \,|\, \mathcal{F}_t\big].$$

2. The forward approach: we model directly the forward curve $f(t,T)$ and the spot price S is defined by
$$S_T = \lim_{t\to T} f(t,T).$$
In general, we can find an equivalence between the two approaches. In Section 4.2, we consider the second approach and we model the forward curve. However, we need a a spike component in the spot. Thus, in Section 4.1, we compute the forward prices starting from the spot modeling and we find the equivalent forward dynamics to have the component Z^β in the spot price.

4.1 The spot price approach

Suppose that we have a model under a risk neutral probability $\mathbb{Q}$ on the spot price process $(S_t)_{t\geq 0}$ of the form $S_t^c + Z_t^\beta$. Forward products with value $f(t,T,\theta)$ at time t delivers electricity during a period θ. By absence of arbitrage opportunity, we must have
$$f(t,T,\theta) = \frac{1}{\theta}\int_T^{T+\theta} f(t,u)\,du.$$

On the market, we do not observe directly $f(t,T)$ but products of the form $f(t,T,\theta)$. For instance, the One Month Ahead (1MAH) is the product delivering during one month, each hour, 1kWh of electricity. The delivery starts the first day of next month. The One Week Ahead (1WAH) is the one delivering during one week with delivery starting next Monday.

Theorem 3.2. *Suppose that the spot price is modelled by $S_t = S_t^c + Z_t^\beta, \quad t \geq 0$, according to (3.2)-(3.3)-(3.4) under a risk-neutral probability $\mathbb{Q}$.*

1. *We have an explicit representation of $f(t,T) = \mathbb{E}^{\mathbb{Q}}\big[S_T \,\big|\, \mathcal{F}_t\big]$ given by*
$$f(t,T) = f^c(t,T) + f^\beta(t,T),$$
with
$$f^c(t,T) = \mathbb{E}^{\mathbb{Q}}\big[S_T^c \,\big|\, \mathcal{F}_t\big]$$
and
$$f^\beta(t,T) = e^{-\beta(T-t)} Z_t^\beta + \frac{\lambda x \star \nu}{\beta}\big(1 - e^{-\beta(T-t)}\big).$$

2. *We also have* $f(t,T,\theta) = f^c(t,T,\theta) + f^\beta(t,T,\theta)$, *with* $f^c(t,T,\theta) = \frac{1}{\theta}\int_T^{T+\theta} f^c(t,u)\,du$ *and*

$$f^\beta(t,T,\theta) = \frac{e^{-\beta(T-t)} - e^{-\beta(T+\theta-t)}}{\beta\theta} Z_t^\beta + \frac{\lambda x \star \nu}{\beta}\big(1 - \frac{e^{-\beta(T-t)} - e^{-\beta(T+\theta-t)}}{\beta\theta}\big).$$

The term $f^\beta(t,T,\theta)$ is then negligible for the computation of forward products if $\frac{\lambda}{\beta}$ and $\frac{1}{\beta\theta}$ are small. In this case, we can estimate the jump process part on the spot and calibrate $(S_t^c)_{t\geq 0}$ on forward products. The proof is a direct consequence of the formula (3.2)-(3.3)-(3.4) and can be readily obtained by integrating.

4.2 The forward approach

From now, instead of putting a model on the spot price, we specifiy a model for the forward prices. Let thus $f(t,T)$ denote the forward time at time t with maturity T. We assume that we have a decomposition of the form

$$f(t,T) = \int_0^t \mu_s ds + f^c(t,T) + f^\beta(t,T),$$

where $df^c(t,T) = f^c(t,T)\left(\sigma_l dW_t^l + \sigma_s e^{-\alpha(T-t)} dW_t^s\right)$ with $(W_t^l, W_t^s)_{t\geq 0}$ a two-dimensional Brownian motions under the historic probability $\mathbb{P}$ with correlation ρ. This dynamics corresponds to the popular two factors model for forward prices of electricity [15]. The short term volatility $\sigma_s e^{-\alpha(T-t)}$ captures the Samuelson effect: the volatility increases when $T-t$ decreases. In order to have the equivalence between $(Z_t^\beta)_{t\geq 0}$ and $f^\beta(t,T)$, according to Theorem 3.2, we model f^β by

$$f^\beta(t,T) = \int_0^t \int_{\mathbb{R}} x e^{-\beta(T-t)} \underline{p}(ds,dx).$$

The spot price is then equal to $S_t = \int_0^t \mu_s ds + S_t^c + Z_t^\beta$ with

$$S_t^c = f^c(0,t)\exp\big(-\frac{\sigma_l^2 t}{2} - \frac{\sigma_s^2}{2}\int_0^t e^{-2\alpha(t-u)}du + \sigma_l dW_t^l + \sigma_s \int_0^t e^{-\alpha(t-u)} dW_u^s\big),$$

having

$$\frac{dS_t^c}{S_t^c} = \big(\partial_t \log\left(f^c(0,t)\right) + \alpha\sigma_s^2 \int_0^t e^{-2\alpha(t-u)}du + \alpha\sigma_s \int_0^t e^{-\alpha(t-u)} dW_u^s\big)dt + \sigma_l dW_t^l + \sigma_s dW_t^s. \quad (3.12)$$

If $\frac{\lambda}{\beta}$ and $\frac{1}{\beta\theta}$ are small, we can approximate $f(t,T,\theta)$ by $f^c(t,T,\theta)$ and estimate $f^c(t,T)$ using the forward products. An estimation procedure is given by Féron and Daboussi [10] to estimate the parameters in this model.

4.3 Change of measure and pricing of options

We have seen that forward products are not impacted by the spikes if $\frac{\lambda}{\beta}$ is small. However, it can have an important impact on options on the spot, for instance strip call options, with payoff of the form $\sum_{i=1}^p (S_{t_i} - K)^+$ for prescribed dates t_i. If we consider an option with payoff having a single component $(S_t - K)^+$, the jump process will have a slow impact on the price of this option. The

probability to have a jump at time t is equal to 0 and even if there is a jump before, it disappears very quickly. However, the jump process may have a significant impact on the value of options with payoff $\sum_{i=1}^{p}\left(S_{t_i}-K\right)^{+}$ because the probability of having spikes on $[0,1]$ is non zero. (Note that only upward spikes will have an impact on the price of these options.)

Unlike spot prices, forward contracts are tradable assets. In the following, we assume absence of arbitrage opportunity. According to the fundamental theorem of asset pricing, there exists a probability measure $\mathbb{Q}$ equivalent to the historical measure $\mathbb{P}$ such that $f(t,T)$ is a local martingale under $\mathbb{Q}$. Because of the presence of jumps, the market is incomplete and $\mathbb{Q}$ is not unique. According to [24, Theorem 2], there exists a predictable process $(\gamma_t)_{t\geq 0}$ and a predictable process $(Y(.,t,x))_{t\geq 0,x\in\mathbb{R}}$ such that:

1) $\mu_t+\gamma_t c_t+\int_0^t\int_{\mathbb{R}} x\lambda Y(t,x)\,e^{-\beta(T-t)}\nu(dx)=0$ ($\mathbb{P}\otimes dt$ almost-surely),

2) $\int_0^1\gamma_s^2 c_s ds<\infty$ almost surely,

3) $\int_0^1\int_{\mathbb{R}}|x|^2\wedge|x|Y(t,x)\,e^{-\beta(T-t)}\lambda\nu(dx)<\infty$ ($\mathbb{P}\otimes dt$ almost -surely).

with c_t equal to $f^c(t,T)\sqrt{\sigma_l^2+\sigma_s^2e^{-\alpha(T-t)}}$ in our case. Under the equivalent measure, $f(t,T)$ is an Itô semi-martingale with drift 0, volatility c and jump measure $p^*=Yp$ following

$$df(t,T)=df^c(t,T)+df^\beta(t,T)$$

with

$$df^c(t,T)=f^c(t,T)\left(\sigma_l dW_t^{l,*}+\sigma_s e^{-\alpha(T-t)}dW_t^{s,*}\right),$$

$$df^\beta(t,T)=\int_{\mathbb{R}} xe^{-\beta(T-t)}\left(\underline{p^*}(dt,dx)-\lambda Y(t,x)\,\nu(dx)\,dt\right)$$

for two Brownian motions $(W^{s,*},W^{l,*})$ under the new measure. The volatility does not change unlike the intensity and the law of jump sizes of the Poisson process.

In order to choose the martingale measure change, one usually choose a certain optimisation criterion. The most common criterion is the local risk-minimisation introduced by Föllmer and Schweizer (see [26] for details). The variance of the cost of the strategy is minimised locally, infinitesimally at each time. This strategy corresponds to choose as risk neutral probability the minimal martingale measure defined in [11]. Under certain assumptions, this measure is a true probability measure and the asset is a local martingale under this measure. Furthermore, the intensity changes and depends on the drift μ, which is also true for most common criteria. Since we work on a finite time framework, the drift is not identifiable and it is not possible to estimate it. We may alternatively choose the historical approach of Merton consisting of picking a change of probability that does not affect the intensity and the jump sizes of the Poisson measure [20]. The equivalent probability measure is defined by

$$\frac{d\mathbb{Q}^M}{d\mathbb{P}}=\exp\Big(-\int_0^1\theta_u d\frac{\sigma_l W_u^l+\sigma_s W_u^s}{\sqrt{\sigma_l^2+\sigma_s^2e^{-2\alpha(T-u)}}}-\frac{1}{2}\int_0^1\theta_u^2du\Big)$$

with $\theta_u=\frac{-(\mu_u+e^{-\beta(T-u)}\int_{\mathbb{R}}x\nu(dx))}{f^c(u,T)\sqrt{\sigma_l^2+\sigma_s^2e^{-2\alpha(T-u)}}}$. The Nikonov condition is satisfied so it defines in turn a genuine probability measure. Under $\mathbb{Q}^M$, the price of the forward contract $f(t,T)$ follows the dynamics $df(t,T)=df^c(t,T)+df^\beta(t,T)$ with

$$df^c(t,T)=f^c(t,T)\left(\sigma_l dW_t^{l,\mathbb{Q}^M}+\sigma_s e^{-\alpha(T-t)}dW_t^{s,\mathbb{Q}^M}\right)$$

and

$$df^{\beta}(t,T) = \int_{\mathbb{R}} xe^{-\beta(T-t)} \left(\underline{p}(dt,dx) - \lambda\nu(dx)\,dt\right),$$

where $W^{l,\mathbb{Q}^M}$ and $W^{s,\mathbb{Q}^M}$ are two $\mathbb{Q}^M$-Brownian motions. Merton chooses this probability considering that the risk associated to the jumps is diversifiable. As noticed in Tankov in [9, Section 10.1], using this strategy leaves one exposed to the risk of the jumps. It only corrects the average effect of jumps (provided the jump component of the electricity price is independent of the other asset, which is the case here: we understand the electricity spikes caused by physical exogenous events; it can in particular be caused by the production capacity and the demand which are not assets (see the structural model of Aid et al. [1] for instance). Finally, the price of an option with payoff $H(S_T) = H(f(T,T))$ at time t is given by

$$\mathbb{E}^{\mathbb{Q}^M}\left[H(S_T) \mid \mathcal{F}_t\right].$$

4.4 Application to the French market with real data

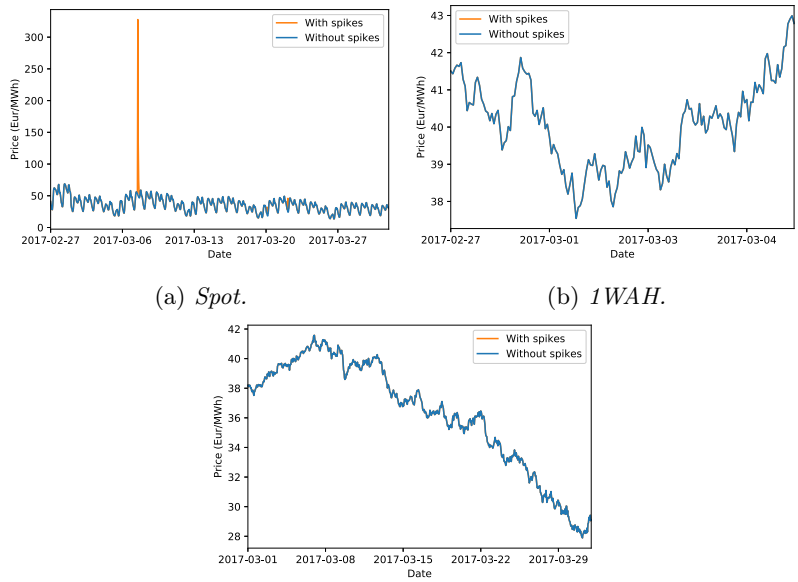
In the following, we focus on the French market and we work on the model of Section 4.2. We dispose of the daily forward prices in 2015 and 2016.

Parameters of f^{β} We use the parameters found in Table 3.6 to calibrate Z^{β} on the spot prices. We model the size of the jumps by its empirical distribution, each jump being estimated with $\Delta^n_{\mathcal{I}_n(q)}X$, knowing that a bias is present.

Parameters of f^c We observe on the market the forward products of the form $f(t,T,\theta)$ (θ is between one week and one year). As $\frac{\lambda}{\beta}$ and $\frac{1}{\beta\theta}$ are small, we can neglect the part of the jump part on the forward prices and consider that the forward products have only a continuous part. We use the method of Féron and Daboussi [10] to calibrate the parameters of f^c on the forward data. We find for the different parameters $\alpha = 12.56\ y^{-1}$, $\sigma_s = 1.03\ y^{-\frac{1}{2}}$, $\sigma_l = 0.25\ y^{-\frac{1}{2}}$ and $\rho = -0.11$.

Forward products In Figure 3.3, we display a simulation of the spot price, the 1WAH and the 1MAH with and in absence of spikes. The parameters of the spike component are the one of Table 3.6 with $\varpi = 0.15$. We observe that the difference between the trajectory of the forward products with and without spikes is very small but significant for the spot price.

Strip options We recall that strip options are of the form $\sum_{i=1}^{p}(S_{t_i} - K)^+$ with price equal to $\mathbb{E}^{\mathbb{Q}^M}\left[\sum_{i=1}^{p}(S_{t_i} - K)^+\right]$. We give in Table 3.7 confidence intervals at level 95% for the price of strip options computed using Monte Carlo method with 10000 simulations. The strip options delivers during one year each hour and we consider the different strikes 100, 200 and 300. We consider the case where there is no spikes and the cases with spikes using the different threshold of the form $C\hat{\sigma}\Delta_n^{0.49}$ with $C = 3$, $C = 4$ and $C = 5$. Considering spikes leads to higher value for the strip options. Furthermore, options valued at zero have now non negligible values. We notice that the choice of the threshold have a low impact on the price of the option. Indeed, a higher C leads to less jumps, but the empirical distribution contains only the larger jumps which are the ones impacting the price for high strikes.



(a) *Spot.* (b) *1WAH.*

(c) *1MAH.*

Figure 3.3: *Simulation of different products in a two factor model with and without spikes between the* 27th *of February 2017 and the* 31st *of March 2017. We illustrate the spot, the 1WAH starting the* 27th *of February 2017 and the 1MAH starting the* 01st *of March 2017.*

5 Proofs

In the following proofs, we set the drift μ to zero. Generalizing to the non-null drift case is done using the usual argument based on the Girsanov theorem.

5.1 Proof of Proposition 3.1

The proof follows the path of [3, Theorem 10.26, p.374] and the one of [19]. We will denote by ξ_ν a random variable distributed according to ν. We set $v_n = \Delta_n^{-\varpi}$. Let

$$\mathcal{A}_n = \big\{ i \in \{1, \dots, n, i \neq i(n,q)\ \forall q \geq 1\} \big\}.$$

Proof under Assumption 3.3 (I)

We first need to prove

$$\mathbb{P}\big(\sup_{i \in \mathcal{A}_n} \tfrac{|\Delta_i^n X|}{\sqrt{\Delta_n}} > v_n \big) \to 0, \tag{3.13}$$

$$\mathbb{P}\big(\inf_{i \in \mathcal{A}_n^c} \tfrac{|\Delta_i^n X|}{\sqrt{\Delta_n}} < v_n \big) \to 0 \tag{3.14}$$

and

$$\mathbb{P}\big(\sup_{i \in \{1,\dots,n\}} \Delta_i^n N \geq 2 \big) \to 0. \tag{3.15}$$

Model / Strike	100	200	300
Without spike	$[1716.22, 1806.83]$	$[0.0089, 0.063]$	$[0, 0]$
Spikes, $C = 3$	$[2482.57, 2576.17]$	$[434.21, 450.26]$	$[264.14, 274.00]$
Spikes, $C = 4$	$[2442.66, 2536.26]$	$[412.98, 428.44]$	$[251.72, 262.29]$
Spikes, $C = 5$	$[2417.24, 2510.79]$	$[397.74, 412.76]$	$[242.04, 252.41]$

Table 3.7: *Confidence intervals at level* 95% *for the price of strip options computed using Monte Carlo method with* 10000 *simulations for different strikes and different models.*

Step 1). We have

$$\frac{\Delta_i^n X}{\sqrt{\Delta_n}} = \frac{-\beta_n \int_{t_{i-1}}^{t_i} Z_s^\beta ds}{\sqrt{\Delta_n}} + \frac{\Delta_i^n X'}{\sqrt{\Delta_n}}$$

with

$$X_t' = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{\mathbb{R}} x \underline{p}\left(ds, dx\right).$$

By [3, Equation (10.71), p.374] we have $\mathbb{P}\big(\sup_{i\in\mathcal{A}_n} \frac{|\Delta_i^n X'|}{\sqrt{\Delta_n}} > v_n\big) \to 0$, Therefore, in order to prove (3.13), we need to show that

$$\mathbb{P}\big(\sup_{i\in\mathcal{A}_n} \frac{|\beta_n \int_{t_{i-1}}^{t_i} Z_s^\beta ds|}{\sqrt{\Delta_n}} > v_n\big) \to 0. \tag{3.16}$$

For $i \in \mathcal{A}_n$, we have $|\beta_n \int_{t_{i-1}}^{t_i} Z_s^\beta ds| = \left(1 - e^{-\beta_n \Delta_n}\right) |Z_{t_{i-1}}^\beta|$ and $\sup_{i\in\mathcal{A}_n} \frac{|\beta_n \int_{t_{i-1}}^{t_i} Z_s^\beta ds|}{\sqrt{\Delta_n}}$ is dominated by $\beta_n \sqrt{\Delta_n} \sup_{t\in[0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}\left(ds, dx\right)$. Also $\mathbb{P}\big(\sup_{t\in[0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}\left(ds, dx\right) > \frac{v_n}{\beta_n \sqrt{\Delta_n}}\big)$ equals the probability that $\sup_{t\in[0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t_i - s)} \underline{p}\left(ds, dx\right)$ crosses the barrier $\frac{v_n}{\beta_n \sqrt{\Delta_n}}$ at least one time. By Markov's inequality it is further bounded by $2\lambda_n \mathbb{P}\big(|\xi_\nu| > \frac{v_n}{\beta_n \sqrt{\Delta_n}}\big)$ according to [7, Equation (10)] which converges to 0 under Assumption 3.3 (ii). Thus

$$\mathbb{P}\big(v_n \sqrt{\Delta_n} \leq \sup_{i\in\{1,\ldots,n\}} \beta_n \Delta_n |Z_{t_i}^\beta|\big) \to 0 \tag{3.17}$$

and (3.16) follows which completes the proof of (3.13).

Step 2). We next turn to (3.15). The left hand side of (3.15) is equal to $\mathbb{P}\left(\cup_{i=1}^n \Delta_i^n N \geq 2\right) \lesssim \lambda_n^2 \Delta_n$ which converges to 0 if $\lambda_n \sqrt{\Delta_n} \to 0$. With no loss of generality we may (and will) work on the set $\{\sup_{i\in\{1,\ldots,n\}} \Delta_i^n N \leq 1\}$. In the interval $\left(\left(i\left(n,q\right) - 1\right) \Delta_n, i\left(n,q\right) \Delta_n\right]$, there is only one jump and we have

$$\Delta_{i(n,q)}^n Z^\beta = -\left(1 - e^{-\beta_n \Delta_n}\right) Z_{t_{i(n,q)-1}} + e^{-\beta_n (i(n,q) - T_q)} \Delta X_{T_q}$$

for all $q \geq 1$, therefore

$$\begin{aligned}
\frac{|\Delta_{i(n,q)}^n X|}{\sqrt{\Delta_n}} &\geq \frac{e^{-\beta_n \Delta_n} |\Delta X_{T_q}|}{\sqrt{\Delta_n}} - \frac{|\Delta_{i(n,q)}^n X^c|}{\sqrt{\Delta_n}} - \frac{|\left(1 - e^{-\beta_n \Delta_n}\right) Z_{t_{i(n,q)-1}}^\beta|}{\sqrt{\Delta_n}} \\
&\geq \inf_{q\in\{1,\ldots,N_1\}} \frac{e^{-\beta_n \Delta_n} |\Delta X_{T_q}|}{\sqrt{\Delta_n}} - \sup_{i\in\bar{\mathcal{A}}_n} \frac{|\Delta_i^n X^c|}{\sqrt{\Delta_n}} - \sup_{i\in\{1,\ldots,n\}} \frac{\beta_n \Delta_n |Z_{t_i}^\beta|}{\sqrt{\Delta_n}}.
\end{aligned}$$

It follows that $\mathbb{P}\big(\inf_{i\in\bar{\mathcal{A}}_n}\frac{|\Delta_i^n X|}{\sqrt{\Delta_n}}\leq v_n\big)$ is dominated by the sum of the three following terms:

$$\mathbb{P}\big(\inf_{q\in\{1,\dots,N_1\}}|\Delta X_{T_q}|e^{-\beta_n\Delta_n}\leq 3v_n\sqrt{\Delta_n}\big), \tag{3.18}$$

$$\mathbb{P}\big(\inf_{q\in\{1,\dots,N_1\}}|\Delta X_{T_q}|e^{-\beta_n\Delta_n}\leq 3\sup_{i\in\bar{\mathcal{A}}_n}|\Delta_i^n X^c|\big) \tag{3.19}$$

and

$$\mathbb{P}\big(\inf_{q\in\{1,\dots,N_1\}}|\Delta X_{T_q}|e^{-\beta_n\Delta_n}\leq 3\sup_{i\in\{1,\dots,n\}}\beta_n\Delta_n|Z_{t_i}^\beta|\big). \tag{3.20}$$

The term (3.18) equals

$$\mathbb{E}\big[1-\big(\mathbb{P}\big(|\xi_\nu|>3v_n\sqrt{\Delta_n}e^{\beta_n\Delta_n}\big)\big)^{N_1}\big]=1-\exp\big(-\lambda_n\mathbb{P}(|\xi_\nu|\leq 3v_n\sqrt{\Delta_n}e^{\beta_n\Delta_n})\big)$$

and converges to 0 under the assumption $\lambda_n\mathbb{P}\big(|\xi_\nu|\leq\Delta_n^{\frac{1}{2}-\varpi}\big)\to 0$.The term (3.19) is dominated by

$$\mathbb{P}\left(\inf_{q\in\{1,\dots,N_1\}}|\Delta X_{T_q}|e^{-\beta_n\Delta_n}\leq 3v_n\sqrt{\Delta_n}\right)+\mathbb{P}\left(v_n\sqrt{\Delta_n}\leq\sup_{i\in\bar{\mathcal{A}}_n}|\Delta_i^n X^c|\right). \tag{3.21}$$

The left hand side of (3.21) is equal to (3.18) and converges to 0. According to [19, Corollary 3.3], for $i\in\{1,...,n\}$,

$$\mathbb{P}\left(\Delta_i^n X^c>v_n\sqrt{\Delta_n}\right)\leq 2e^{-\frac{v_n^2}{2\bar{\sigma}^2}}.$$

The right hand side of (3.21) is then dominated by

$$\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{P}\left(|\Delta_i^n X^c|\geq v_n\sqrt{\Delta_n}\right)\right)\leq\mathbb{E}\left(N_1\right)2e^{-\frac{v_n^2}{2\bar{\sigma}^2}}=2\lambda_n e^{-\frac{v_n^2}{2\bar{\sigma}^2}}\to 0.$$

The term (3.20) is dominated by

$$\mathbb{P}\left(\inf_{q\in\{1,\dots,N_1\}}|\Delta X_{T_q}|e^{-\beta_n\Delta_n}\leq 3v_n\sqrt{\Delta_n}\right)+\mathbb{P}\left(v_n\sqrt{\Delta_n}\leq\sup_{i\in\{1,\dots,n\}}\beta_n\Delta_n|Z_{t_i}^\beta|\right). \tag{3.22}$$

The left hand side of (3.21) is equal to (3.18) and converges to 0. The right hand side of (3.21) converges also to 0 according to (3.17).

Work under Assumption 3.3 (II)

First, we have

$$\mathbb{P}\left(\sup_{i\in\{0,\dots,n-k_n\}}|N_{(i+k_n)\Delta_n}-N_{i\Delta_n}|\geq 2\right)\lesssim\lambda_n^2\Delta_n k_n^2\to 0.$$

Thus, we can work on the set $\{\sup_{i\in\{0,\dots,n-k_n\}}|N_{(i+k_n)\Delta_n}-N_{i\Delta_n}|\leq 1\}$. We need to prove:

$$\mathbb{P}\left(\exists i\in\mathcal{A}_n:\frac{|\Delta_i^n X|}{\sqrt{\Delta_n}}>v_n\text{ and }\Delta_i^n X\Delta_{i+1}^n X<0\right)\to 0 \tag{3.23}$$

and

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n^c : \frac{|\Delta_i^n X|}{\sqrt{\Delta_n}} < v_n \text{ or } \Delta_i^n X \Delta_{i+1}^n X > 0\right) \to 0. \tag{3.24}$$

Step 1). The left hand side of (3.23) is dominated by the sum of

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n : \frac{|\Delta_i^n X^c|}{\sqrt{\Delta_n}} > \frac{v_n}{2} \text{ and } \Delta_i^n X \Delta_{i+1}^n X < 0\right) \tag{3.25}$$

and

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n : \frac{|\Delta_i^n Z^\beta|}{\sqrt{\Delta_n}} > \frac{v_n}{2} \text{ and } \Delta_i^n X \Delta_{i+1}^n X < 0\right). \tag{3.26}$$

(3.25) is dominated by

$$\mathbb{P}\left(\sup_{i\in\{1,..,n\}} |\Delta_n^i X^c| > \frac{v_n}{2\sqrt{\Delta_n}}\right) \leq 2ne^{-\frac{v_n^2}{2\bar{\sigma}^2}}$$

and converges to 0. We then only have to show that (3.26) converges to 0. We need to consider two cases: the one where there is a jump happening during a certain period before i and the one where there is no jumps during this period. (3.26) is dominated by the sum of

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n : \frac{|\Delta_i^n Z^\beta|}{\sqrt{\Delta_n}} > \frac{v_n}{2},\ \Delta_i^n X \Delta_{i+1}^n X < 0 \text{ and } \exists q \in \{1, \max(k_n, i)\} : i - q \in \mathcal{A}_n^c\}\right) \tag{3.27}$$

and

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n : \frac{|\Delta_i^n Z^\beta|}{\sqrt{\Delta_n}} > \frac{v_n}{2},\ \Delta_i^n X \Delta_{i+1}^n X < 0 \text{ and } \forall q \in \{1, \max(k_n, i)\} : i - q \in \mathcal{A}_n\}\right). \tag{3.28}$$

The case (3.27) corresponds to the case when there is a jump and then because we are on the set $\{\sup_{i\in\{0,\dots,n-k_n\}} |N_{(i+k_n)\Delta_n} - N_{i\Delta_n}| \leq 1\}$, $i+1 \in \mathcal{A}_n$ and

$$\Delta_{i+1}^n Z^\beta = -\left(1 - e^{-\beta_n \Delta_n}\right) Z_i^\beta = e^{-\beta_n \Delta_n} \Delta_i^n Z^\beta.$$

Thus, (3.27) is dominated by

$$\begin{aligned}
&\mathbb{P}\left(-|\Delta_i^n X^c||\Delta_{i+1}^n X^c| - |\Delta_i^n X^c| \frac{v_n\sqrt{\Delta_n}}{2} - |\Delta_{i+1}^n X^c| \frac{v_n\sqrt{\Delta_n}}{2} + \left(\frac{v_n\sqrt{\Delta_n}}{2}\right)^2 < 0\right) \\
&\leq 2\mathbb{P}\left(\sup_{i\in\{1,..,n\}} |\Delta_n^i X^c| > \frac{v_n\sqrt{\Delta_n}}{6}\right) + \mathbb{P}\left(\left(\sup_{i\in\{1,..,n\}} |\Delta_n^i X^c|\right)^2 > \frac{1}{3}\left(\frac{v_n\sqrt{\Delta_n}}{2}\right)^2\right) \to 0.
\end{aligned}$$

Concerning (3.28), we have in the case where there are no jumps that

$$|\Delta_i^n Z^\beta| = \left(1 - e^{-\beta_n \Delta_n}\right) e^{-\beta_n \Delta_n k_n} |Z^\beta_{(i-k_n)\Delta_n}| \leq \beta_n \Delta_n \sup_{t\in[0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}(ds, dx)\, e^{-\beta_n \Delta_n k_n}$$

and (3.28) is dominated by

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}(ds, dx) > e^{\beta_n \Delta_n k_n} \frac{v_n}{\beta_n \sqrt{\Delta_n}}\right) \\
&\leq \lambda_n \mathbb{P}\left(|\xi_\nu| > e^{\beta_n \Delta_n k_n} \frac{v_n}{\beta_n \sqrt{\Delta_n}}\right) \to 0
\end{aligned}$$

using the same argument as for (3.16) and (3.23) follows.

Step 2). The left hand side of (3.24) is dominated by the sum of

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n^c : \beta_n \Delta_n \frac{|\Delta_i^n X|}{\sqrt{\Delta_n}} \geq v_n\right), \tag{3.29}$$

and

$$\mathbb{P}\left(\exists i \in \mathcal{A}_n^c : \Delta_i^n X \Delta_{i+1}^n X \geq 0\right). \tag{3.30}$$

To prove that (3.29) converges to 0, the proof is similar to the one of (3.14). The only difference is that $\Delta_i^n Z$ is equal to $\left(1 - e^{-\beta_n \Delta_n}\right) e^{-\beta_n \Delta_n k_n}$ and instead of having the term

$$\mathbb{P}\left(v_n \sqrt{\Delta_n} \leq 3 \sup_{i \in \{1,\ldots,n\}} \beta_n \Delta_n |Z_{t_i}^\beta|\right),$$

we have the term

$$\mathbb{P}\left(v_n \sqrt{\Delta_n} \leq 3 \sup_{i \in \{1,\ldots,n\}} \beta_n \Delta_n e^{-\beta_n \Delta_n k_n} |Z_{t_i}^\beta|\right) \leq \lambda_n \mathbb{P}\left(|\xi_\nu| > e^{\beta_n \Delta_n k_n} \frac{v_n \sqrt{\Delta_n}}{\beta_n \Delta_n}\right) \to 0.$$

To prove the convergence of (3.30) to 0, it is sufficient to prove that $\Delta_{i(n,q)}^n X$ has the same sign than ΔX_{T_q} and that $\Delta_{i(n,q)+1}^n X$ has the opposite sign with probability one. We then have to prove that

$$\mathbb{P}\left(\inf_{q \in \{1,..,N_1\}} \Delta_{i(n,q)}^n X \Delta X_{T_q} < 0\right) \to 0 \tag{3.31}$$

and

$$\mathbb{P}\left(\sup_{q \in \{1,..,N_1\}} \Delta_{i(n,q)+1}^n X \Delta X_{T_q} > 0\right) \to 0. \tag{3.32}$$

We have $\Delta_{i(n,q)}^n X \Delta X_{T_q}$ that dominates the sum of

$$-\sup_{q \in \{1,..,N_1\}} |\Delta_{i(n,q)}^n X^c| |\Delta X_{T_q}|,$$

$$-\left(1 - e^{-\beta_n \Delta_n}\right) e^{-\beta_n \Delta_n k_n} \sup_{t \in [0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}\left(ds, dx\right) |\Delta X_{T_q}|$$

and

$$e^{-\beta_n \Delta_n} |\Delta X_{T_q}| \inf_{q \in \{1,..,N_1\}} |\Delta X_{T_q}|.$$

Thus, we have (3.31) dominated by the probability of the event

$$\Big\{ e^{-\beta_n \Delta_n} \inf_{1 \leq q \leq N_1} |\Delta X_{T_q}| < \sup_{1 \leq q \leq N_1} |\Delta_{i(n,q)}^n X^c| + \left(1 - e^{-\beta_n \Delta_n}\right) e^{-\beta_n \Delta_n k_n} \sup_{t \in [0,1]} \int_0^t \int_{\mathbb{R}} |x| e^{-\beta_n (t-s)} \underline{p}\left(ds, dx\right) \Big\}$$

that converges to 0 if we use a similar proof than the one of (3.14). The proof of (3.32) is similar because there is no jumps in the interval and $(i(n,q), i(n,q)+1]$ and

$$\begin{aligned}\Delta^n_{i(n,q)+1} Z^\beta &= -\left(1-e^{-\beta_n\Delta_n}\right) Z^\beta_{i(n,q)\Delta_n} \\ &= -\left(1-e^{-\beta_n\Delta_n}\right)\Delta^n_{i(n,q)} Z^\beta - \left(1-e^{-\beta_n\Delta_n}\right) Z^\beta_{(i(n,q)-1)\Delta_n}.\end{aligned}$$

The term (3.32) is then dominated by the probability of the event

$$\begin{aligned}\Big\{e^{-\beta_n\Delta_n}\left(1-e^{-\beta\Delta_n}\right)\inf_{q\in\{1,..,N_1\}}|\Delta X_{T_q}| < \sup_{q\in\{1,..,N_1\}}|\Delta^n_{i(n,q)}X^c| \\ + \left(1-e^{-\beta_n\Delta_n}\right)e^{-\beta_n\Delta_n k_n}\sup_{t\in[0,1]}\int_0^t\int_{\mathbb{R}}|x|e^{-\beta_n(t-s)}\underline{p}(ds,dx)\Big\}.\end{aligned}$$

The proof is then the same except we have to control the term

$$\mathbb{P}\left(\sup_{q\in\{1,..,N_1\}}|\Delta^n_{i(n,q)}X^c| > v_n\left(1-e^{-\beta\Delta_n}\right)\right)$$

which is bounded by $\lambda_n e^{-\frac{\left(v_n\left(1-e^{-\beta\Delta_n}\right)\right)^2}{2\bar{\sigma}}}$ and converges to 0 if $\lambda_n e^{-\left(\beta\Delta_n^{1-\varpi}\right)}\to 0$ and the term

$$\mathbb{P}\left(\sup_{t\in[0,1]}\int_0^t\int_{\mathbb{R}}|x|e^{-\beta_n(t-s)}\underline{p}(ds,dx) > e^{\beta_n\Delta_n k_n}v_n\sqrt{\Delta_n}\right)$$

that converges to 0 if $\lambda_n\mathbb{P}\left(|\xi_\nu| > e^{\beta_n\Delta_n k_n}\Delta_n^{\frac{1}{2}-\varpi}\right)\to 0$ and the first part of the Proposition follows.

5.2 Estimator of β_n in the case where the jump times and sizes are known

In order to prove Theorem 3.1, we start by giving an estimator of β_n when the jump times and sizes are known.

Proposition 3.3. *Let*

$$\widehat{\beta}^c_n = -\frac{1}{\Delta_n}\log\left(\left(1+\frac{\sum_{q\in E_n}\operatorname{sgn}(\Delta X_{T_q})\left(\Delta^n_{i(n,q)+1}X + 2\Delta_n\sum_{j=1}^{q-1}\Delta X_{T_j}\right)}{\sum_{q\in E_n}\operatorname{sgn}(\Delta X_{T_q})\Delta^n_{i(n,q)}X}\mathbf{1}_{N_1>0}\right)\vee\Delta_n\right).$$

with $E_n = \{q\in\{1,..,N_1\},\ i(n,q)+1\in\mathcal{A}_n$ *and* $i(n,q) < i(n,q+1)\}$. *We have under Assumption 3.1, Assumption 3.2 and* $\frac{1}{\beta_n\sqrt{\lambda_n\Delta_n}}\to 0$*:*

$$\frac{\widehat{\beta}^c_n-\beta_n}{\beta_n} = M_{c,n} + V_n^T J_n.$$

on the set $\{N_1>0\}$ *with*

$$M_{c,n} = e^{\beta_n\Delta_n}\frac{\lambda_n}{\beta_n}\frac{x\star\nu\operatorname{sgn}(x)\star\nu}{|x|\star\nu}\left(\frac{\left(e^{\beta_n\Delta_n}-1\right)}{\beta_n\Delta_n} - \frac{\beta_n\Delta_n}{1-e^{-\beta_n\Delta_n}}\right),$$

$$V_n = (V_{1,n}, V_{2,n}, V_{3,n}, V_{4,n})^T ,$$

$$\begin{cases} V_{1,n} &= e^{\beta_n \Delta_n} \frac{\sqrt{\lambda_n}}{\sqrt{3}\beta_n |x|\star\nu} \sqrt{(\mathrm{sgn}\,(x)\star\nu)^2 |x|^2\star\nu + (x\star\nu)^2 - 2\mathrm{sgn}\,(x)\star\nu|x|^2\star\nu} \\ V_{2,n} &= e^{\beta_n \Delta_n} \min\left(\sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(1-e^{-2\beta_n\Delta_n})}{2\beta_n\Delta_n}}, \frac{\lambda_n}{\beta_n}\right) \\ V_{3,n} &= e^{\beta_n \Delta_n} \frac{\sqrt{\int_0^1 \sigma_s^2 ds}}{|x|\star\nu\sqrt{\lambda_n}\beta_n\sqrt{\Delta_n}} \\ V_{4,n} &= e^{\beta_n \Delta_n} \left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right) \frac{\sqrt{\int_0^1 \sigma_s^2 ds}\sqrt{\Delta_n}}{|x|\star\nu\sqrt{\lambda_n}} \end{cases}$$

$$J_n = (J_{1,n}, J_{2,n}, J_{3,n}, J_{4,n})^T = O_p(1) ,$$

Furthermore, under the additional assumption $\lambda_n \to \infty$, we have

$$(J_{3,n}, J_{4,n})^T \xrightarrow{d} \mathcal{N}(0, I_2) .$$

Adding assumptions $|x|^3 \star \nu < \infty$ and

$$(\mathrm{sgn}\,(x)\star\nu)^2 |x|^2\star\nu + (x\star\nu)^2 - 2\mathrm{sgn}\,(x)\star\nu|x|^2\star\nu \neq 0,$$

we have

$$(J_{1,n}, J_{3,n}, J_{4,n})^T \xrightarrow{d} \mathcal{N}(0, I_3) .$$

Adding assumptions $\frac{\beta_n}{\lambda_n^2} \to 0$ and $|x|^4 \star \nu < \infty$, we have

$$J_n \xrightarrow{d} \mathcal{N}(0, I_4) .$$

The proof is given here. For simplification of the proof, we assume in the following that $(\sigma_t)_{0\le t\le 1}$ is a deterministic function. The generalization to the stochastic case can be done easily by taking all the elements of the proof. Only central limit theorems involving Brownian increments need further technical tools for their proof , which can be found in [14, Section 2.2.5].

The estimator of the slope

$$\widehat{s}_n^c = -\frac{\sum_{q\in E_n} \mathrm{sgn}(\Delta X_{T_q}) \left(\Delta^n_{i(n,q)+1} X + 2\Delta_n \sum_{j=1}^{q-1} \Delta X_{T_j}\right)}{\sum_{q\in E_n} \mathrm{sgn}\left(\Delta X_{T_q}\right) \Delta^n_{i(n,q)} X}$$

is the sum of the two following terms:

$$-\frac{\sum_{q\in E_n} \mathrm{sgn}\left(\Delta X_{T_q}\right) \left(\Delta^n_{i(n,q)+1} Z^\beta + 2\Delta_n \sum_{j=1}^{q-1} \Delta X_{T_j}\right)}{\sum_{q\in E_n} \mathrm{sgn}\left(\Delta X_{T_q}\right) \Delta^n_{i(n,q)} X} \mathbf{1}_{N_1>0} \tag{3.33}$$

and

$$-\frac{\sum_{q\in E_n} \Delta^n_{i(n,q)+1} X^c \mathrm{sgn}\left(\Delta X_{T_q}\right)}{\sum_{q\in E_n} \mathrm{sgn}\left(\Delta X_{T_q}\right) \Delta^n_{i(n,q)} X} \mathbf{1}_{N_1>0}. \tag{3.34}$$

If $i \in \mathcal{A}_n$,

$$\begin{aligned} -\Delta_i^n Z^\beta &= \left(1 - e^{-\beta_n\Delta_n}\right) Z^\beta_{t_{i-1}} \\ &= \left(1 - e^{-\beta_n\Delta_n}\right) \left(\Delta^n_{i-1} Z^\beta + Z^\beta_{t_{i-2}}\right) . \end{aligned}$$

Thus, we can write (3.33) as the sum of:

$$\left(1-e^{-\beta_n\Delta_n}\right)\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}Z^\beta}{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(\Delta^n_{i(n,q)}Z^\beta+\Delta^n_{i(n,q)}X^c\right)} \tag{3.35}$$

and

$$\left(1-e^{-\beta_n\Delta_n}\right)\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^\beta_{t_{i(n,q)-1}}-\frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right)}{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X}\mathbf{1}_{N_1>0} \tag{3.36}$$

$\widehat{s}^c_n$ is then the sum of (3.34), (3.35) and (3.36). From now, we work on the set $\{N_1>0\}$.

Convergence of the denominator

First, let us study the error we make approximating the denominator by $\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X$. The difference between the denominator and $\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X$ the two corresponds to the error we make by taking into account the interval of times with more than one jumps or the interval of the ones with a jump happening at the next interval. Thus

$$\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X-\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X$$

is equal to

$$\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n}$$

which is equal to the sum of

$$\sum_{q=1}^{N_1}\Delta^n_{i(n,q)}X^c\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n} \tag{3.37}$$

and

$$\sum_{q=1}^{N_1}\Delta^n_{i(n,q)}Z^\beta\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n}. \tag{3.38}$$

The expectation of (3.37) is equal to:

$$\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{E}\left(\Delta^n_{i(n,q)}X^c\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n}|N_1\right)\right).$$

However, X^c is independent from $\underline{p}$. Thus, the expectation of (3.37) is equal to

$$\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{E}\left(\Delta^n_{i(n,q)}X^c\right)\mathbb{E}\left(\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n}|N_1\right)\right)=0.$$

The variance is equal to

$$\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{E}\left(\left(\Delta^n_{i(n,q)}X^c\right)^2|\mathcal{F}_{t_{i(n,q)}}\right)\mathbb{E}\left(\mathbf{1}_{T_{q+1}-T_q<t_{i(n,q)}-T_q+\Delta_n}|N_1\right)\right) \tag{3.39}$$

and is dominated by

$$\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{E}\left(\left(\Delta^n_{i(n,q)}X^c\right)^2|\mathcal{F}_{t_{i(n,q)}}\right)\mathbb{E}\left(\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}|N_1\right)\right).$$

Conditionally on N_1, $T_{q+1}-T_q$ has a Beta $(1,N_1)$ distribution for $i<N_1$. Then,

$$\mathbb{E}\left(\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}|N_1\right)=\int_0^{2\Delta_n}N_1\left(1-t\right)^{N_1-1}dt=1-\left(1-2\Delta_n\right)^{N_1} \tag{3.40}$$

and (3.39) is dominated by

$$\mathbb{E}\left(\left(1-\left(1-2\Delta_n\right)^{N_1}\right)\mathbb{E}\left(\sum_{q=1}^{N_1}\mathbb{E}\left(\left(\Delta^n_{i(n,q)}X^c\right)^2|\mathcal{F}_{t_{i(n,q)}}\right)\right)\right).$$

and

$$\begin{aligned}\mathbb{E}\left(\left(\Delta^n_{i(n,q)}X^c\right)^2|\mathcal{F}_{t_{i(n,q)}}\right)&=\int_{t_{i(n,q)-1}}^{t_{i(n,q)}}\sigma_s^2ds\\&\leq\bar{\sigma}\Delta_n.\end{aligned}$$

Thus, (3.39) is dominated by

$$\bar{\sigma}\Delta_n\mathbb{E}\left(N_1\left(1-\left(1-2\Delta_n\right)^{N_1}\right)\right)\leq 2\bar{\sigma}\left(\lambda_n\Delta_n\right)^2$$

and (3.37) is equal to $O_p\left(\lambda_n\Delta_n\right)$. Concerning (3.38), it is dominated by

$$\sum_{q=1}^{N_1}|\Delta^n_{i(n,q)}Z^\beta|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}$$

and we have

$$\begin{aligned}|\Delta^n_{i(n,q)}Z^\beta|&=|-\left(1-e^{-\beta_n\Delta_n}\right)Z_{t_{i(n,q)-1}}+\sum_{j=1}^{\Delta^n_{i(n,q)}N}\Delta^j_{i(n,q)}Xe^{-\beta_n\left(t_{i(n,q)}-T^j_{i(n,q)}\right)}|\\&\leq\beta_n\Delta_n|Z_{t_{i(n,q)-1}}|+\sum_{j=1}^{\Delta^n_{i(n,q)}N}|\Delta X^j_{i(n,q)}|\end{aligned}$$

with ΔX^j_i the jth jump in $\left((i-1)\Delta_n,i\Delta_n\right]$ and T^j_i the time when it happens. We also have

$$\sum_{q=1}^{N_1}|Z^\beta_{t_{i(n,q)-1}}|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}$$

that is dominated by

$$\sum_{i=1}^{n}\int_0^{t_{i-1}}|x|e^{-\beta_n(t_{i-1}-t)}\underline{p}\left(dt,dx\right)\Delta_i^n N\left(\mathbf{1}_{\Delta_i^n N\geq 2}+\mathbf{1}_{\Delta_{i+1}^n N\geq 1}+\mathbf{1}_{\Delta_{i+2}^n N\geq 1}\right)$$

with the convention $\Delta_{n+1}^n N=\Delta_{n+2}^n N=0$. Because of the independence of $\Delta_i^n N$, $\Delta_{i+1}^n N$ and $\Delta_{i+2}^n N$ with $\mathcal{F}_{t_{i-1}}$, the expectation of the right hand side of this equation is equal to

$$\sum_{i=1}^{n}\mathbb{E}\left(\int_0^{t_{i-1}}|x|e^{-\beta_n(t_{i-1}-t)}\underline{p}\left(dt,dx\right)\right)\mathbb{E}\left(\Delta_i^n N\left(\mathbf{1}_{\Delta_i^n N\geq 2}+\mathbf{1}_{\Delta_{i+1}^n N\geq 1}+\mathbf{1}_{\Delta_{i+2}^n N\geq 1}\right)\right).$$

As

$$\mathbb{E}\left(\int_0^{t_{i-1}}|x|e^{-\beta_n(t_{i-1}-t)}\underline{p}\left(dt,dx\right)\right)=O\left(\frac{\lambda_n}{\beta_n}\right)$$

and

$$\mathbb{E}\left(\Delta_i^n N\left(\mathbf{1}_{\Delta_i^n N\geq 2}+\mathbf{1}_{\Delta_{i+1}^n N\geq 1}+\mathbf{1}_{\Delta_{i+2}^n N\geq 1}\right)\right)=O\left(\lambda_n^2\Delta_n^2\right),$$

we have

$$\sum_{i=1}^{n}\int_0^{t_{i-1}}|x|e^{-\beta_n(t_{i-1}-t)}\underline{p}\left(dt,dx\right)\Delta_i^n N\left(\mathbf{1}_{\Delta_i^n N\geq 2}+\mathbf{1}_{\Delta_{i+1}^n N\geq 1}+\mathbf{1}_{\Delta_{i+2}^n N\geq 1}\right)=O_p\left(\frac{\lambda_n^3\Delta_n}{\beta_n}\right).$$

Thus,

$$\sum_{q=1}^{N_1}|Z^{\beta}_{t_{i(n,q)-1}}|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}=O_p\left(\frac{\lambda_n^3\Delta_n}{\beta_n}\right). \tag{3.41}$$

We can also show that

$$\begin{aligned}\sum_{q=1}^{N_1}\sum_{j=1}^{\Delta_{i(n,q)}^n N}|\Delta X^j_{i(n,q)}|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}&\leq\sum_{i=1}^{n}\Delta_i^n N\sum_{j=1}^{\Delta_i^n N}|\Delta X_i^j|\left(\mathbf{1}_{\Delta_i^n N\geq 2}+\mathbf{1}_{\Delta_{i+1}^n N\geq 1}+\mathbf{1}_{\Delta_{i+2}^n N\geq 1}\right)\\&=O_p\left(\lambda_n^2\Delta_n\right).\end{aligned} \tag{3.42}$$

Combining (3.41) and (3.42), (3.38) is equal to

$$O_p\left(\lambda_n^2\Delta_n\right).$$

Summing the errors of (3.37) and (3.38), we find

$$\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X=\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X+O_p\left(\lambda_n^2\Delta_n\right). \tag{3.43}$$

We need to study the convergence of $\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X$ in order to study the one of the denominator. This sum is equal to the sum of

$$\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X^c \tag{3.44}$$

and

$$\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n Z^\beta. \tag{3.45}$$

(3.44) can be rewritten as

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{\Delta_i^n N} \operatorname{sgn}\left(\Delta X_i^j\right) \right) \Delta_i^n X^c \mathbf{1}_{\Delta_i^n N \geq 1}$$

which is equal to

$$\sum_{i=1}^{n} \operatorname{sgn}\left(\Delta X_i^1\right) \Delta_i^n X^c \mathbf{1}_{\Delta_i^n N=1} + \sum_{i=1}^{n} \left(\sum_{j=1}^{\Delta_i^n N} \operatorname{sgn}\left(\Delta X_i^j\right) \right) \Delta_i^n X^c \mathbf{1}_{\Delta_i^n N \geq 2} \tag{3.46}$$

As $\mathbb{E}\left(\Delta_i^n X^c\right) = 0$ and X^c is independent of $\underline{p}$, the left hand side of (3.46) has expectation 0 and conditional variance equal to

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{\Delta_i^n N=1} \left(\Delta_i^n X^c\right)^2 |\mathcal{F}_{i-1}\right) &= \sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{\Delta_i^n N=1} |\mathcal{F}_{i-1}\right) \mathbb{E}\left(\left(\Delta_i^n X^c\right)^2 |\mathcal{F}_{i-1}\right) \\
&= \lambda_n \Delta_n \sum_{i=1}^{n} \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^2 ds \\
&= \lambda_n \Delta_n \int_0^1 \sigma_s^2 ds.
\end{aligned}$$

Thus, the left hand side of (3.46) is equal to $O_p\left(\sqrt{\lambda_n \Delta_n}\right)$. In the same way, we show that the right hand side of (3.46) is equal to $O_p\left(\lambda_n \Delta_n\right)$. (3.45) is equal to

$$\begin{aligned}
&\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{\Delta_{i(n,q)}^n N} \Delta X_{i(n,q)}^j e^{-\beta_n \left(\left(\lfloor \frac{T_q}{\Delta_n} \rfloor + 1\right)\Delta_n - T_{i(n,q)}^j\right)} \\
&- \left(1 - e^{-\beta_n \Delta_n}\right) \sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) Z_{t_{i(n,q)-1}}.
\end{aligned} \tag{3.47}$$

The first term of (3.47) is equal to the sum of

$$\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta X_{T_q} e^{-\beta_n \left(\left(\lfloor \frac{T_q}{\Delta_n} \rfloor + 1\right)\Delta_n - T_q\right)}$$

and

$$\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{\substack{j=1, \\ T_{i(n,q)}^j \neq T_q}}^{\Delta_{i(n,q)}^n N} \Delta X_q^i e^{-\beta_n \left(\left(\lfloor \frac{T_q}{\Delta_n} \rfloor + 1\right)\Delta_n - T_{i(n,q)}^j\right)} \mathbf{1}_{\Delta_{i(n,q)}^n N \geq 2}$$

which is equal to

$$\int_0^1 \int_{\mathbb{R}} |x| e^{-\beta_n \left(\left(\lfloor \frac{t}{\Delta_n} \rfloor + 1\right)\Delta_n - t\right)} \underline{p}\left(dt, dx\right) + O_p\left(\lambda_n^2 \Delta_n\right).$$

The second term of (3.47) is dominated in absolute value by

$$\beta_n\Delta_n\sum_{q=1}^{N_1}\int_0^{t_{i(n,q)-1}} e^{-\beta_n\left(t_{i(n,q)-1}-t\right)}|x|\underline{p}\left(dt,dx\right)\leq\beta_n\Delta_n e^{\beta_n\Delta_n}\sum_{q=1}^{N_1}\int_0^{T_q^-} e^{-\beta_n\left(T_q^- -t\right)}|x|\underline{p}\left(dt,dx\right)$$

which is equal to

$$\beta_n\Delta_n e^{\beta_n\Delta_n}\int_0^1\int_0^t |y|e^{-\beta_n(t-s)}\underline{p}\left(ds,dy\right)\underline{p}\left(dt,dx\right).$$

This term has expectation equal to $O\left(\beta_n\Delta_n\frac{\lambda_n^2}{\beta_n}\right)$; the second term of (3.47) is then equal to $O_p\left(\lambda_n^2\Delta_n\right)$. Finally we have:

$$\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X=\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)+O_p\left(\sqrt{\lambda_n\Delta_n}+\lambda_n^2\Delta_n\right). \tag{3.48}$$

Combining (3.43) and (3.48), we have for the denominator

$$\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X=\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)+O_p\left(\lambda_n^2\Delta_n+\sqrt{\lambda_n\Delta_n}\right).$$

When $\lambda_n=O\left(1\right)$,

$$\frac{1}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}=O_p\left(1\right)$$

and when $\lambda_n\to\infty$, using $\int_0^1 e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}dt=\left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right)$,

$$\frac{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}{\lambda_n|x|\star\nu\left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right)}\to 1. \tag{3.49}$$

Thus, we have

$$\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}=1+O_p\left(\lambda_n\Delta_n+\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right). \tag{3.50}$$

Convergence of the jump process part (3.36)

According to (3.50), (3.36) is equal to

$$\left(1-e^{-\beta_n\Delta_n}\right)\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z_{t_{i(n,q)-1}}^{\beta}-\frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}\times$$
$$\left(1+O_p\left(\lambda_n\Delta_n+\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)\right).$$

As for the proof of (3.43), we show that the difference between the numerator

$$\sum_{q\in E_n} \operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^{\beta}_{t_{i(n,q)-1}} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right) \tag{3.51}$$

and

$$\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^{\beta}_{t_{i(n,q)-1}} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right)$$

is dominated by

$$\sum_{q=1}^{N_1}\left(|Z^{\beta}_{t_{i(n,q)-1}}| + \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}|\Delta X_{T_j}|\right)\mathbf{1}_{T_{q+1}-T_q<2\Delta_n}.$$

According to (3.41), we have

$$\sum_{q=1}^{N_1}|Z^{\beta}_{t_{i(n,q)-1}}|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n} = O_p\left(\frac{\lambda_n^3\Delta_n}{\beta_n}\right).$$

We can also show that

$$\sum_{q=1}^{N_1}\frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}|\Delta X_{T_j}|\mathbf{1}_{T_{q+1}-T_q<2\Delta_n} = O_p\left(\frac{\lambda_n^3\Delta_n}{\beta_n}\right).$$

Finally, we have that (3.51) is equal to

$$\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^{\beta}_{t_{i(n,q)-1}} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right) + O_p\left(\frac{\lambda_n^3\Delta_n}{\beta_n}\right)$$

and because the denominator is of order λ_n, (3.36) is equal to

$$\begin{aligned}\left(1-e^{-\beta_n\Delta_n}\right)&\left(\frac{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^{\beta}_{t_{i(n,q)-1}} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)} + O_p\left(\frac{\lambda_n}{\beta_n}\lambda_n\Delta_n\right)\right)\times \\ &\left(1+O_p\left(\lambda_n\Delta_n+\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)\right).\end{aligned} \tag{3.52}$$

Thus, let us study the convergence of

$$\frac{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(Z^{\beta}_{t_{i(n,q)-1}} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\sum_{j=1}^{q-1}\Delta X_{T_j}\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\lfloor+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}. \tag{3.53}$$

As $Z^{\beta}_{t_{i(n,q)-1}} = e^{-\beta_n\left(t_{i(n,q)-1}-T_q\right)} Z^{\beta}_{T_{q^-}}$ when there is only one jump in $(i(n,q)-1, i(n,q)]$, the numerator of (3.53) is equal to

$$\int_{[0,1]^2\times\mathbb{R}^2} y\mathrm{sgn}(x)\,\mathbf{1}_{t>s}\left(e^{-\beta_n(t-s)}e^{-\beta_n\left(\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n-t\right)} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)\underline{p}(ds,dy)\,\underline{p}(dt,dx) \tag{3.54}$$
$$+ O_p\left(\frac{\lambda_n}{\beta_n}\lambda_n^2\Delta_n\right).$$

The term $O_p\left(\frac{\lambda_n}{\beta_n}\lambda_n^2\Delta_n\right)$ corresponds to the case with more than one jump. Using Fubini's theorem for Poisson measures, we can write the left hand side of (3.54) as the sum of

$$M_{1,n},$$

$$\lambda_n\int_0^1\int_{\mathbb{R}} g_{1,n}(t,x)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)$$

and

$$\int_{[0,1]^2\times\mathbb{R}^2} f_n(t,x,s,y)\left(\underline{p}-\lambda_n\underline{q}\right)(ds,dy)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)$$

with

$$\begin{aligned}M_{1,n} &= \lambda_n^2 x\star\nu\,\mathrm{sgn}(x)\star\nu\int_{[0,1]^2}\mathbf{1}_{t>s}\left(e^{-\beta_n(t-s)}e^{-\beta_n\left(\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n-t\right)} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)dtds\\ &= \frac{\lambda_n^2}{\beta_n}x\star\nu\,\mathrm{sgn}(x)\star\nu\left(\frac{\left(e^{\beta_n\Delta_n}-1\right)}{\beta_n\Delta_n} - \Delta_n\frac{1-e^{-\beta_n}}{1-e^{-\beta_n\Delta_n}} - \frac{\beta_n\Delta_n}{1-e^{-\beta_n\Delta_n}}\right),\end{aligned}$$

$$\begin{aligned}g_{1,n}(t,x) &= x\mathrm{sgn}(y)\star\nu\int_{[0,1]}\mathbf{1}_{s>t}\left(e^{-\beta_n(s-t)}e^{-\beta_n\left(\lfloor\frac{s}{\Delta_n}\rfloor\Delta_n-s\right)} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)ds\\ &\quad + y\star\nu\mathrm{sgn}(x)\int_{[0,1]}\mathbf{1}_{t>s}\left(e^{-\beta_n(t-s)}e^{-\beta_n\left(\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n-t\right)} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)ds\end{aligned}$$

and

$$f_n(t,x,s,y) = y\mathrm{sgn}(x)\,\mathbf{1}_{t>s}\left(e^{-\beta_n(t-s)}e^{-\beta_n\left(\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n-t\right)} - \frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right).$$

(3.53) is then equal to the sum of

$$M_{2,n} = \frac{M_{1,n}}{\lambda_n|x|\star\nu\left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right)},$$

$$\frac{\left(\lambda_n\int_0^1\int_{\mathbb{R}} g_n(t,x)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx) + \int_{[0,1]^2\times\mathbb{R}^2} f_n(t,x,s,y)\left(\underline{p}-\lambda_n\underline{q}\right)(ds,dy)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)}\underline{p}(dt,dx)}$$

with

$$g_n(t,x) = -\frac{M_{2,n}}{\lambda_n}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}\rfloor+1\right)\Delta_n-t\right)} + g_{1,n}(t,x)$$

and

$$O_p\left(\frac{\lambda_n}{\beta_n}\lambda_n\Delta_n\right).$$

The term $\int_0^1\int_{\mathbb{R}} g_n(t,x)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)$ has expectation 0 and variance

$$\lambda_n\int_0^1\int_{\mathbb{R}} g_n^2(t,x)\,\nu(dx)\,dt$$

which is equal to

$$\frac{\lambda_n}{3\beta_n^2}\left((\mathrm{sgn}(x)\star\nu)^2|x|^2\star\nu+(x\star\nu)^2-2\mathrm{sgn}(x)\star\nu|x|^2\star\nu\right)+O\left(\frac{\lambda_n\Delta_n}{\beta_n}+\frac{\lambda_n}{\beta_n^4}\right).$$

Remark 3.1. *The computation of the variance is not detailed because it is technical in term of calculus and does not present any interest.*

Thus, we have

$$\frac{\lambda_n\int_0^1\int_{\mathbb{R}} g_n(t,x)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)}{\lambda_n|x|\star\nu}=V_{g,n}J_{1,s,n}$$

with

$$J_{1,s,n}=O_p(1)$$

and

$$V_{g,n}=\frac{\sqrt{\lambda_n}}{\sqrt{3}\beta_n|x|\star\nu}\sqrt{(\mathrm{sgn}(x)\star\nu)^2|x|^2\star\nu+(x\star\nu)^2-2\mathrm{sgn}(x)\star\nu|x|^2\star\nu}$$
$$+O\left(\frac{\sqrt{\lambda_n\Delta_n}}{\sqrt{\beta_n}}+\frac{\sqrt{\lambda_n}}{\beta_n^2}\right).$$

Remark 3.2. *When $\beta_n\asymp\Delta_n$, $\frac{\sqrt{\lambda_n\Delta_n}}{\sqrt{\beta_n}}\asymp\frac{\sqrt{\lambda_n}}{\beta_n}$ and it is of the same order than the principal term. However, in this case, $\frac{\sqrt{\lambda_n}}{\beta_n}=o\left(\frac{1}{\sqrt{\beta_n}}\right)$ and we will see that $\frac{1}{\sqrt{\beta_n}}$ is the order of*

$$\frac{\int_0^1\int_{\mathbb{R}}\int_0^t\int_{\mathbb{R}} f_n(t,x,s,y)\left(\underline{p}-\lambda_n\underline{q}\right)(ds,dy)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)}{\lambda_n}.$$

The term $\int_0^1\int_0^t\int_{\mathbb{R}^2} f_n(t,x,s,y)\left(\underline{p}-\lambda_n\underline{q}\right)(ds,dy)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)$ has expectation 0 and variance equal to

$$\lambda_n^2\int_0^1\int_0^1\int_{\mathbb{R}^2} f_n^2(t,x,s,y)\,\nu(dy)\,ds\nu(dx)\,dt$$

which is equal to

$$\lambda_n^2|x|^2\star\nu\int_{[0,1]^2}\mathbf{1}_{t>s}\left(e^{-\beta_n(t-s)}e^{-\beta_n\left(\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n-t\right)}-\left(\frac{2\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)\right)^2dtds$$

and which is equivalent to

$$|x|^2\star\nu\frac{\lambda_n^2}{2\beta_n}\frac{\left(e^{2\beta_n\Delta_n}-1\right)}{2\beta_n\Delta_n}.$$

Furthermore, we can show that

$$\mathbb{E}\left(\left|\int_0^1\int_{\mathbb{R}}\int_0^t\int_{\mathbb{R}} f_n\left(t,x,s,y\right)\left(\underline{p}-\lambda_n\underline{q}\right)\left(ds,dy\right)\left(\underline{p}-\lambda_n\underline{q}\right)\left(dt,dx\right)\right|\right)=O\left(\frac{\lambda_n^2}{\beta_n}\right).$$

As we have (3.49), we can finally write (3.53) as

$$M_{2,n}+\frac{\beta_n\Delta_n}{\left(1-e^{-\beta_n\Delta_n}\right)}V_gJ_{1,s,n}+\frac{\beta_n\Delta_n}{\left(1-e^{-\beta_n\Delta_n}\right)}\min\left(\sqrt{\frac{|x|^2\star\nu}{\left(|x|\star\nu\right)^2}\frac{1}{2\beta_n}\frac{\left(e^{2\beta_n\Delta_n}-1\right)}{2\beta_n\Delta_n}},\frac{\lambda_n}{\beta_n}\right)J_{2,s,n} \tag{3.55}$$

with

$$J_{2,s,n}=O_p\left(1\right).$$

Central limit theorem To have a central limit theorem, we will use the results of [23, Theorem 3]. Let us suppose that $|x|^3\star\nu<\infty$ assuring the existence of $\int_0^1\int_{\mathbb{R}}g_n^3\left(t,x\right)\nu\left(dx\right)dt$. We also suppose that

$$\left(\operatorname{sgn}\left(x\right)\star\nu\right)^2|x|^2\star\nu+\left(x\star\nu\right)^2-2\operatorname{sgn}\left(x\right)\star\nu|x|^2\star\nu\neq 0.$$

In the case where it is equal to 0, the order of the term before $J_{1,s,n}$ will be negligible compared to the one before $J_{2,s,n}$ and we do not need a central limit theorem of $J_{1,s,n}$. To have a central limit theorem on $J_{1,s,n}$, according to [23, Theorem 3], we need:

$$\frac{\lambda_n\int_0^1\int_{\mathbb{R}}g_n^3\left(t,x\right)\nu\left(dx\right)dt}{\left(\lambda_n\int_0^1\int_{\mathbb{R}}g_n^2\left(t,x\right)\nu\left(dx\right)dt\right)^{\frac{3}{2}}}\to 0. \tag{3.56}$$

If this condition is fulfilled, we have $J_{1,s,n}\to\mathcal{N}\left(0,1\right)$. We have

$$\lambda_n\int_0^1\int_{\mathbb{R}}g_n^2\left(t,x\right)\nu\left(dx\right)dt\asymp\frac{\lambda_n}{\beta_n^2}. \tag{3.57}$$

We also have, separating the case $\beta_n\Delta_n\asymp 1$ and $\beta_n\Delta_n=o\left(1\right)$, that $|g_n\left(t,x\right)|$ is bounded by

$$e^{\beta\Delta_n}\left(|x|\int_{[0,1]}\mathbf{1}_{s>t}\left(e^{-\beta_n(s-t)}+\frac{\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)ds+\int_{[0,1]}\mathbf{1}_{s>t}\left(e^{-\beta_n(s-t)}+\frac{\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)ds\right)$$

and then by $\max\left(|x|,1\right)O\left(\frac{1}{\beta_n}\right)$. Thus, the left hand side of (3.56) is bounded by:

$$\frac{\lambda_n\int_0^1\int_{\mathbb{R}}max\left(|x|,1\right)g_n^2\left(t,x\right)\nu\left(dx\right)dt}{\left(\lambda_n\int_0^1\int_{\mathbb{R}}g_n^2\left(t,x\right)\nu\left(dx\right)dt\right)^{\frac{3}{2}}}O\left(\frac{1}{\beta_n}\right)$$

which is equal, using (3.57), to $O\left(\frac{1}{\lambda_n}\right)$ and that converges to 0 when $\lambda_n\to\infty$. Thus, (3.56) is verified if $\lambda_n\to\infty$ and in this case we have

$$J_{1,s,n}\to\mathcal{N}\left(0,1\right).$$

Let us suppose that $|x|^4 \star \nu < \infty$ assuring the existence of $\int_0^1 \int_0^1 \int_{\mathbb{R}^2} f_n^4(t,x,s,y)\,\nu(dy)\,ds\nu(dx)\,dt$. To have a central limit theorem on $\int_{[0,1]^2\times\mathbb{R}^2} f_n(t,x,s,y)\left(\underline{p}-\lambda_n\underline{q}\right)(ds,dy)\left(\underline{p}-\lambda_n\underline{q}\right)(dt,dx)$, according to [23, Theorem 3], we need:

$$\frac{\lambda_n^2 \int_0^1 \int_0^1 \int_{\mathbb{R}^2} f_n^4(t,x,s,y)\,\nu(dy)\,ds\nu(dx)\,dt}{\left(\lambda_n^2 \int_0^1 \int_0^1 \int_{\mathbb{R}^2} f_n^2(t,x,s,y)\,\nu(dy)\,ds\nu(dx)\,dt\right)^2} \to 0. \tag{3.58}$$

We can show that

$$\int_0^1 \int_0^1 \int_{\mathbb{R}^2} f_n^4(t,x,s,y)\,\nu(dy)\,ds\nu(dx)\,dt \asymp \frac{1}{4\beta_n}\frac{\left(e^{4\beta_n\Delta_n}-1\right)}{4\beta_n\Delta_n}.$$

Thus, (3.58) is verified if $\frac{\beta_n}{\lambda_n^2} \to 0$ and in this case

$$\min\left(\sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(e^{2\beta_n\Delta_n}-1)}{2\beta_n\Delta_n}},\frac{\lambda_n}{\beta_n}\right) = \sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(e^{2\beta_n\Delta_n}-1)}{2\beta_n\Delta_n}}$$

and we have

$$\left(J_{1,s,n}, J_{2,s,n}\right)^T \to \mathcal{N}\left(0, \mathrm{Id}_2\right).$$

Finally, combining (3.52) and (3.53), we have (3.36) equal to

$$\beta_n\Delta_n M_{s,n} + \beta_n\Delta_n V_g J_{1,s,n} + \beta_n\Delta_n \min\left(\sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(e^{2\beta_n\Delta_n}-1)}{2\beta_n\Delta_n}},\frac{\lambda_n}{\beta_n}\right) J_{2,s,n} \tag{3.59}$$

with

$$\begin{aligned} M_{s,n} &= \left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right) M_{2,n} \\ &= \frac{\lambda_n}{\beta_n}\frac{x\star\nu\,\mathrm{sgn}(x)\star\nu}{|x|\star\nu}\left(\frac{\left(e^{\beta_n\Delta_n}-1\right)}{\beta_n\Delta_n} - \frac{\beta_n\Delta_n}{1-e^{-\beta_n\Delta_n}} - \Delta_n\frac{1-e^{-\beta_n}}{1-e^{-\beta_n\Delta_n}}\right). \end{aligned}$$

Convergence of the Brownian part (3.34)

As for the proof of the convergence of the denominator, we show that the numerator of (3.34) is equal to

$$\sum_{q=1}^{N_1} \Delta_{i(n,q)+1}^n X^c \mathrm{sgn}\left(\Delta X_{T_q}\right) + O_p\left(\lambda_n\Delta_n\right)$$

and (3.34) is equal to, using (3.50),

$$\frac{\sum_{q=1}^{N_1}\Delta_{i(n,q)+1}^n X^c\mathrm{sgn}\left(\Delta X_{T_q}\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}\underline{p}(dt,dx)}\left(1+O_p\left(\lambda_n\Delta_n+\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)\right)+O_p\left(\Delta_n\right).$$

Let us study the convergence of

$$\frac{\sum_{q=1}^{N_1}\Delta_{i(n,q)+1}^n X^c\mathrm{sgn}\left(\Delta X_{T_q}\right)}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}\underline{p}(dt,dx)}. \tag{3.60}$$

We can write the numerator of (3.60) as

$$\sum_{i=2}^{n} \operatorname{sgn}\left(\Delta X_{i-1}^{1}\right) \Delta_i^n X^c \mathbf{1}_{\Delta_{i-1}^n N \geq 1} + \sum_{i=2}^{n} \left(\sum_{j=2}^{\Delta_{i-1}^n N} \operatorname{sgn}\left(\Delta X_i^j\right) \right) \Delta_i^n X^c \mathbf{1}_{\Delta_i^n N \geq 2} \tag{3.61}$$

with ΔX_i^j the jth jump in $((i-1)\Delta_n, i\Delta_n]$. Let us study the first term of the sum of (3.61). We have:

$$\begin{aligned}\mathbb{E}\left(\Delta_i^n X^c \mathbf{1}_{\Delta_{i-1}^n N \geq 1} \operatorname{sgn}\left(\Delta X_{i-1}^1\right) | \mathcal{F}_{i-1}\right) &= \mathbf{1}_{\Delta_{i-1}^n N \geq 1} \operatorname{sgn}\left(\Delta X_{i-1}^1\right) \mathbb{E}\left(\Delta_i^n X^c | \mathcal{F}_{i-1}\right) \\ &= 0\end{aligned}$$

as X^c is a $\mathcal{F}$ martingale. We also have

$$\begin{aligned}V_n^2 &= \sum_{i=2}^{n} \mathbb{E}\left(\left(\Delta_i^n X^c\right)^2 \mathbf{1}_{\Delta_{i-1}^n N \geq 1} | \mathcal{F}_{i-1}\right) \\ &= \sum_{i=2}^{n} \mathbf{1}_{\Delta_{i-1}^n N \geq 1} \mathbb{E}\left(\left(\Delta_i^n X^c\right)^2 | \mathcal{F}_{i-1}\right) \\ &= \sum_{i=2}^{n} \mathbf{1}_{\Delta_{i-1}^n N \geq 1} \int_{t_{i-1}}^{t_i} \sigma_s^2 ds.\end{aligned}$$

The expectation of $\sum_{i=2}^{n} \operatorname{sgn}\left(\Delta X_{i-1}^1\right) \Delta_i^n X^c \mathbf{1}_{\Delta_{i-1}^n N \geq 1}$ is then equal to 0 and its variance to $\mathbb{E}\left(V_n^2\right) = \lambda_n \Delta_n \int_{\Delta_n}^{1} \sigma_s^2 ds + O\left(\left(\lambda_n \Delta_n\right)^2\right)$. We can show that the second term of the sum of (3.61) is equal to $O_p\left(\lambda_n \Delta_n\right)$ which is negligible. When $\lambda_n = O(1)$, (3.60) is equal to $O_p\left(\sqrt{\Delta_n}\right) = O_p\left(\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)$ and when $\lambda_n \to \infty$, $\int_0^1 \int_{\mathbb{R}} |x| e^{-\beta_n\left(\left(\lfloor \frac{t}{\Delta_n} + 1\right)\Delta_n - t\right)} \underline{p}(dt, dx)$ is equivalent to $\lambda_n \Delta_n |x| \star \nu$ and (3.60) is equal to $O_p\left(\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)$. Finally, we then can write (3.34) as

$$\frac{\sqrt{\int_0^1 \sigma_s^2 ds} \sqrt{\Delta_n}}{|x| \star \nu \sqrt{\lambda_n}} J_{3,s,n} \tag{3.62}$$

with

$$J_{3,s,n} = O_p(1).$$

Central limit theorem To have a central limit theorem, we need a Lindeberg condition. Let $\eta > 0$. As we work on the set $\{N_1 > 0\}$, V_n is strictly positice. We have:

$$\begin{aligned}&\sum_{i=2}^{n} \mathbb{E}\left(\frac{\left(\Delta_i^n X^c\right)^2 \mathbf{1}_{\Delta_{i-1}^n N \geq 1}}{V_n^2} \mathbf{1}_{\frac{|\Delta_i^n X^c| \mathbf{1}_{\Delta_{i-1}^n N \geq 1}}{V_n} \geq \eta}\right) \\ &= \sum_{i=2}^{n} \mathbb{E}\left(\frac{\left(\Delta_i^n X^c\right)^2}{V_n^2} \mathbf{1}_{\Delta_{i-1}^n N \geq 1} \mathbf{1}_{\frac{|\Delta_i^n X^c|}{V_n} \geq \eta}\right) \\ &= \sum_{i=2}^{n} \mathbb{E}\left(\frac{\mathbf{1}_{\Delta_{i-1}^n N \geq 1}}{V_n^2} \mathbb{E}\left(\left(\Delta_i^n X^c\right)^2 \mathbf{1}_{\frac{|\Delta_i^n X^c|}{V_n} \geq \eta} | \sigma\left(\left(N_s\right)_{0 \leq s \leq 1}\right)\right)\right).\end{aligned} \tag{3.63}$$

Using Cauchy Schwartz inequality, we find that

$$\mathbb{E}\left((\Delta_i^n X^c)^2 \mathbf{1}_{\frac{|\Delta_i^n X^c|}{V_n}\geq\eta}|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right) \leq \mathbb{E}\left((\Delta_i^n X^c)^4\right)^{\frac{1}{2}} \mathbb{E}\left(\mathbf{1}_{\frac{|\Delta_i^n X^c|}{V_n}\geq\eta}|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right)^{\frac{1}{2}}$$
$$= \mathbb{E}\left((\Delta_i^n X^c)^4\right)^{\frac{1}{2}} \mathbb{P}\left(\frac{|\Delta_i^n X^c|}{V_n}\geq\eta|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right)^{\frac{1}{2}}.$$

We can dominate the term $\mathbb{P}\left(\frac{|\Delta_i^n X^c|}{V_n}\geq\eta|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right)$ using Markov inequality:

$$\mathbb{P}\left(\frac{|\Delta_i^n X^c|}{V_n}\geq\eta|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right) \leq \frac{\mathbb{E}\left((\Delta_i^n X^c)^2\right)}{\eta^2 V_n^2}$$
$$= \frac{\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_s^2 ds}{\eta^2 V_n^2}$$
$$\leq \frac{\bar{\sigma}^2\Delta_n}{\eta^2 V_n^2}.$$

We also can dominate the term $\mathbb{E}\left((\Delta_i^n X^c)^4\right)$ using the Burkholder-Davis-Gundy inequality:

$$\mathbb{E}\left((\Delta_i^n X^c)^4\right) \leq C_2\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_s^2 ds\right)^2$$
$$\leq C_2\bar{\sigma}^4\Delta_n^2.$$

Thus,

$$\mathbb{E}\left((\Delta_i^n X^c)^2 \mathbf{1}_{\frac{|\Delta_i^n X^c|}{V_n}\geq\eta}|\sigma\left((N_s)_{0\leq s\leq 1}\right)\right) \leq \frac{\sqrt{C_2}\bar{\sigma}^3\Delta_n^{\frac{3}{2}}}{\eta V_n}$$

and (3.63) is dominated by

$$\sum_{i=2}^{n}\frac{\sqrt{C_2}\bar{\sigma}^3\Delta_n^{\frac{3}{2}}}{\eta}\mathbb{E}\left(\frac{\mathbf{1}_{\Delta_{i-1}^n N}\geq 1}{V_n}\right)$$

Furthermore,

$$V_n^2 \geq \underline{\sigma}^2\Delta_n N_{1-\Delta_n}. \tag{3.64}$$

Then, (3.63) is dominated by

$$\frac{\sqrt{C_2}\bar{\sigma}^3\Delta_n^{\frac{1}{2}}e^{\frac{\lambda_n\Delta_n}{2}}}{\eta\underline{\sigma}}\mathbb{E}\left(\sqrt{N_{1-\Delta_n}}\right). \tag{3.65}$$

Using Jensen inequality,

$$\mathbb{E}\left(\sqrt{N_{1-\Delta_n}}\right) \leq \sqrt{\mathbb{E}\left(N_{1-\Delta_n}\right)} = \sqrt{\lambda_n\left(1-\Delta_n\right)}.$$

Thus (3.65) is dominated by

$$\frac{\sqrt{C_2}\bar{\sigma}^3}{\eta\underline{\sigma}}\sqrt{\lambda_n\Delta_n}$$

that converges to 0. Thus, the Lindeberg condition is verified and

$$\frac{1}{V_n}\sum_{i=2}^{n}\Delta_i^n X^c \operatorname{sgn}\left(\Delta_j^1 X\right)\mathbf{1}_{\Delta_{i-1}^n N=1} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

We want to prove that

$$\frac{V_n^2}{\lambda_n\Delta_n} \xrightarrow{p} \int_0^1 \sigma_s^2 ds. \tag{3.66}$$

As $\frac{1}{\lambda_n}\int_0^1 \sigma_s^2 dN_s \xrightarrow{p} \int_0^1 \sigma_s^2 ds$, it is sufficient to show that $|\frac{V_n^2}{\lambda_n\Delta_n} - \frac{1}{\lambda_n}\int_0^1 \sigma_s^2 dN_s| \xrightarrow{p} 0$.

$$V_n^2 = \sum_{q=1}^{N_{1-\Delta_n}} \int_{t_{i(n,q)}}^{t_{i(n,q)+1}} \sigma_s^2$$

and we have:

$$|\frac{V_n^2}{\lambda_n\Delta_n} - \frac{1}{\lambda_n}\int_0^1 \sigma_s^2 dN_s| \leq \frac{1}{\lambda_n\Delta_n}\sum_{q=1}^{N_{1-\Delta_n}} \int_{t_{i(n,q)}}^{t_{i(n,q)+1}} |\sigma_s^2 - \sigma_{T_q}^2| ds$$

The expectation of the right hand side is equal to

$$\frac{1}{\Delta_n}\int_0^{1-\Delta_n}\int_{\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n+\Delta_n}^{\lfloor\frac{t}{\Delta_n}\rfloor\Delta_n+2\Delta_n} |\sigma_s - \sigma_t| ds dt$$

and converges to 0 as σ^2 is continuous on the compact $[0,1]$. (3.66) is then verified and

$$\frac{1}{\sqrt{\lambda_n\Delta_n\int_0^1\sigma_s^2 ds}}\sum_{q=1}^{N_1}\Delta_{i(n,q)+1}^n X^c \operatorname{sgn}\left(\Delta X_{T_q}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

We finally have:

$$J_{3,s,n} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

Convergence of (3.35)

(3.35) is equal to

$$\begin{aligned}
&= 1 - e^{-\beta_n\Delta_n} - \left(1 - e^{-\beta_n\Delta_n}\right)\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X^c}{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(\Delta_{i(n,q)}^n Z^\beta + \Delta_{i(n,q)}^n X^c\right)} \\
&= 1 - e^{-\beta_n\Delta_n} - \left(1 - e^{-\beta_n\Delta_n}\right)\frac{\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X^c}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n - t\right)}\underline{p}\left(dt,dx\right)}\left(1 + O_p\left(\lambda_n\Delta_n + \frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)\right)
\end{aligned}$$

according to (3.50) and using the same proof than for the convergence of the denominator on the Brownian part, we have

$$\sum_{q\in E_n}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta_{i(n,q)}^n X^c = \sum_{q=1}^{N_1}\Delta_{i(n,q)+1}^n X^c + O_p\left(\lambda_n\Delta_n\right).$$

Thus, (3.35) is equal to

$$1-e^{-\beta_n\Delta_n}-\left(1-e^{-\beta_n\Delta_n}\right)\frac{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X^c}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}\left(1+O_p\left(\lambda_n\Delta_n+\frac{\sqrt{\Delta_n}}{\sqrt{\lambda_n}}\right)\right)$$
$$+O_p\left(\beta_n\Delta_n\Delta_n\right).$$

We then are interested in the convergence of

$$\frac{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X^c}{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\left(\lfloor\frac{t}{\Delta_n}+1\right)\Delta_n-t\right)}\underline{p}\left(dt,dx\right)}. \tag{3.67}$$

The numerator is equal to (3.44) and we have seen that it is equal to

$$\sum_{i=1}^n\operatorname{sgn}\left(\Delta X_i^1\right)\Delta_i^nX^c\mathbf{1}_{\Delta_i^nN\geq1}+O_p\left(\lambda_n\Delta_n\right). \tag{3.68}$$

$\sum_{i=1}^n\Delta_i^nX^c\mathbf{1}_{\Delta_i^nN\geq1}$ is a sum of martingale with conditional variance equal to

$$\lambda_n\Delta_n\int_0^1\sigma_s^2ds.$$

As for the convergence of (3.34), the Lindeberg condition is verified when $\lambda_n\to\infty$. We then have (3.67) which is equal to

$$\frac{\sqrt{\int_0^1\sigma_s^2ds}\sqrt{\Delta_n}}{|x|\star\nu\sqrt{\lambda_n}}J_{4,s,n}$$

with

$$J_{4,s,n}=O_p\left(1\right)$$

and

$$J_{4,s,n}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,1\right)$$

when $\lambda_n\to\infty$.
Concerning the convergence of $(J_{3,s,n},J_{4,s,n})$ we need to study

$$\sum_{i=2}^n\mathbb{E}\left(\left(\Delta_i^nX^c\right)^2\mathbf{1}_{\Delta_{i-1}^nN\geq1,\Delta_i^nN\geq1}\mathcal{F}_{i-1}\right)$$

which is equal to

$$\left(\lambda_n\Delta_n+O\left(\left(\lambda_n\Delta_n\right)^2\right)\right)\sum_{i=2}^n\mathbf{1}_{\Delta_{i-1}^nN\geq1}\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_s^2ds.$$

This term is negligible compared the term of variances and then, when $\lambda_n\to\infty$,

$$\left(J_{3,s,n},J_{4,s,n}\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\mathrm{Id}_2\right).$$

We also show, when $\frac{\beta_n}{\lambda_n^2}\to0$, that

$$\left(J_{1,s,n},J_{2,s,n},J_{3,s,n},J_{4,s,n}\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\mathrm{Id}_3\right).$$

Finally, (3.35) is equal to

$$1-e^{-\beta_n\Delta_n}+\left(1-e^{-\beta_n\Delta_n}\right)\frac{\sqrt{\int_0^1\sigma_s^2ds}\sqrt{\Delta_n}}{\sqrt{\lambda_n}|x|\star\nu}J_{4,s,n}. \tag{3.69}$$

Combining (3.59), (3.62) and (3.69) we have

$$\widehat{s}_n^c=\left(1-e^{-\beta_n\Delta_n}\right)+\beta_n\Delta_nM_{s,n}+\beta_n\Delta_nV_{s,n}^TJ_{s,n}$$

with

$$V_s=\left(V_{1,s,n},V_{2,s,n},V_{3,s,n}\right)^T,$$

$$\begin{cases} V_{1,s,n} &= V_{g,n}\\ V_{2,s,n} &= \min\left(\sqrt{\frac{|x|^2\star\nu}{(|x|\star\nu)^2}\frac{1}{2\beta_n}\frac{(e^{2\beta_n\Delta_n}-1)}{2\beta_n\Delta_n}},\frac{\lambda_n}{\beta_n}\right)\\ V_{3,s,n} &= \frac{\sqrt{\int_0^1\sigma_s^2ds}}{|x|\star\nu\sqrt{\lambda_n}\beta_n\sqrt{\Delta_n}}\\ V_{4,s,n} &= \left(\frac{1-e^{-\beta_n\Delta_n}}{\beta_n\Delta_n}\right)\frac{\sqrt{\int_0^1\sigma_s^2ds}\sqrt{\Delta_n}}{|x|\star\nu\sqrt{\lambda_n}}\end{cases}$$

and

$$J_{s,n}=\left(J_{1,s,n},J_{2,s,n},J_{3,s,n},J_{4,s,n}\right)^T.$$

Let

$$\widehat{\beta}_n^c=-\frac{1}{\Delta_n}\log\left(\left(1-\widehat{s}_n^c\right)\vee\Delta_n\right).$$

For n enough large, if $\frac{1}{\sqrt{\lambda_n\Delta_n}\beta_n}\to 0$, $\widehat{\beta}_n^c=-\frac{1}{\Delta_n}\log\left(1-\widehat{s}_n^c\right)$ and using Taylor's expansion, we can write

$$\widehat{\beta}_n^c=\beta_n+\beta_ne^{\beta_n\Delta_n}\left(M_{s,n}+V_{s,n}^TJ_n\right)$$

with J_n a vector of size four having the same asymptotic properties than $J_{s,n}$. We set $M_{c,n}=e^{\beta_n\Delta_n}M_{s,n}$ and $V_n=e^{\beta_n\Delta_n}V_{s,n}$. We have

$$M_{c,n}=e^{\beta_n\Delta_n}\frac{\lambda_n}{\beta_n}\frac{x\star\nu\,\mathrm{sgn}\left(x\right)\star\nu}{|x|\star\nu}\left(\frac{\left(e^{\beta_n\Delta_n}-1\right)}{\beta_n\Delta_n}-\frac{\beta_n\Delta_n}{1-e^{-\beta_n\Delta_n}}\right)+O\left(\frac{\lambda_n}{\beta_n^2}\right)$$

with

$$\frac{\lambda_n}{\beta_n^2}=o\left(V_{2,n}\right)$$

so we can omit this term in $M_{c,n}$. We also have

$$\begin{aligned}V_{1,n}=&e^{\beta_n\Delta_n}\frac{\sqrt{\lambda_n}}{\sqrt{3}\beta_n|x|\star\nu}\sqrt{\left(\mathrm{sgn}\left(x\right)\star\nu\right)^2|x|^2\star\nu+\left(x\star\nu\right)^2-2\mathrm{sgn}\left(x\right)\star\nu|x|^2\star\nu}\\&+O\left(\frac{\sqrt{\lambda_n}}{\beta_n^2}+\frac{\sqrt{\lambda_n\Delta_n}}{\sqrt{\beta_n}}\right)\end{aligned}$$

with

$$O\left(\frac{\sqrt{\lambda_n}}{\beta_n^2}+\frac{\sqrt{\lambda_n\Delta_n}}{\sqrt{\beta_n}}\right)=o\left(V_{2,n}+V_{3,n}\right);$$

this term can also be omitted. The proof of Proposition 3.3 is achieved.

5.3 Proof of Theorem 3.1

As for the proof of Theorem 3.3, we consider the case where the volatility is deterministic. The generalization to the stochastic case can be easily done.

First, in order to prove Theorem 3.1, we first need Lemma 3.1.

Lemma 3.1. *We have*

$$\mathbb{P}\left(\operatorname{sgn}\left(\Delta X_{T_q}\right)=\operatorname{sgn}\left(\Delta^n_{i(n,q)}X\right)\ \forall q\in\{1,...,N_1\}\right)\to 1. \tag{3.70}$$

Proof We work on $\Omega_n\cap\{\sup_{|t-s|<2\Delta_n}|N_t-N_s|=1\}$ with probability converging to 1 if $\lambda_n^2\Delta_n\to 0$. We first consider the case where Assumption 3.3 (I) is verified. In this case, $\beta_n\Delta_n=o(1)$ and we have to prove that

$$\mathbb{P}\left(\inf_{q\in\{1,...,N_1\}}\frac{\Delta^n_{i(n,q)}X}{\Delta X_{T_q}}\leq 0\right)\to 0. \tag{3.71}$$

For all $q\geq 1$, we have:

$$\begin{aligned}\frac{\Delta^n_{i(n,q)}X}{\Delta X_{T_q}}&=1+\frac{\Delta^n_{i(n,q)}X^c}{\Delta X_{T_q}}-\frac{\left(1-e^{-\beta_n\Delta_n}\right)Z^\beta_{i(n,q)-1}}{\Delta X_{T_q}}\\&\geq 1-\frac{|\Delta^n_{i(n,q)}X^c|}{|\Delta X_{T_q}|}-\frac{\beta_n\Delta_n|Z^\beta_{i(n,q)-1}|}{|\Delta X_{T_q}|}\\&\geq 1-\frac{\sup_{i\in\mathcal{A}_n^c}|\Delta^n_iX^c|}{\inf_{i\in\{1,...,N_1\}}|\Delta X_{T_i}|}-\frac{\beta_n\Delta_n\sup_{i\in\{1,...,n\}}|Z^\beta_{t_i}|}{\inf_{i\in\{1,...,N_1\}}|\Delta X_{T_i}|}.\end{aligned}$$

Thus, the right hand side of (3.71) is dominated by

$$\mathbb{P}\left(\sup_{i\in\mathcal{A}_n^c}|\Delta^n_iX^c|+\beta\Delta_n\sup_{i\in\{1,...,n\}}|Z^\beta_{t_i}|\geq\inf_{i\in\{1,...,N_1\}}|\Delta X_{T_i}|\right)$$

which is dominated by the sum of (3.19) and (3.20), both converging to 0.

The case where Assumption 3.3 (II) corresponds to (3.31) which is already proved. □

Combining the results of Proposition 3.1, Proposition 3.3 and Lemma 3.1, we can now prove Theorem 3.1.

Let

$$\widehat{s}_n=-\frac{\sum_{q=1}^{\widehat{\lambda}_n}\operatorname{sgn}(\Delta^n_{I(n,q)}X)\left(\Delta^n_{I(n,q)+1}X-2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{I(n,j)}X\right)}{\sum_{q=1}^{\widehat{R}_n}|\Delta^n_{I(n,q)}X|}\mathbf{1}_{\widehat{R}_n>0}$$

which is an estimator of the slope of Z^β.

We work on the set $\Omega_n\cap\{\operatorname{sgn}\left(\Delta X_{T_q}\right)=\operatorname{sgn}\left(\Delta^n_{i(n,q)}X\right)\ \forall q\in\{1,...,N_1\}\}\cap\{\sup_{|t-s|<2\Delta_n}|N_t-N_s|=1\}$. On this set, we have:

$$\widehat{s}_n=-\frac{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\left(\Delta^n_{i(n,q)}X+2\Delta_n\sum_{j=1}^{q-1}\Delta^n_{i(n,j)}X\right)}{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X}$$

and

$$\widehat{s}_n^c = -\frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right)\left(\Delta_{i(n,q)}^n X + 2\Delta_n \sum_{j=1}^{q-1} \Delta X_{T_q}\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X}.$$

Indeed, We then have

$$\widehat{s}_n = \widehat{s}_n^c + 2\Delta_n \frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{q-1}\left(\Delta X_{T_j} - \Delta_{i(n,j)}^n X\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X}$$

and the second term of the right hand side is equal to the sum of

$$2\Delta_n \frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{q-1} \Delta_{i(n,j)}^n \int_0^t \left(1 - e^{-\beta_n (t-s)}\right) x \underline{p}\left(dt, dx\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X} \tag{3.72}$$

and

$$-2\Delta_n \frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{q-1} \Delta_{i(n,j)}^n X^c}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X}. \tag{3.73}$$

(3.72) is equal to the sum of

$$2\Delta_n \frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{q-1} \Delta X_{T_j}\left(1 - e^{-\beta_n\left(t_{i(n,j)} - T_j\right)}\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X} \tag{3.74}$$

and

$$2\Delta_n \frac{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \sum_{j=1}^{q-1} Z_{t_{i(n,j)-1}}\left(1 - e^{-\beta_n \Delta_n}\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X} = O_p\left(\beta_n \Delta_n^2 \frac{\lambda_n^2}{\beta_n}\right). \tag{3.75}$$

The numerator of (3.74) is equal to

$$\int_0^1 \int_0^{t^-} \int_{\mathbb{R}^2} \operatorname{sgn}\left(x\right) y \left(1 - e^{-\beta_n\left(\lfloor \frac{s}{\Delta_n} \rfloor \Delta_n + \Delta_n - s\right)}\right) \underline{p}\left(dt, dx\right) \underline{p}\left(ds, dy\right).$$

(3.74) can be written as the sum of a bias and a variance term with the bias equal to

$$\begin{aligned} M_{n,3} &= \frac{2\Delta_n \lambda_n \operatorname{sgn}\left(x\right) \star \nu\ x \star \nu \int_0^1 \int_0^t e^{-\beta_n\left(\lfloor \frac{s}{\Delta_n} \rfloor \Delta_n + \Delta_n - s\right)} ds dt}{|x| \star \nu \left(\frac{1 - e^{-\beta_n \Delta_n}}{\beta_n \Delta_n}\right)} \\ &= \frac{2\Delta_n \lambda_n \operatorname{sgn}\left(x\right) \star \nu\ x \star \nu \left(\frac{1}{2} - \frac{\left(1 - e^{-\beta_n \Delta_n}\right)}{2\beta_n \Delta_n} + \frac{\left(1 + e^{-\beta_n \Delta_n}\right)}{2\beta_n} - \frac{1 - e^{-\beta_n \Delta_n}}{\beta_n^2 \Delta_n}\right)}{|x| \star \nu \left(\frac{1 - e^{-\beta_n \Delta_n}}{\beta_n \Delta_n}\right)}. \end{aligned}$$

The variance term is the sum of the following terms

$$\frac{2\Delta_n \int_0^1 \int_0^t \int_{\mathbb{R}^2} \operatorname{sgn}\left(x\right) y \left(1 - e^{-\beta_n\left(\lfloor \frac{s}{\Delta_n} \rfloor \Delta_n + \Delta_n - s\right)}\right) \left(\underline{p} - \lambda_n q\left(dt, dx\right)\right) \left(\underline{p} - \lambda_n q\left(ds, dy\right)\right)}{\sum_{q=1}^{N_1} \operatorname{sgn}\left(\Delta X_{T_q}\right) \Delta_{i(n,q)}^n X} = O_p\left(\beta_n \Delta_n^2\right),$$

$$\frac{2\lambda_n\Delta_n x\star\nu\int_0^1\int_{\mathbb{R}}\operatorname{sgn}(x)\int_0^1\mathbf{1}_{t>s}\left(1-e^{-\beta_n\left(\lfloor\frac{s}{\Delta_n}\rfloor\Delta_n+\Delta_n-s\right)}\right)ds\left(\underline{p}-\lambda_n q(dt,dx)\right)}{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X}=O_p\left(\sqrt{\lambda_n}\beta_n\Delta_n^2\right),$$

$$\frac{2\lambda_n\Delta_n\operatorname{sgn}(x)\star\nu\int_0^1\int_{\mathbb{R}}y\int_0^1\mathbf{1}_{t>s}dt\left(1-e^{-\beta_n\left(\lfloor\frac{s}{\Delta_n}\rfloor\Delta_n+\Delta_n-s\right)}\right)\left(\underline{p}-\lambda_n q(ds,dy)\right)}{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X}=O_p\left(\sqrt{\lambda_n}\beta_n\Delta_n^2\right),$$

and

$$-M_{n,3}\frac{\int_0^1\int_{\mathbb{R}}|x|e^{-\beta_n\left(\lfloor\frac{s}{\Delta_n}\rfloor\Delta_n+\Delta_n-s\right)}\left(\underline{p}-\lambda_n q(ds,dx)\right)+O_p\left(\lambda_n^2\Delta_n+\sqrt{\lambda_n\Delta_n}\right)}{\sum_{q=1}^{N_1}\operatorname{sgn}\left(\Delta X_{T_q}\right)\Delta^n_{i(n,q)}X}=O_p\left(\sqrt{\lambda_n}\beta_n\Delta_n^2\right)$$

using (3.50). (3.72) is then equal to

$$M_{n,3}+O_p\left(\beta_n\Delta_n\sqrt{\lambda_n}\Delta_n\right).$$

The numerator of (3.73) is equal to $O_p\left(\Delta_n\lambda_n\sqrt{\Delta_n}\right)$. Finally, we have

$$\widehat{s}_n=\widehat{s}_n^c+M_{n,3}+O_p\left(\beta_n\Delta_n\left(\sqrt{\lambda_n}\Delta_n+\frac{\sqrt{\Delta_n}}{\beta_n}\right)\right)\tag{3.76}$$

and

$$\widehat{s}_n^c=M_{s,n}+\beta_n\Delta_n V_{s,n}J_{s,n}.$$

Let

$$\widehat{\beta}_n=-\frac{1}{\Delta_n}\log\left(\left(1-\widehat{s}_n\right)\vee\Delta_n\right)$$

which is equal to $-\frac{1}{\Delta_n}\log\left(1-\widehat{s}_n\right)$ for n enough large. We apply Taylor's expansion in order to have the results of Theorem 3.1 with $M_n=M_{c,n}+e^{\beta\Delta_n}\frac{M_{n,3}}{\beta_n\Delta_n}$. The error of order $O_p\left(\beta_n\Delta_n\left(\sqrt{\lambda_n}\Delta_n+\frac{\sqrt{\Delta_n}}{\beta_n}\right)\right)$ in (3.76) becomes an error of order $O_p\left(\sqrt{\lambda_n}\Delta_n+\frac{\sqrt{\Delta_n}}{\beta_n}\right)$ for $\frac{\widehat{\beta}_n-\beta_n}{\beta_n}$ which is negligible compared to the errors of the variance terms. The term M_n is equal to

$$e^{\beta_n\Delta_n}\frac{\lambda_n}{\beta_n}\frac{x\star\nu\operatorname{sgn}(x)\star\nu}{|x|\star\nu}\left(\frac{e^{\beta_n\Delta_n}-1}{\beta_n\Delta_n}-1\right)+O\left(\frac{\lambda}{\beta^2}\right)$$

with

$$O\left(\frac{\lambda}{\beta^2}\right)=o\left(V_{2,n}\right)$$

and can be omitted.

Bibliography

[1] René Aid, Luciano Campi, Adrien Nguyen Huu, and Nizar Touzi. A structural risk-neutral model of electricity prices. *International Journal of Theoretical and Applied Finance*, 12(07):925–947, 2009.

[2] Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009.

[3] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[4] Ole Eiler Barndorff-Nielsen, Neil Shephard, and Matthias Winkel. Limit theorems for multipower variation in the presence of jumps. *Stochastic processes and their applications*, 116(5):796–806, 2006.

[5] Fred Espen Benth, Jan Kallsen, and Thilo Meyer-Brandis. A non-gaussian ornstein–uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2):153–169, 2007.

[6] Fred Espen Benth, Rüdiger Kiesel, and Anna Nazarova. A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5):1589–1616, 2012.

[7] Hermine Biermé and Agnes Desolneux. A fourier approach for the level crossings of shot noise processes with jumps. *Journal of Applied Probability*, pages 100–113, 2012.

[8] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[9] Rama Cont and Peter Tankov. *Financial modelling with jump processes*, volume 2. CRC press, 2003.

[10] Olivier Féron and Elias Daboussi. *Commodities, Energy and Environmental Finance*, chapter Calibration of electricity price models, pages 183–207. Springer, 2015.

[11] Hans Föllmer and Martin Schweizer. Hedging of contingent claims. *Applied stochastic analysis*, 5:389, 1991.

[12] Hélyette Geman and Andrea Roncoroni. Understanding the fine structure of electricity prices. *The Journal of Business*, 79(3):1225–1261, 2006.

[13] Jhonny Gonzalez, John Moriarty, and Jan Palczewski. Bayesian calibration and number of jump components in electricity spot price models. *Available at SSRN 2714382*, 2016.

[14] Jean Jacod and Philip Protter. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.

[15] Rüdiger Kiesel, Gero Schindlmayr, and Reik H Börger. A two-factor model for the electricity forward market. *Quantitative Finance*, 9(3):279–287, 2009.

[16] Jia Li, Viktor Todorov, and George Tauchen. Volatility occupation times. *The Annals of Statistics*, 41(4):1865–1891, 2013.

[17] Jia Li, Viktor Todorov, and George Tauchen. Jump regressions. *Econometrica*, 85(1):173–195, 2017.

[18] Julio J Lucia and Eduardo S Schwartz. Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of derivatives research*, 5(1):5–50, 2002.

[19] Cecilia Mancini. Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52, 2004.

[20] Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

[21] Thilo Meyer-Brandis and Peter Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(05):503–528, 2008.

[22] Manuel Moreno, Pedro Serrano, and Winfried Stute. Statistical properties and economic implications of jump-diffusion processes with shot-noise effects. *European Journal of Operational Research*, 214(3):656–664, 2011.

[23] Giovanni Peccati and Murad S Taqqu. Central limit theorems for double poisson integrals. *Bernoulli*, 14(3):791–821, 2008.

[24] Philip Protter and Kazuhiro Shimbo. No arbitrage and general semimartingales. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, pages 267–283. Institute of Mathematical Statistics, 2008.

[25] Thorsten Schmidt. Modelling energy markets with extreme spikes. In *Mathematical Control Theory and Finance*, pages 359–375. Springer, 2008.

[26] Martin Schweizer. *Handbooks in mathematical finance: Option pricing, interest rates and risk management*, chapter A guided tour through quadratic hedging approaches, pages 538–574. Cambridge University Press, 2001.

[27] Almut Elisabeth Dorothea Veraart. Inference for the jump part of quadratic variation of Itô semimartingales. *Econometric Theory*, 26(2):331–368, 2010.

Chapter 4

Local polynomial estimation of a doubly stochastic Poisson process

Abstract

We consider a doubly stochastic Poisson process with stochastic intensity $\lambda_t = nq(X_t)$ where X is a continuous Itô semimartingale and n is an integer. Both processes are observed continuously over a fixed period $[0, T]$. An estimation procedure is proposed in a non parametrical setting for the function q on an interval I where X is sufficiently observed using a local polynomial estimator. A method to select the bandwidth in a non asymptotic framework is proposed, leading to an oracle inequality. If m is the degree of the chosen polynomial, the accuracy of our estimator over the Hölder class of order β is $n^{\frac{-\beta}{2\beta+1}}$ if $m \geq \lfloor \beta \rfloor$ and $n^{\frac{-m}{2m+1}}$ if $m < \lfloor \beta \rfloor$ and is optimal in the minimax sense if $m \geq \lfloor \beta \rfloor$. A parametrical test is also proposed to test if q belongs to some parametrical family. Those results are applied to French temperature and electricity spot prices data where we infer the intensity of electricity spot spikes as a function of the temperature.

1 Introduction

1.1 Motivation

Jump processes are used in several domains such as finance, insurance or neuroscience. They are a particular case of point processes, see [19] for more details. In finance, they allow to model discontinuities in equity prices time series and heavy tails in asset returns [12]. An application for insurance is the model of extreme events such that occurrence times of earthquakes [24]. In neuroscience, these processes model spikes which is a potential difference in the membrane of a neuron [29]. Spikes is a high increase of the potential in the membrane followed by a quick reversion to the initial level of the potential. Spikes are also present in electricity spot prices time series and can be both negative and positive: price level is very high or very low during a short time period before coming back to its original level [5]. Spikes are often modeled by mean reverting jump processes [9; 22].

In all these areas, the frequency of the jumps can be explained by an exogenous variable. Modeling these dependences can not be omitted, because they have an impact on risk management or prediction and help us to understand some behaviors. In [29], the author explains the neural spiking activity with three king of covariates: the previous spikes, exogenous stimuli and concurrent neural activity. In [3], the authors propose a model for financial contagion. Financial contagion is the fact that a large price move in a market causes large price moves in other markets ; jumps are explained by a first jump and in this case we often use multidimensional Hawkes process which are mutually exciting processes, see [3]. For electricity spot prices, spikes are often caused by abnormal temperatures which are not modeled by a jump process. In a general case, when the covariate is not an other jump process, we often use doubly stochastic Poisson processes, which are Poisson processes with stochastic intensity. Two of the most famous models are the Aalen multiplicative model introduced in [1] where the intensity process is of the form $\alpha_t Y_t$ with α_t a function of time and Y_t a stochastic process and the Cox regression model introduced in [13] where the intensity process is of the form $\alpha_t \exp\left(\beta^T Z\right)$ with Z a multi-dimensional random variable. Both are used mainly for life times modeling.

A large literature is dedicated on methods of estimation for intensity estimation of Poisson process, especially in a non parametric setting. In the case of inhomogeneous Poisson processes, [27] and [28] use projection estimators and model selection techniques. Several finite dimensional spaces called models are considered to find projection estimators and they propose a penalty criterion in order to select a model. They work in a non asymptotic framework and a concentration inequality is found. Furthermore, minimax rates are found over several classes of functions. A different approach for the estimation of the intensity is the use of kernel methods as in [15] and [8]. In [15] and [8] , asymptotic properties of the kernel estimator are studied ; in [8], methods to select the bandwidth is proposed. In [33], the intensity can be stochastic and is also estimated as a function of time with a kernel estimator in an asymptotic framework. In the context of Cox and Aalen processes, [11] also proposes model selection techniques with projection estimators ; local polynomial estimator, which is a generalization of kernel estimators, is proposed by [10] and studied in an asymptotic framework. A method of estimation in asymptotic framework for Cox regression with a time dependent covariable is established in [23].

1.2 Objectives and results

In our case, we are interested in a doubly stochastic Poisson process denoted by N where its intensity λ is a function of an exogenous covariate which is a stochastic process X_t

$$\lambda_t = q\left(X_t\right)$$

and our goal is to estimate the function q. In this case, conditionally on X, N is a inhomogeneous Poisson process. We assume that we observe N and X over a time horizon $[0, T]$. We can think of the example of the frequency of electricity spot prices spikes as a function of the temperature depending also on time. This framework has already been studied. Indeed, [31] proposes a kernel estimator of the function q in the case where T goes to ∞ and when X satisfies some conditions, which can be for instance stationarity. In [14], X_t corresponds to the fractional part of a Brownian motion and the doubly stochastic Poisson process is used to model the limit order book ; an estimation procedure is proposed in an asymptotic framework. We consider a different framework where X is a continuous Itô semimartingale. X must also have a local time l_T^x for x in $\mathbb{R}$ measuring the time spend by X around x before time T verifying properties of Proposition 4.1 which can be insured by low restrictions given in Assumption 4.1. The function q can be estimated at point x only if X takes this value before time T, or if $l_T^x > 0$. We estimate q on an arbitrarily interval I and we need to work conditionally on the event

$$D\left(I, \nu\right) = \{\omega \in \Omega,\ \inf_{x \in I} l_T^x\left(\omega\right) \geq \frac{\nu T}{|I|}\}$$

with $\nu \in (0, 1]$. We choose to work in a non parametric framework and in a non asymptotic framework. To our knowledge, inferencing the intensity of a doubly stochastic Poisson process as a function of a continuous Itô semimartingale in a non-asymptotic framework is not present in the literature.

A local polynomial estimator $\hat{q}_h$ of q is proposed in Section 3 with h a bandwidth parameter. The criteria used to evaluate the performance of our estimator is the L_2 norm on I, $\|\cdot\|_I$. We also gives a method to select a bandwidth over a finite set $\mathcal{H}$. We adapt the method of [20] used for density estimation with i.i.d. observations to our context of intensity estimation for doubly stochastic Poisson process. The method consists in approximating the bias of $\hat{q}_h$ by an estimator of $\|q_h - q_{h_{\min}}\|_I^2$ where $q_h = \mathbb{E}\left(\hat{q}_h|X\right)$ and $h_{\min} = \min\ \mathcal{H}$. Indeed, if $h_{\min}$ is sufficiently small, the bias of $\hat{q}_{h_{\min}}$ is negligible and $\|q_h - q_{h_{\min}}\|_I^2 \approx \|q - q_h\|_I^2$. This method is an extension of the one of Goldenshluger and Lepski [16]. One of our main results is the oracle inequality of Proposition 4.6. If we write

$$\lambda_t = nq\left(X_t\right)$$

with $n \in \mathbb{N},\ n \geq 1$, we obtain

$$\begin{aligned}\mathbb{E}\left(\|q - \hat{q}_{\hat{h}}\|_I^2 | D\left(I, \nu\right)\right) \leq & \left(\lambda \vee \frac{1}{\lambda} + O\left(\log\left(n\right)^{-1}\right)\right) \min_{h \in \mathcal{H}} \mathbb{E}\left(\|q - \hat{q}_h\|_I^2 | D\left(I, \nu\right)\right) \\ & + O\left(\log\left(n\right) \mathbb{E}\left(\|q - q_{h_{\min}}\|_I^2 | D\left(I, \nu\right)\right)\right) + O\left(\frac{\log\left(n \vee |\mathcal{H}|\right)^6}{n}\right)\end{aligned}$$

where $\hat{h}$ is the bandwidth obtained using our selection method and $\lambda > 0$ a parameter chosen by the statistician present in the penalty criteria. Furthermore, if we consider that q belongs the

Hölder class on I with parameter β, our estimator is optimal in the minimax rate sense and the minimax rate of convergence is $n^{-\frac{\beta}{2\beta+1}}$ if the degree of the polynomial is greater than $\lfloor\beta\rfloor$, see Proposition 4.7. In addition to give a method of estimation for the intensity as a function of X, we have also shown that the method of [20] is adapted to inhomogeneous Poisson process because the case $X_t = t$ respects the assumptions.

A second objective is to test if our function q belongs to some class of parametrical model. Indeed, non parametric estimators are not convenient for operational applications. This objective is achieved in Section 4 where a test is proposed. We test

$$H_0 : q = q_\theta \text{ on } I \text{ for some } \theta \in \Theta$$

against

$$H_1 : q \neq q_\theta \text{ on } I \text{ for all } \theta \in \Theta$$

where $\Theta \in \mathbb{R}^d$. We consider the contrast $M_n(\theta)$ defined in (4.10) that is an unbiased estimator of the distance between q and g_θ based on our estimator $\hat{q}_{\hat{h}}$ which is an estimator of q under both hypothesis. Under H_0, the minimum of the contrast gives an estimator of θ, $\hat{\theta}_n$, converging at the rate $\sqrt{n}$ and a central limit theorem is given when $n \to \infty$, see Proposition 4.8 (i). Furthermore, the quantity $\hat{h}^{\frac{1}{2}} n M_n\left(\hat{\theta}_n\right)$ converges in law towards a normal random variable under H_0, see Proposition 4.8 (ii), but to ∞ under H_1, see Proposition 4.8 (iii). This allows us to propose a critical region for the test.

In Section 5, our estimation procedure is applied on electricity prices and temperature data in order to model the dependence between the spikes frequency of electricity prices and the temperature. In Section 6, results on simulated data are given.

Proof of the oracle inequality is given in Section 7 and other proofs in Section 8.

2 Statistical setting

Let $(X_t)_{0\leq t\leq T}$ be a real valued continuous semimartingale of the form

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \tag{4.1}$$

defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_s)_{0\leq s\leq T}, \mathbb{P}\right)$ where $(W_t)_{0\leq t\leq T}$ is a standard Brownian motion, $(b_t)_{0\leq t\leq T}$ and $(\sigma_t)_{0\leq t\leq T}$ are cádlág, progressively measurables and verify $\int_0^T \left(|\mu_s| + \sigma_s^2\right) ds < \infty$ almost surely. Let $(N_t)_{0\leq t\leq T}$ be a doubly stochastic Poisson process with intensity $(\lambda_t)_{0\leq t\leq T}$ also defined on $\left(\Omega, \mathcal{F}, (\mathcal{F}_s)_{0\leq s\leq T}, \mathbb{P}\right)$. We observe the two processes on $[0, T]$ with T finite. We assume that the intensity, which is the function of interest, is of the form

$$\lambda_t = nq(X_t)$$

where $n \in \mathbb{N}$, $n \geq 1$ corresponds to the asymptotic. As we observe the counting process on a finite time horizon, we need to have a sufficient number of jumps during this finite period, which is of order $\int_0^T \lambda_u du$ and then in our case of order n. We denote by $\Lambda_\cdot = \int_0^\cdot \lambda_u du$ the compensator of N and by $M = N - \Lambda$ the compensated Poisson process. Conditionally on X, N is an inhomogeneous Poisson process with deterministic intensity at time t $nq(X_t)$. Our aim is to estimate the function $(q(x),\ x \in I)$ for an arbitrary compact interval I.

Assumption 4.1. *We assume that one of the following assumptions is true:*

(i) $\inf_{0\leq s\leq T}\sigma_s \geq \underline{\sigma}$ *with* $\underline{\sigma}>0$ *a deterministic constant and*

$$\mathbb{E}\left(\int_0^T |\mu_s|ds + \sup_{0\leq t\leq T}|\int_0^T \sigma_s dW_s|\right) < \infty,$$

(ii) $X_t = t$ *for all* t *in* $[0,T]$.

Remark 4.1. *Assumption 4.1 (ii) can be more general. These assumptions are sufficient but not necessary conditions for existence of a local time. In the case* μ *deterministic and* $\sigma = 0$*, we need for all* $0 \leq s \leq T$*,* $\mu_s \neq 0$ *but only the case* $\mu = 1$ *is interesting. Existence of local time with a stochastic drift and a null volatility could also be considered but existing results about local times for absolute continuous processes are not enough in the literature to consider it.*

Proposition 4.1. *Let* X *the process defined in (4.1). Under Assumption 4.1, there exists a function defined on* $\mathbb{R}\times[0,T]$ *and denoted* $(x,t)\mapsto l_t^x$ *verifying*

(i) an occupation time formula of the form

$$\int_0^t f(X_s)\,ds = \int_{\mathbb{R}} f(x)\,l_t^x dx,\ 0\leq t\leq T$$

for any measurable function f *on* $\Omega\times\mathbb{R}$*,*

(ii) $\mathbb{E}\left(\sup_{x\in\mathbb{R}} l_T^x\right) < \infty$ *and*

(iii) $x\mapsto l_T^x$ *is continuous on* $\mathbb{R}$ *under Assumption 4.1 (i) and has one point of discontinuity under Assumption 4.1 (ii).*

As noticed in [17], the estimation of $q(x)$ at point $x\in I$ is meaningful only if the process X hits the point x before time T, or if $l_T^x > 0$. Indeed, l_T^x is equal to

$$\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^T \mathbf{1}_{|X_s-x|\leq\epsilon}ds$$

and measures the time spend by X arround the point x. For $\nu\in(0,1]$, let us define the event

$$D(I,\nu) = \{\omega\in\Omega,\ \inf_{x\in I} l_T^x(\omega) \geq \frac{\nu T}{|I|}\}$$

with $|I|$ the Lebesgue measure of I. From now, we work on the event $D(I,\nu)$. Under (ii), if $X_t = t$, the natural choice of I is $[0,T]$ and $\nu = 1$.

Remark 4.2. *We choose* ν *being dimensionless, justifying the normalization by* $\frac{|I|}{T}$*. Indeed, the local time have dimension equal to time times the inverse of the dimension of* X*. Furthermore, if* $I = \left[\underline{I},\overline{I}\right]$*, and if we do the mapping* $X'_t = \frac{X_{tT}-\underline{I}}{|I|}$ *for* $t\in[0,1]$*, we have* $\inf_{x\in[0,1]} l_1^x(X') = \frac{|I|}{T}\inf_{x\in I} l_T^x$ *where* $l_1^x(X')$ *is the local time of* X' *at time* 1 *and point* x*. As* $\int_I l_T^x dx \geq \inf_{x\in I} l_T^x|I|$ *and* $\int_I l_T^x dx = \int_0^T \mathbf{1}_{X_s\in I}ds \leq T$*,* $\inf_{x\in I} l_T^x \leq \frac{T}{|I|}$ *and* ν *has to be bounded by 1.*

3 Local polynomial estimation

Let $m > 0$ be an integer, K a kernel function and $\mathcal{S}_{m+1}^{+}$ the set of positive definite matrix of $\mathbb{R}^{m+1\times m+1}$. Let

$$K_h(u) = h^{-1}K\left(\frac{u}{h}\right),\ u \in \mathbb{R},\ h > 0.$$

Let us consider the local polynomial estimator for $h > 0$ and $x \in \mathbb{R}$:

$$\hat{q}_h(x) = \frac{1}{n}\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} dN_s \tag{4.2}$$

with

$$w(x,h,z) = U^T(0)\,B(x,h)^{-1}\,U(z)\,\mathbf{1}_{B(x,h)\in\mathcal{S}_{m+1}^{+}},\ z \in \mathbb{R}, \tag{4.3}$$

$$U(x) = \left(1, x, \frac{x^2}{2!}, .., \frac{x^m}{m!}\right)^T,$$

and

$$B(x,h) = \int_0^T U\left(\frac{X_s - x}{h}\right) U^T\left(\frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} ds. \tag{4.4}$$

If $B(x,h) \in \mathcal{S}_{m+1}^{+}$, the estimator $\hat{q}_h(x)$ is equal to $U^T(0)\,\hat{\theta}_h(x)$ with

$$\begin{aligned}\hat{\theta}_h(x) = \underset{\theta\in\mathbb{R}^{m+1}}{argmin} &- \frac{2}{n}\theta^T\int_0^T U\left(\frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} dN_s \\ &+ \theta^T\int_0^T U\left(\frac{X_s - x}{h}\right) U^T\left(\frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} ds\theta.\end{aligned} \tag{4.5}$$

The case $l = 0$ corresponds to the classical Nadaraya Watson estimator. The term $\mathbf{1}_{X_s\in I}$ allows us to avoid issues at the boundaries. We denote by q_h the conditional expectation of $\hat{q}_h$ given X:

$$q_h(x) = \int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} q(X_s)\, ds.$$

If q is a polynomial function of degree m on I, $q_h(x)$ is equal to $q(x)$, see Proposition 4.2.

Proposition 4.2. *Let $x \in I$ and $h > 0$ such that $B(x,h) \in \mathcal{S}_{m+1}^{+}$ where B is defined in (4.4). Let Q be a polynomial function of degree $\leq m$. For any realization of the process X, we have*

$$\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} Q(X_s)\, ds = Q(x).$$

In particular,

$$\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} ds = 1,$$

$$\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\,\mathbf{1}_{X_s\in I} (X_s - x)^k\, ds = 0,\ 1 \leq k \leq m.$$

Proof of Proposition 4.2 is immediate using the Taylor's expansion of a polynomial function and (4.5) ; it is similar to the proof of [30, Proposition 1.12].
For $B(x,h)^{-1}$ to be defined, the positive matrix $B(x,h)$ must be definite. Assumption 4.2 is sufficient for $B(x,h)^{-1}$ to be well defined on the event $D(I,\nu)$, see Proposition 4.3.

Assumption 4.2. *We assume that:*

(i) there exists $K_{\min} > 0$ and $\Delta > 0$ such that $K(u) \geq K_{\min}\mathbf{1}_{|u|\leq\Delta}$ for all u in $\mathbb{R}$,

(ii) K has a compact support belonging to $[-1,1]$ and $\|K\|_\infty = \sup_{x\in\mathbb{R}}|K(x)| < \infty$.

Proposition 4.3. *Let $0 < h \leq \frac{2}{3}\Delta|I|$. Under Assumption 4.2, on the event $D(I,\nu)$, the matrix $B(x,h)$ defined in (4.4) belongs to $\mathcal{S}^+_{m+1}$ and for $x \in I$, $z \in \mathbb{R}$,*

$$|w(x,h,z)\,\mathbf{1}_{|z|\leq 1}| \leq \frac{|I|}{A_K \nu T}$$

with w defined in Equation (4.3) and A_K a constant depending on K.

3.1 Method for bandwidth selection

Our objective is to propose a method in order to choose the bandwidth h. We want for this bandwidth to minimize the L_2 loss on the interval I, $\mathbb{E}\left(\|q-\hat{q}_h\|_I^2|D(I,\nu)\right)$, with $\|f\|_I^2 = \int_I f(x)^2\,dx$ for $f \in L_2(I)$. This loss is equal to the sum of a bias term, $\mathbb{E}\left(\|q-q_h\|_I^2|D(I,\nu)\right)$, which depends on the regularity of q and is usually increasing with h, and a variance term $\mathbb{E}\left(\|\hat{q}_h-q_h\|_I^2|D(I,\nu)\right)$, decreasing with h. The theoretical bandwidth minimizing this quantity depends on the function q itself which is unknown and is called the oracle. One wants to find an estimator of this oracle, $\hat{h}$, such that $\mathbb{E}\left(\|q-\hat{q}_{\hat{h}}\|_I^2|D(I,\nu)\right) \leq (1+o(1))\min_{h\in\mathcal{H}}\mathbb{E}\left(\|\hat{q}_h-q\|_I^2|D(I,\nu)\right)+o(1)$ when $n\to\infty$, in order to have a loss with $\hat{h}$ close to the minimal one ; this type of inequality is called an oracle inequality. The usual method to find this estimator of the oracle, which is done in this section, is to find an unbiased estimator of the bias and of the variance and to consider the bandwidth h minimizing the sum of the two estimators. In the following, we will also denote by $<\cdot,\cdot>_I$ the associated scalar product of $\|\cdot\|_I$.

In order to select the bandwidth parameter, we use the approach of [20] which is used for density estimation. We consider a finite set $\mathcal{H}$ of $(0,\infty)$ and $h_{\min} = \min\mathcal{H}$. The idea is to approximate the bias by $\mathbb{E}\left(\|q_h-q_{h_{\min}}\|_I^2\right)$ with $h_{\min}$ sufficiently small. A natural estimator of $\mathbb{E}\left(\|q_h-q_{h_{\min}}\|_I^2\right)$ is

$$\|\hat{q}_h-\hat{q}_{h_{\min}}\|_I^2.$$

This estimator induces a bias equal to the expectation of

$$\frac{-1}{n}\int_I\int_0^T\left(w\left(x,h,\frac{X_s-x}{h}\right)K_h(X_s-x)-w\left(x,h_{\min},\frac{X_s-x}{h_{\min}}\right)K_{h_{\min}}(X_s-x)\right)^2 dx q(X_s)\,ds$$

which can be estimated by

$$-\frac{1}{n^2}\int_I\int_0^T\left(w\left(x,h,\frac{X_s-x}{h}\right)K_h(X_s-x)-w\left(x,h_{\min},\frac{X_s-x}{h_{\min}}\right)K_{h_{\min}}(X_s-x)\right)^2 dx dN_s.$$

This estimator can be written as

$$-\hat{V}_h - \hat{V}_{h_{\min}} + 2\hat{V}_{h,h_{\min}}$$

where

$$\hat{V}_h = \frac{1}{n^2}\int_0^T \int_I \left(w\left(x, h, \frac{X_s - x}{h}\right) K_h\left(X_s - x\right)\right)^2 \mathbf{1}_{X_s \in I} dx dN_s$$

and

$$\hat{V}_{h,h_{\min}} = \frac{1}{n^2}\int_0^T \int_I w\left(x, h, \frac{X_s - x}{h}\right) K_h\left(X_s - x\right) w\left(x, h_{\min}, \frac{X_s - x}{h_{\min}}\right) K_{h_{\min}}\left(X_s - x\right) \mathbf{1}_{X_s \in I} dx dN_s. \quad (4.6)$$

An unbiased estimator of the variance which is equal to

$$\mathbb{E}\left(\frac{1}{n}\int_0^T \int_I \left(w\left(x, h, \frac{X_s - x}{h}\right) K_h\left(X_s - x\right)\right)^2 \mathbf{1}_{X_s \in I} dx q\left(X_s\right) ds\right)$$

is given by

$$\hat{V}_h = \frac{1}{n^2}\int_0^T \int_I \left(w\left(x, h, \frac{X_s - x}{h}\right) K_h\left(X_s - x\right)\right)^2 \mathbf{1}_{X_s \in I} dx dN_s.$$

In order to choose the bandwidth, we use the criteria

$$\|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2 + pen\left(h\right)$$

where

$$pen_\lambda\left(h\right) = \lambda\hat{V}_h - \hat{V}_h - \hat{V}_{h_{\min}} + 2\hat{V}_{h,h_{\min}}, \text{ with } \lambda > 0. \quad (4.7)$$

The optimal bandwidth $\hat{h}$ is given by

$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2 + pen\left(h\right). \quad (4.8)$$

In the following, we want to derive an oracle inequality for the estimator $\hat{q}_{\hat{h}}$ which has not be done to our knowledge.

3.2 Concentration inequalities

In order to compute this oracle inequality, we first need the two following concentration inequalities, from [28] and [18]. The concentration inequality of Proposition 4.4 is a weak Bernstein inequality, the one of Proposition 4.5 is an inequality for the Poisson U-statistic. These inequalities will be useful in our case because N is an inhomogeneous Poisson process conditionally on X.

Proposition 4.4. *[28, Equation (2.2)] Let $T > 0$. Let N be an inhomogeneous Poisson process with intensity $\lambda_\cdot$, $\Lambda_\cdot = \int_0^\cdot \lambda_u du$ and $M = N - \Lambda$. For all $u \geq 0$, with probability larger than $1 - e^{-u}$*

$$|\int_0^T f\left(s\right) dM_s| \leq \sqrt{2u\int_0^T f^2\left(s\right) d\Lambda_s} + \frac{\sup_{x\in[0,T]} |f\left(x\right)| u}{3}.$$

Proposition 4.5. *[18, Theorem 4.2] Let $T > 0$. Let N be an inhomogeneous Poisson process with intensity $\lambda_\cdot$, $\Lambda_\cdot = \int_0^\cdot \lambda_u du$ and $M = N - \Lambda$. For all $\epsilon, u \geq 0$, with probability larger than $1 - 2.77e^{-u}$*

$$\int_0^T \int_0^{s^-} f(u,s)\, dM_u dM_s \leq 2(1+\epsilon)^{\frac{3}{2}} C\sqrt{u} + 2\nu(\epsilon) Du + \beta(\epsilon) Bu^{\frac{3}{2}} + \gamma(\epsilon) Au^2$$

where

$$\nu(\epsilon) = \sqrt{2\kappa}\left(2+\epsilon+\epsilon^{-1}\right),\ \beta(\epsilon) = e\left(1+\epsilon^{-1}\right)^2 \kappa(\epsilon) + \left(\sqrt{2\kappa}\left(2+\epsilon+\epsilon^{-1}\right)\right) \vee \frac{(1+\epsilon)^2}{\sqrt{2}},$$

$$\gamma(\epsilon) = \left(e\left(1+\epsilon^{-1}\right)^2 \kappa(\epsilon)\right) \vee \frac{(1+\epsilon)^2}{3},\ \kappa = 6,\ \kappa(\epsilon) = 1.25 + \frac{32}{\epsilon}$$

and

$$A = \sup_{(u,s)\in[0,T]^2} f(u,s),\ B^2 = \max\{\sup_{s\leq T} \int_0^s f(u,s)^2 d\Lambda_u, \sup_{u\leq T} \int_0^u f(u,s)^2 d\Lambda_s\},$$

$$C^2 = \int_0^T \int_0^s f(u,s)^2 d\Lambda_u d\Lambda_s,\ D = \sup_{\int_0^T a_u^2 d\Lambda_u = 1, \int_0^T b_s^2 d\Lambda_s = 1} \int_0^T a_u \int_u^T b_s f(u,s)\, d\Lambda_u d\Lambda_s.$$

3.3 Oracle inequality

For a function $f \in L_\infty(I)$, we denote by $\|f\|_{I,\infty}$ the norm $\sup_{x\in I}|f(x)|$. We will also need for the kernel the following norms: $\|\cdot\|_1$, $\|\cdot\| = \|\cdot\|_2$ and $\|\cdot\|_\infty$ corresponding respectively to the L_1, L_2 and L_∞ norms on $\mathbb{R}$, with the L_p norm defined by $\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p\right)^{\frac{1}{p}}$, $p \geq 1$ and $\|f\|_\infty = \sup_{x\in\mathbb{R}}|f(x)|$. Proposition 4.6 gives an oracle inequality for $\hat{q}_{\hat{h}}$.

Proposition 4.6. *Assume 4.1 and 4.2. Let $x \geq 1$, $\epsilon \in (0,1)$. Let $\mathcal{H}$ a finite subset of $(0,\infty)$ such that $\min\mathcal{H} = h_{\min} \geq \frac{\|K\|_\infty \|K\|_1 |I|}{n}$ and $\max\mathcal{H} \leq \frac{2}{3}|I|\Delta$. Let $\hat{q}_h$ the local polynomial estimator defined in (4.2) and $\hat{h}$ the bandwidth defined in (4.8). With conditional probability given X larger than $C_1|\mathcal{H}|e^{-x}$, on the event $D(I,\nu)$,*

$$\|\hat{q}_{\hat{h}} - q\|_I^2 \leq C_0(\epsilon,\lambda) \min_{h\in\mathcal{H}} \|\hat{q}_h - q\|_I^2 + C_2(\epsilon,\lambda)\|q_{h_{\min}} - q\|_I^2 + \frac{C_3(\epsilon,K,\lambda)|I|}{\nu^2 T^2}\left(\frac{(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n} + \frac{x^3|I|}{n^2 h_{\min}}\right)$$

where $C_0(\epsilon,\lambda) = \lambda + \epsilon$ if $\lambda \geq 1$ and $C_0(\epsilon,\lambda) = \frac{1}{\lambda} + \epsilon$ if $0 < \lambda < 1$, $C_2(\epsilon,\lambda)$ is a constant depending only on ϵ and λ and $C_3(\epsilon,K,\lambda)$ is a constant depending only on ϵ, K and λ. Furthermore, $C_2(\epsilon,\lambda) \asymp \frac{1}{\epsilon}$ and $C_3(\epsilon,K,\lambda) \asymp \frac{1}{\epsilon^3}$ when $\epsilon \to 0$.

We also have,

$$\begin{aligned}\mathbb{E}(\|\hat{q}_{\hat{h}}-q\|_I^2|D(I,\nu)) \leq & \left(\lambda\vee\frac{1}{\lambda}+\frac{\tilde{C}_1}{\log(n)}\right)\min_{h\in\mathcal{H}}\mathbb{E}\left(\|\hat{q}_h-q\|_I^2|D(I,\nu)\right)\\ &+\tilde{C}_2(\lambda)\log(n)\,\mathbb{E}\left(\|q_{h_{\min}}-q\|_I^2|D(I,\nu)\right)\\ &+\frac{\tilde{C}_3(K,\lambda)|I|}{\nu^2T^2}\left(\frac{(1+\|q\|_{I,\infty}\mathbb{E}(\|l_T\|_{I,\infty}|D(I,\nu))|I|)\log(n\vee|\mathcal{H}|)^5}{n}+\frac{\log(n\vee|\mathcal{H}|)^6}{n}\right)\\ &+\tilde{C}_4\frac{\|q\|_I^2}{n^4}+\frac{\tilde{C}_5(K)|I|}{\nu^2T^2}\frac{\sqrt{\sum_{i=1}^4(\|q\|_{\infty,I}T)^i}}{n}. \end{aligned} \tag{4.9}$$

where $\tilde{C}_1$ and $\tilde{C}_4$ are constant, $\tilde{C}_2(\lambda)$ is a constant that only depends on λ, $\tilde{C}_3(K,\lambda)$ is a constant that only depends on K and λ and $\tilde{C}_5(K)$ is a constant that only depends on K.

In this inequality, one can see the presence of an error of order $\log(n)\,\mathbb{E}\left(\|q_{h_{\min}}-q\|_I^2|D(I,\nu)\right)$. This error comes from the approximation of the bias $\|q_h-q\|_I^2$ by $\|q_h-q_{h_min}\|_I^2$ and is negligible if $h_{\min}$ is small enough. We also remark that if $\lambda=1$, the constant of the oracle inequality is optimal.

3.4 Adaptative minimax estimation

In this section, we study the performance of the estimator $\hat{q}_{\hat{h}}$ in terms of convergence rate. We now work with the asymptotic $n\to\infty$, meaning the number of jumps becomes large when $n\to\infty$. For $\rho,\beta,L>0$, let $\Lambda_{\rho,\beta}=\{f:I\to\mathbb{R}:\ f(x)\geq\rho,\ \|f\|_{I,\infty}<\infty\}\cap\Sigma(\beta,L,I)$ where $\Sigma(\beta,L,I)$ is the Hölder class on I defined as the set of $l=\lfloor\beta\rfloor$ differentiable functions $f:I\to\mathbb{R}$ whose derivative $f^{(l)}$ verifies

$$|f^{(l)}(x)-f^{(l)}(x')|\leq L|x-x'|^{\beta-l},\ \forall x,\ x'\in I,$$

see [30, Definition 1.2]. We will restrict to the study of $q\in\Lambda_{\rho,\beta}$.

To evaluate the performance of an estimator $\tilde{q}_n$ of q, we consider the minimax risk

$$\mathbf{R}(\tilde{q}_n,\Lambda_{\rho,\beta},\varphi_n)=\sup_{\lambda\in\Lambda_{\rho,\beta}}\mathbb{E}\left(\varphi_n^{-2}\int_I(\tilde{q}_n(N,X,x)-q(x))^2\,dx|D(I,\nu)\right).$$

An estimator $\tilde{q}_n^*$ is said to attain an optimal rate of convergence $\varphi_n(\Lambda_{\rho,\beta})$ if

$$\limsup_{n\to\infty}\mathbf{R}(\tilde{q}_n^*,\Lambda_{\rho,\beta},\varphi_n(\Lambda_{\rho,\beta}))<\infty$$

and no estimator can attain a better rate:

$$\liminf_{n\to\infty}\inf_{\tilde{q}_n}\mathbf{R}(\tilde{q}_n,\Lambda_{\rho,\beta},\varphi_n(\Lambda_{\rho,\beta}))>0$$

where the infimum is taken over all estimators.

Proposition 4.7. *Assume 4.1 and 4.2. Let us consider the set of bandwidth $\mathcal{H}=\{h>0|h\geq\frac{\|K\|_\infty\|K\|_1|I|}{n},\ h\leq\frac{2}{3}|I|\Delta$ and $h^{-1}\in\mathbb{N}\}$. Let $\hat{q}_{\hat{h}}$ the estimator defined in Proposition 4.6 and let m be the degree of the polynomial. In the case where $m\geq\lfloor\beta\rfloor$, $\hat{q}_{\hat{h}}$ is optimal in the minimax sense and the optimal rate of convergence is given by $\varphi(\Lambda_{\rho,\beta})=n^{\frac{-\beta}{2\beta+1}}$. In the case where $m<\lfloor\beta\rfloor$, the rate of convergence of $\hat{q}_h$ is $n^{\frac{-m}{2m+1}}$*

4 Test for a parametric family

Let us consider the parametric family $P = \{g_\theta(\cdot),\ \theta \in \Theta\}$ with Θ a subset of $\mathbb{R}^d$, $d \geq 1$. Our objective is to test if the intensity function q is in P. Let us consider the two following hypothesis:

$$\begin{cases} H_0 : \ \exists \theta_0 \in \Theta,\ q(\cdot) = g_\theta(\cdot) \\ H_1 : \ q \notin P \end{cases}.$$

We want to test H_0 against H_1. Under both hypothesis, one way to estimate q is to use the local polynomial estimator $\hat{q}_{\hat{h}}$. As we work in an asymptotic framework, we denote by h_n the optimal bandwidth $\hat{h}$. Under H_0, to estimate the parameter θ, let us consider the contrast

$$\begin{aligned} M_n(\theta) = \|\hat{q}_{h_n}(\cdot) - \int_0^T w\left(\cdot, h_n, \frac{X_s - \cdot}{h_n}\right) K_{h_n}(X_s - \cdot)\, \mathbf{1}_{X_s \in I} g_\theta(X_s)\, ds\|_I^2 \\ - \frac{1}{n^2} \int_I \int_0^T w^2\left(x, h_n, \frac{X_s - x}{h_n}\right) K_{h_n}^2(X_s - x)\, \mathbf{1}_{X_s \in I} dN_s dx. \end{aligned} \tag{4.10}$$

The contrast defined in (4.10) is similar to the one in [2], used in the case of the estimation of the drift and the volatility of an Itô diffusion. However, in [2], the norm is weighted by the density of X, that is l_T^x is our case. As it is important to have a good estimate of q everywhere on I, the norm is not weighted in our case. The second term in the right hand side of (4.10) is a correction of the bias. We can also notice that we use the function $\int_0^T w\left(\cdot, h_n, \frac{X_s - \cdot}{h_n}\right) K_{h_n}(X_s - \cdot)\, \mathbf{1}_{X_s \in I} g_\theta(X_s)\, ds$ and not directly g_θ in order to eliminate the bias part of the norm. An estimator of θ under H_0 is then

$$\hat{\theta}_n = \inf_{\theta \in \Theta} M_n(\theta). \tag{4.11}$$

Under classical assumptions 4.3, this estimator is consistent at a speed rate of $\sqrt{n}$, see Proposition 4.8 (i). The idea of the test is that under H_0, $M_n\left(\hat{\theta}_n\right)$ is close to $M(\theta_0)$ which is equal to 0. The rate of convergence is of order $n\sqrt{h_n}$, see Proposition 4.8 (ii). However, under H_1, $M_n\left(\hat{\theta}_n\right)$ converges to $\inf_{\theta \in \Theta} \|q - g_\theta\|_I^2$ which is different from 0 and then $nh_n M_n\left(\hat{\theta}_n\right)$ goes to ∞, see Proposition 4.8 (iii).

Assumption 4.3. *We assume that*

(i) The set Θ is compact in $\mathbb{R}^d$.

(ii) For some $M = M_I > 0$,

$$\sup_{x \in I} |g_{\theta_1}(x) - g_{\theta_2}(x)| \leq M \|\theta_1 - \theta_2\|_d \text{ for } \theta_1, \theta_2 \in \Theta$$

where $\|\cdot\|_d$ is the $\mathbb{R}^d$ Euclidian norm.

(iii) For all x in I, $\theta \mapsto g_\theta(x)$ is three times continuously differentiable. Furthermore, $x \mapsto g_\theta(x)$, $x \mapsto \partial_\theta g_\theta(x)$ are continuous on I and $\frac{\partial_{2,\theta} g_\theta}{\partial \theta_i \partial \theta_j}$, $\frac{\partial_{3,\theta} g_\theta}{\partial \theta_i \partial \theta_j \partial \theta_k}$ are bounded on I for i, j, $k = 1, .., d$ for all $\theta \in \Theta$.

(iv) *For some* $\eta = \eta_I > 0$,

$$\inf_{(x,\theta)\in I\times\theta} \lambda_{\min}\left(\partial_\theta g_\theta(x)\,\partial_\theta g_\theta(x)^T\right) \geq \eta$$

where $\lambda_{\min}(A)$ *is the smallest eigenvalue of a matrix* A *f*

(v) *The equality* $g_{\theta_1} = g_{\theta_2}$ *on* I *implies* $\theta_1 = \theta_2$.

(vi) $h_n \to 0$ *and* $n\sqrt{h_n} \to \infty$.

(vii) *The intensity function* q *is continuous on* I *and the kernel function* K *is continuous on* $\mathbb{R}$.

Proposition 4.8. *Let* $M_n(\theta)$ *and* $\hat{\theta}_n$ *defined respectively by* (4.10) *and* (4.11). *We work under Assumption 4.1, Assumption 4.2 and Assumption 4.3. On the event* $D(I,\nu)$, *under* H_0,

(i)

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\int_I \partial_\theta g_{\theta_0}(u)\,\partial_\theta g_{\theta_0}(u)^T du\right)^{-1} \int_I \frac{\partial_\theta g_{\theta_0}(u)\,\partial_\theta g_{\theta_0}(u)^T g_{\theta_0}(u)}{l_T^u} du\right),$$

(ii)

$$n\sqrt{h_n} M_n\left(\hat{\theta}_n\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 2\int_{\mathbb{R}}\left(\int_{\mathbb{R}} w(u)\,w(u+p)\,K(u)\,K(u+p)\,du\right)^2 dp \int_I \frac{(g_{\theta_0}(y))^2}{(l_T^y)^2} dy\right)$$

with

$$w(u) = U^T(0)\left(\int_{\mathbb{R}} U(z)\,U^T(z)\,K(z)\,dz\right)^{-1} U(u)$$

and under H_1,

(iii)

$$|n\sqrt{h_n} M_n\left(\hat{\theta}_n\right)| \xrightarrow{p} \infty.$$

In order to test the null hypothesis at level α, we reject H_0 when

$$|M_n(\theta_n)| \geq \hat{c}(\alpha) = n^{-1} h_n^{-\frac{1}{2}} \sqrt{\hat{V}_n}\,\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

where

$$\hat{V}_n = C(K)\int_I \left(\frac{g_{\hat{\theta}_n}(y)}{\int_0^T K_{h_n}(y - X_s)\,\mathbf{1}_{X_s\in I} ds}\right)^2 dy,$$

$$C(K) = 2\left(\int_{\mathbb{R}} K(u)\,du\right)^2 \int_{\mathbb{R}}\left(\int_{\mathbb{R}} w(u)\,w(u+p)\,K(u)\,K(u+p)\,du\right)^2 dp$$

only depends on K and Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ random variable.

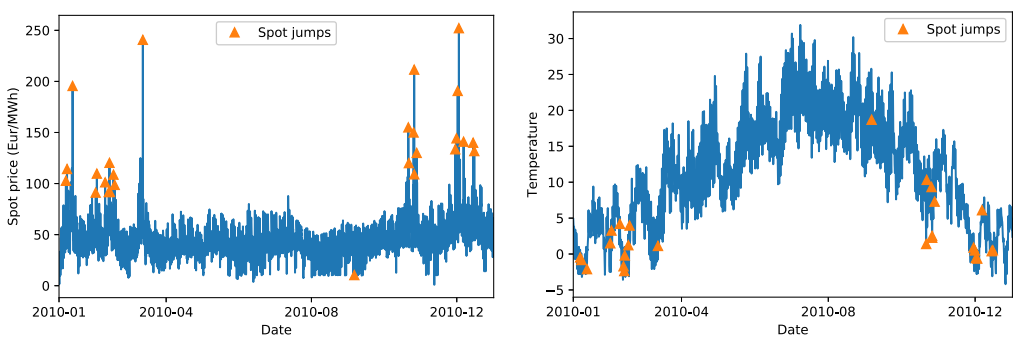
5 Dependence between the frequency of electricity spot spikes and temperature

5.1 Data

We dispose of

- the hourly French spot price between the first January of 2006 and the first of January 2017 not included,
- the hourly French temperature, which is an spatial average of the temperature over 32 cities, between the first January of 2006 and the first of January 2017 not included.

The times series are given in Figure 4.1 for the year 2010.



(a) *Spot price.* (b) *Temperature.*

Figure 4.1: *French spot price and temperature during 2010.*

5.2 Detection of the jumps

In the spot price time series, we observe spikes that are characteristic of the electricity spot market. A spike can be defined as a jump with a strong mean reversion. We then assume that the spot price S has the following dynamic:

$$S_t = Y_t + Z_t$$

with Y_t a continuous Itô semi-martingale and

$$Z_t = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}\,(dt, dx)$$

with $\underline{p}$ a Poisson measure on $\mathbb{R}^+ \times \mathbb{R}$ with compensator $\underline{q} = \lambda_t dt \otimes \nu\,(dx)$. Indeed, Z has the dynamic

$$dZ_t = -\beta Z_t dt + \int_{\mathbb{R}} x \underline{p}\,(dt, dx)$$

corresponding to a mean reverting Poisson process. Let us consider N the Poisson process associated to the jump times of X. N has intensity $(\lambda_t)_{0 \leq t \leq T}$ which is the intensity we want to estimate as

a function of the temperature. In order to detect the jumps, we use the method of Chapter 3 with a threshold equal to $5\hat{\sigma}\Delta_n^{0.49}$ where Δ_n is the frequency of observations and $\hat{\sigma}$ is the multipower variation estimator of order 20. We keep only the increments verifying $\Delta_i^n S \Delta_{i+1}^n S < 0$ as jumps. The data are segmented in periods of one year for the detection of the jumps in order to avoid too much change in the volatility. In Chapter 3, the intensity of the Poisson process is constant. Assuming that λ is bounded below and above, the results can easily be extended to the case where λ is stochastic. Jump times are represented in Figure 4.1. In the following, we consider that we observe N but we are aware that we only have an estimator of it.

5.3 Dependence with temperature

In this section, we estimate the intensity of the jump process as a function of the temperature, using the method of Section 3. In addition to the statistic interest, quanto options are financial options with temperature and spot price as underlying. They can be used for instance to hedge both volume and price risks. In order to price these options, it is necessary to capture the dependence between the temperature and the spot price. More details are given in [6] where the dependence between the two is only modeled by a correlation and the spikes are not represented.

The temperature is illustrated in Figure 4.1 along with the jump times. The spikes seems to happen more often for high and low temperatures. The observed temperature belongs to the interval $[-8.50, 33.95]$. The temperature is not observed continuously but because of the high frequency of the data and the long range of observation, we pretend that the error due to the discretization is negligible. One wants to estimate the intensity of the spike process as a function of the temperature on the interval $[-5, 33]$ where the temperature is sufficiently observed. To estimate the intensity function, we consider the Epanechnikov kernel $K(u) = \frac{3}{4}\left(1-u^2\right)\mathbf{1}_{|u|\leq 1}$ and the local polynomial estimator with degree 1 considered in Section 3. We choose $h_{\min}$ equals to $\frac{|I|\|K\|_1\|K\|_\infty|I|}{N_I} = 0.13$, where $\|K\|_1 = 1$, $\|K\|_\infty = \frac{3}{4}$ and $N_I = 219$ is the number of jumps in the interval I. The tuning parameter of the estimation procedure λ is chosen equal to 1. The optimal bandwidth is selected among the set $\mathcal{H} = \{h = h_{\min} + 0.1i,\ h_{\min} \leq 11\}$. The minimum of the criteria is achieved for $h = 8.73$ and the estimator for this value of h is given in Figure 4.2.
This result confirms our intuition: spikes happen more often when temperature is low. We now want to test the hypothesis that the intensity is a quadratic function of the temperature:

$$q(x) = a_0 \exp(a_1 x) \text{ for } x \in I$$

with $a_0 > 0$, $a_1 < 0$. The constant

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} w(u)\, w(u+p)\, K(u)\, K(u+p)\, du \right)^2 dp = \frac{413113}{985600}$$

is needed for the test. Using the test of Section 4, we find that the null hypothesis is not rejected for a level of confidence at 95% (with a p value equals to 0.083) and that the estimated parameters are $(\hat{a}_0, \hat{a}_1) = (1033.8, -0.2)$. Figure 4.2 includes the parametric estimator of the intensity as a function of the temperature. We also test if the intensity function is independent from the temperature, corresponding to q constant: the test is rejected for a level of confidence at 95% (with a p value equals to 0).

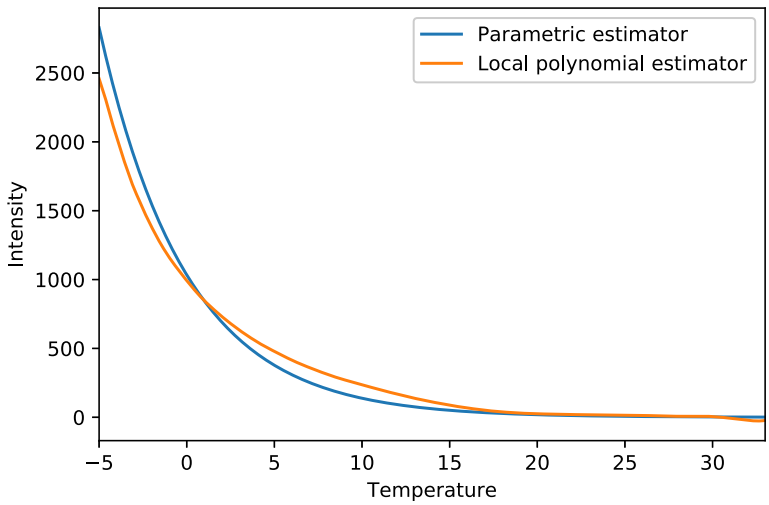


Figure 4.2: *Local polynomial estimator and parametric estimator of the intensity as a function of the temperature.*

6 Numerical results

In order to evaluate the performance of our estimation procedure, we present some simulation results. To be consistent with data, let us consider a model reproducing the temperature data and the spike times.

As in [7], we model the temperature θ_t as the sum of a trend seasonality function

$$\Gamma_t = a + bt + c_1 \sin\left(\frac{2\pi t + \tau_1}{365 \times 24}\right) + c_2 \sin\left(\frac{2\pi t + \tau_2}{24}\right)$$

corresponding to yearly and daily seasonality and a diffusion X_t having dynamics

$$dX_t = -\alpha X_t dt + \sigma dW_t$$

where W_t is a standard Brownian motion. In [7], the temperature is modeled by a CARMA process with stochastic seasonal volatility but for simplicity we consider the simplest one corresponding to an Ornstein Uhlenbeck process. Using classical estimation procedures, we find $a = 12.06$, $b = 0.0000072$, $c_1 = 7.81$, $c_2 = -3.18$, $\tau_1 = -16924.50$, $tau_2 = 10.84$, $\alpha = 0.011$, $\sigma = 0.46$.

The spike intensity λ_t is considered as an exponential function of the temperature

$$\lambda_t = a_0 \exp\left(a_1 \theta_t\right)$$

with $a_0 = 1033.8$ and $a_1 = -0.2$. As the quality of the estimation procedure depends on the interval of estimation, we consider three intervals: $[-1, 29]$, $[-3, 31]$ and $[-5, 33]$. As for the estimation on data, we consider a local polynomial estimator of degree 1 with a Epanechnikov kernel. On each interval, we apply our estimation procedure and our bandwidth selection method with $\lambda = 1$,

$h_{\min} = \frac{|I|\|K\|_1\|K\|_\infty|I|}{200}$ and $\mathcal{H} = \{h = h_{\min} + 0.1i,\ h_{\min} \leq 11\}$. To evaluate the performance on our estimator, we consider the error

$$e = \mathbb{E}\left(\frac{\int_I \left(q(x) - \hat{q}_{\hat{h}}(x)\right)^2 dx}{\int_I q^2(x)\, dx}\right)$$

where $\hat{h}$ is the optimal bandwidth given by our estimation procedure and we compare it to the oracle error

$$e_o = \min_{h \in \mathcal{H}} \mathbb{E}\left(\frac{\int_I \left(q(x) - \hat{q}_h(x)\right)^2 dx}{\int_I q^2(x)\, dx}\right).$$

In practice, we consider estimators of these errors, $\hat{e}$ and $\hat{e}_o$. In a second time, we test if the intensity is of the form $a_0 \exp(a_1 x)$ and if it is constant, that is independent of the temperature.

Results are given in Table 4.1 where 500 simulations of the model during 6 years with a step time of one hour are considered. % converged corresponds to the percentage estimators that have been computed, meaning that their local time was large enough and the matrix B invertible. This percentage diminishes with the length of the interval, and is very low for the last interval. However, this interval corresponds to the one we have estimated the parameters. One explanation is that the model does not capture all the features of the data. For instance, seasonal volatility is not modeled whereas it impacts impact the number of high and low values taken by the temperature. The two errors $\hat{e}$ and $\hat{e}_0$ increases with the length of the interval, caused by boundary effects: less values of the temperature are observed near the bounds. Furthermore, the ratio between $\hat{e}$ and $\hat{e}_0$ increases: the bandwidth selection procedure is less efficient for larger interval. This corresponds to the term $\frac{1}{\nu^2}$ in the oracle inequality. Columns % exponential and constant corresponds to the percentage of simulation for which the corresponding test has not been rejected at level 95%. Results are satisfying both for exponential and constant test. Estimators of a_0 and a_1 are consistent with the true parameters but present a small bias, probably due to the form of $M_n(\theta)$ that adds a bias term inside the norm in order to suppress the bias. The mean of $\hat{q}_{\hat{h}}$ is represented in Figure 4.3 for each interval I. One can see that there is a bias in the lower boundary for each interval.

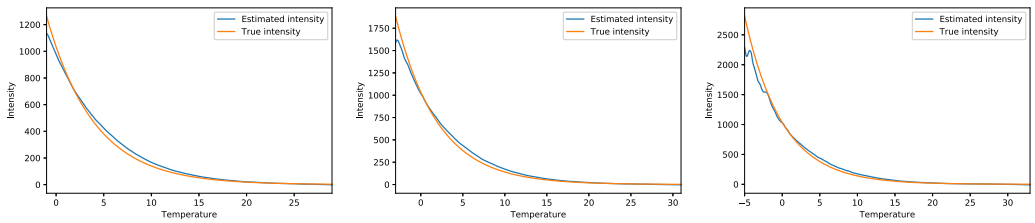
Interval	$\hat{e}$	$\hat{e}_o$	% converged	% exponential	% constant	a_0	a_1
$[-1, 29]$	0.055	0.026	100	97	0	$[1030.60, 1059, 65]$	$[-0.210, -0.202]$
$[-3, 31]$	0.082	0.03	81	97	0.25	$[1015.67, 1041.91]$	$[-0.223, -0.202]$
$[-5, 33]$	0.19	0.04	12.6	73	0	$[935.77, 973.50]$	$[-0.253, -0.2212]$

Table 4.1: *Performance of the local polynomial estimation procedure and parametrical test on different intervals for simulated data.*

7 Proof of Proposition 4.6

In order to prove Proposition 4.6, we need Proposition 4.9 and Proposition 4.10. Proposition 4.9 gives an approximation of the error by the bias and the variance. This proposition is similar to the one of [21, Proposition 4.1] in the context of density estimation.

During the proof, $\tilde{C}$ denotes a constant that can change from line to line. $\tilde{C}(\cdot)$ denotes a constant depending on $\cdot$ that can also change from line to line.



(a) $I = [-1, 29]$. (b) $I = [-3, 31]$. (c) $I = [-5, 33]$.

Figure 4.3: *Mean of local polynomial and parametric estimator for different intervals with simulated data.*

Proposition 4.9. *Assume 4.1 and 4.2. Let $x \geq 1$, $\eta \in (0, 1]$. Let $\mathcal{H}$ a finite subset of $(0, \infty)$ such that $\min \mathcal{H} = h_{\min} \geq \frac{\|K\|_\infty \|K\|_1 |I|}{n}$ and $\max \mathcal{H} \leq \frac{2}{3}|I|\Delta$. Let $\hat{q}_h$ the local polynomial estimator defined in (4.2) and $\hat{h}$ the bandwidth defined in (4.8). With conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on the event $D(I, \nu)$, for any $h \in \mathcal{H}$,*

$$\|q - \hat{q}_h\|_I^2 \leq (1 + \eta) \left(\|q - q_h\|_I^2 + V_h \right) + \tilde{C} \frac{|I| \left(1 + 2\|q\|_{I,\infty} \|l_T\|_{I,\infty} |I| \|K\|_1^2 \right) x^2}{n A_K^2 \nu^2 T^2 \eta^3}$$

and

$$\|q - q_h\|_I^2 + V_h \leq (1 + \eta) \|q - \hat{q}_h\|_I^2 + \tilde{C} \frac{|I| \left(1 + 2\|q\|_{I,\infty} \|l_T\|_{I,\infty} \| |I| K \|_1^2 \right) x^2}{n A_K^2 \nu^2 T^2 \eta^3}$$

where

$$V_h = \frac{1}{n} \int_0^T \left(\int_I \left(w \left(x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) \right)^2 \mathbf{1}_{X_s \in I} dx \right) q(X_s)\, ds$$

and A_K is a constant depending on K which is introduced in Proposition 4.4.

Proposition 4.10. *Assume 4.1 and 4.2. Let $x \geq 1$, $\theta \in (0, 1)$. Let $\mathcal{H}$ a finite subset of $(0, \infty)$ such that $\min \mathcal{H} = h_{\min} \geq \frac{\|K\|_\infty \|K\|_1 |I|}{n}$ and $\min \mathcal{H} = h_{\min} \geq \frac{\|K\|_\infty \|K\|_1}{n}$. Let $\hat{q}_h$ the local polynomial estimator defined in (4.2) and $\hat{h}$ the bandwidth defined in (4.8). Let $\hat{V}_{h,h_{\min}}$ defined in Equation (4.6) and $pen(\lambda)$ defined in Equation (4.7). With conditional probability given X larger than $1 - C_1|\mathcal{H}|e^{-x}$, on the event $D(I, \nu)$, for any $h \in \mathcal{H}$,*

$$\begin{aligned}(1 - \theta) \|\hat{q}_{\hat{h}} - q\|_I^2 \leq (1 + \theta) \|\hat{q}_h - q\|_I^2 + \left(pen_\lambda(h) - 2\hat{V}_{h,h_{\min}} \right) - \left(pen_\lambda\left(\hat{h}\right) - 2\hat{V}_{\hat{h},h_{\min}} \right) \\ + \frac{C_2}{\theta} \|q_{h_{\min}} - q\|_I^2 + \frac{C(K)|I|}{\nu^2 T^2 \theta} \left(\frac{(1 + \|q\|_{I,\infty} \|l_T\|_{I,\infty} |I|) x^2}{n} + \frac{x^3 |I|}{n^2 h_{\min}} \right) \quad (4.12)\end{aligned}$$

where C_1 and C_2 are constant and $C(K)$ is a constant depending on K.

7.1 Proof of Proposition 4.9

In the following, we work conditionally on X. We also work on the event $D(I, \nu)$. Conditionally on X, the process N is a inhomogeneous Poisson process. Proposition 4.4 and Proposition 4.5 are then verified taking the conditional probability given X.

The norm $\|q-\hat{q}_h\|_I^2$ is the sum of $\|q-q_h\|_I^2$ which is a bias term, $\|\hat{q}_h-q_h\|_I$ which is a variance term and the cross term $2<q-q_h,q_h-\hat{q}_h>_I$. In order to control $\|q-\hat{q}_h\|_I^2$ by $\|q-q_h\|_I^2+V_h$, we will control the variance term by V_h and the cross term by $\|q-q_h\|_I^2+V_h$.

Control of the variance term First, let us control the term $\|q_h-\hat{q}_h\|_I^2$. This term is equal to

$$\frac{1}{n^2}\int_I\left(\int_0^T w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)\mathbf{1}_{X_s\in I}dM_s\right)^2dx$$

which can be written as the sum of

$$\frac{1}{n^2}\int_0^T\int_I w^2\left(x,h,\frac{X_s-x}{h}\right)K_h^2\left(X_s-x\right)\mathbf{1}_{X_s\in I}dxdN_s \tag{4.13}$$

and

$$\frac{2}{n^2}\int_0^T\int_0^{s^-}\int_I w\left(x,h,\frac{X_s-x}{h}\right)w\left(x,h,\frac{X_u-x}{h}\right)K_h\left(X_u-x\right)K_h\left(X_s-x\right)dx\mathbf{1}_{X_s\in I}\mathbf{1}_{X_u\in I}dM_udM_s. \tag{4.14}$$

The term (4.13) is a simple Poisson integral and can be controlled with Proposition 4.4. As

$$\begin{aligned}\frac{1}{n^2}\int_I w^2\left(x,h,\frac{X_s-x}{h}\right)K_h^2\left(X_s-x\right)\mathbf{1}_{X_s\in I}dx &\leq \frac{\|K\|_1\|K\|_\infty|I|^2}{n^2A_K^2\nu^2T^2h}\\ &\leq \frac{|I|}{nA_K^2\nu^2T^2}\end{aligned}$$

and

$$\frac{1}{n^4}\int_0^T\left(\int_I w^2\left(x,h,\frac{X_s-x}{h}\right)K_h^2\left(X_s-x\right)\mathbf{1}_{X_s\in I}dx\right)^2d\Lambda_s$$

is bounded by

$$\begin{aligned}\frac{1}{n^2}\sup_{z\in I}\int_I w^2\left(x,h,\frac{z-x}{h}\right)K_h^2\left(z-x\right)dxV_h &\leq \frac{\|K\|_1\|K\|_\infty V_h|I|^2}{A_K^2\nu^2T^2hn^2}\\ &\leq \frac{V_h|I|}{nA_K^2\nu^2T^2},\end{aligned}$$

with conditional probability given X larger than $1-2|\mathcal{H}|e^{-u}$, on $D\left(I,\nu\right)$,

$$|\frac{1}{n^2}\int_0^T\int_I w^2\left(x,h,\frac{X_s-x}{h}\right)K_h^2\left(X_s-x\right)dx\mathbf{1}_{X_s\in I}dN_s-V_h|\leq\sqrt{\frac{2|I|x}{nA_K^2\nu^2T^2}V_h}+\frac{|I|x}{3nA_K^2\nu^2T^2}.$$

Using Young's inequality

$$2ab\leq\theta a^2+\frac{b^2}{\theta},\ \text{for all }\theta\in(0,1],$$

we find

$$|\frac{1}{n^2}\int_0^T\int_I w^2\left(x,h,\frac{X_s-x}{h}\right)K_h^2\left(X_s-x\right)dx\mathbf{1}_{X_s\in I}dN_s-V_h|\leq\theta V_h+\tilde{C}\frac{x|I|}{nA_K^2\nu^2T^2\theta}. \tag{4.15}$$

The term (4.14) is an U-statistics which can be controlled with Proposition 4.5 ; it is dominated in absolute value by

$$\tilde{C}\left(C\sqrt{x}+Dx+Bx^{\frac{3}{2}}+Ax^2\right).$$

We have

$$\begin{aligned}
A &= \frac{1}{n^2}\sup_{(u,s)\in[0,T]^2}\int_I w\left(x,h,\frac{X_s-x}{h}\right)w\left(x,h,\frac{X_u-x}{h}\right)K_h\left(X_u-x\right)K_h\left(X_s-x\right)\mathbf{1}_{X_s\in I}\mathbf{1}_{X_u\in I}dx\\
&\leq \frac{\|K\|_\infty\|K\|_1|I|^2}{n^2A_K^2h\nu^2T^2}\\
&\leq \frac{|I|}{nA_K^2\nu^2T^2}.
\end{aligned}$$

By the occupation time formula, B^2 is bounded by

$$\begin{aligned}
&\frac{|I|^4\sup_{(v,z)\in I^2}\int_I K_h\left(z-x\right)K_h\left(v-x\right)dx\sup_{u\in[0,T]}\int_I\int_I K_h\left(z-x\right)K_h\left(X_u-x\right)\mathbf{1}_{X_u\in I}dxl_T^znq\left(z\right)dz}{n^4A_K^4\nu^4T^4}\\
&\leq \frac{\|K\|_\infty\|K\|_1^3\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^4}{hn^3A_K^4\nu^4T^4}\\
&\leq \frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^3}{n^2A_K^4\nu^4T^4}.
\end{aligned}$$

Using again the occupation time formula and Young's inequality for convolutions,

$$\begin{aligned}
C^2 &\leq \frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{n^3A_K^2\nu^2T^2}\int_0^T\int_I\left(\int_I w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)\mathbf{1}_{X_s\in I}K_h\left(z-x\right)dx\right)^2dzd\Lambda_s\\
&\leq \frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{n^3A_K^2\nu^2T^2}\int_0^T\|w\left(\cdot,h,\frac{X_s-\cdot}{h}\right)K_h\left(X_s-\cdot\right)\mathbf{1}_{X_s\in I}\mathbf{1}_{\cdot\in I}*K_h\left(\cdot\right)\|^2d\Lambda_s\\
&\leq \frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{n^3A_K^2\nu^2T^2}\int_0^T\|w\left(\cdot,h,\frac{X_s-\cdot}{h}\right)K_h\left(X_s-\cdot\right)\mathbf{1}_{X_s\in I}\mathbf{1}_{\cdot\in I}\|^2\|K_h\|_1^2d\Lambda_s\\
&= \frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2V_h|I|^2}{nA_K^2\nu^2T^2}.
\end{aligned}$$

Using twice Cauchy Schwarz inequality, using the fact that $\int_0^T a_u^2d\Lambda_u=1$ and $\int_0^T b_s^2d\Lambda_s=1$, D can be bounded in the same way than C and

$$D\leq\sqrt{\frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2V_h|I|^2}{nA_K^2\nu^2T^2}}.$$

We then have for $x\geq 1$, for all $h\in\mathcal{H}$, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D\left(I,\nu\right)$,

$$\begin{aligned}
&|\int_0^T\int_0^{s^-}\int_I w\left(x,h,\frac{X_u-x}{h}\right)w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_u-x\right)\mathbf{1}_{X_u\in I}K_h\left(X_s-x\right)\mathbf{1}_{X_s\in I}dxdM_udM_s|\\
&\leq \theta V_h+\tilde{C}\frac{|I|\left(1+2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{\nu^2T^2A_K^2n\theta}.
\end{aligned}\tag{4.16}$$

Combining (4.15) and (4.16) with $\theta = \frac{\eta}{2}$, with conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$,

$$|\|q_h - \hat{q}_h\|_I^2 - V_h| \leq \eta V_h + \tilde{C}\frac{|I|\left(1 + 2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{nA_K^2\nu^2T^2\eta}. \tag{4.17}$$

Control of the cross term The cross term $< \hat{q}_h - q_h, q_h - q >_I$ is equal to

$$\frac{1}{n}\int_0^T \left(\int_I (q_h(x) - q(x))\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx\right) dM_s$$

and then it can be controlled using Proposition 4.4. Using Cauchy Schwarz,

$$\begin{aligned}\frac{1}{n}\int_I (q_h(x) - q(x))\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx &\leq \frac{1}{n}\|q_h - q\|_I \sqrt{\frac{\|K\|_1\|K\|_\infty|I|^2}{hA_K^2\nu^2T^2}} \\ &\leq \|q_h - q\|_I \sqrt{\frac{|I|}{nA_K^2\nu^2T^2}}\end{aligned}$$

and

$$\|q_h - q\|_I \sqrt{\frac{|I|}{nA_K^2\nu^2T^2}} x \leq \frac{\theta\|q_h - q\|_I^2}{2} + \frac{|I|x^2}{2n\theta A_K^2\nu^2T^2}.$$

Using Cauchy Schwarz inequality again,

$$\frac{1}{n^2}\int_0^T \left(\int_I (q_h(x) - q(x))\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx\right)^2 d\Lambda_s \leq \|q_h - q\|_I^2 V_h.$$

We can also bound this term using the occupation time formula and Young's inequality for convolution by

$$\frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}}{nA_K^2\nu^2T^2}\|K_h(\cdot) * (q_h - q)(\cdot)\, \mathbf{1}_{\cdot \in I}\|^2 \leq \frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2|I|^2}{nA_K^2\nu^2T^2}\|q_h - q\|_I^2$$

Thus,

$$\frac{1}{n^2}\int_0^T \left(\int_I (q_h(x) - q(x))\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx\right)^2 d\Lambda_s$$

is bounded by

$$\|q - q_h\|_I^2 \sqrt{V_h} \sqrt{\frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2|I|^2}{nA_K^2\nu^2T^2}}$$

and for $\theta, u > 0$,

$$\sqrt{\frac{2}{n^2}\int_0^T \left(\int_I (q_h(x) - q(x))\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx\right)^2 d\Lambda_s x}$$

is bounded by

$$\theta\|q-q_h\|_I^2+\frac{1}{2\theta}\sqrt{V_h}\sqrt{\frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2|I|^2x}{nA_K^2n\nu^2T^2}}\leq\theta\|q-q_h\|_I^2+\frac{uV_h}{\theta}+\frac{x\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2|I|^2}{16\theta uA_K^2n\nu^2T^2}.$$

Thus, for $\theta\in(0,1)$, $x\geq 1$, with $u=\theta^2$, we obtain, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h\in\mathcal{H}$,

$$\sqrt{\frac{2}{n^2}\int_0^T\left(\int_I\left(q_h(x)-q(x)\right)w\left(x,h,\frac{X_s-x}{h}\right)K_h\left(X_s-x\right)\mathbf{1}_{X_s\in I}dx\right)^2d\Lambda_s x}$$

is bounded by

$$\theta\left(\|q-q_h\|_I^2+V_h\right)+\tilde{C}\frac{x\|l_T\|_{I,\infty}\|q\|_{I,\infty}\|K\|_1^2|I|^2}{\theta^3nA_K^2\nu^2T^2}$$

and using Proposition 4.4, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h\in\mathcal{H}$,

$$2|<q_h-\hat{q}_h,q_h-q>_I|\leq 3\theta\left(\|q-q_h\|_I^2+V_h\right)+\tilde{C}\frac{|I|\left(1+2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{nA_K^2\nu^2T^2\theta^3}. \quad (4.18)$$

Combining (4.17) and (4.18), we find, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h\in\mathcal{H}$,

$$\|q-\hat{q}_h\|_I-\|q-q_h\|_I^2\leq 3\theta\|q-q_h\|_I^2+(1+4\theta)V_h+\tilde{C}\frac{|I|\left(1+2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{nA_K^2\nu^2T^2\theta^3}$$

and

$$\|q-\hat{q}_h\|_I-\|q-q_h\|_I^2\geq -3\theta\left(\|q-q_h\|_I^2+V_h\right)+(1-\theta)V_h-\tilde{C}\frac{|I|\left(1+2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{nA_K^2\nu^2T^2\theta^3}.$$

Taking $\theta=\frac{\eta}{4}$, the proof is achieved.

7.2 Proof of Proposition 4.10

We continue to work conditionally on X and we work on the event $D(I,\nu)$. The beginning of the proof is similar to the one of [20, Theorem 9]. Let $\theta'\in(0,1)$. For any $h\in\mathcal{H}$,

$$\begin{aligned}\|\hat{q}_{\hat{h}}-q\|_I^2+\mathrm{pen}_\lambda\left(\hat{h}\right)&=\|\hat{q}_{\hat{h}}-\hat{q}_{h_{\min}}\|_I^2+\mathrm{pen}_\lambda\left(\hat{h}\right)+\|\hat{q}_{h_{\min}}-q\|_I^2+2<\hat{q}_{\hat{h}}-\hat{q}_{h_{\min}},\hat{q}_{h_{\min}}-q>_I\\&\leq\|\hat{q}_h-\hat{q}_{h_{\min}}\|_I^2+\mathrm{pen}_\lambda(h)+\|\hat{q}_{h_{\min}}-q\|_I^2+2<\hat{q}_{\hat{h}}-\hat{q}_{h_{\min}},\hat{q}_{h_{\min}}-q>_I\\&\leq\|\hat{q}_h-q\|_I^2+2\|q-\hat{q}_{h_{\min}}\|_I^2+2<\hat{q}_h-q,q-\hat{q}_{h_{\min}}>_I+\mathrm{pen}_\lambda(h)\\&\quad+2<\hat{q}_{\hat{h}}-\hat{q}_{h_{\min}},\hat{q}_{h_{\min}}-q>_I.\end{aligned}$$

We then have

$$\begin{aligned}\|\hat{q}_{\hat{h}}-q\|_I^2\leq\|\hat{q}_h-q\|_I^2&+\left(\mathrm{pen}_\lambda(h)-2<\hat{q}_h-q,\hat{q}_{h_{\min}}-q>_I\right)\\&-\left(\mathrm{pen}_\lambda\left(\hat{h}\right)-2<\hat{q}_{\hat{h}}-q,\hat{q}_{h_{\min}}-q>_I\right).\end{aligned} \quad (4.19)$$

We then want to approach $< \hat{q}_h - q, \hat{q}_{h_{\min}} - q >_I$ by $\hat{V}_{h,h_{\min}}$. Let us define

$$S(h,h') =< \hat{q}_h - q_h, q_{h'} - q >_I .$$

We have

$$\begin{aligned} < \hat{q}_h - q, \hat{q}_{h_{\min}} - q >_I &=< \hat{q}_h - q_h + q_h - q, \hat{q}_{h_{\min}} - q_{h_{\min}} + q_{h_{\min}} - q >_I \\ &=< \hat{q}_h - q_h, \hat{q}_{h_{\min}} - q_{h_{\min}} >_I + S(h, h_{\min}) + S(h_{\min}, h) \\ &+ < q_h - q, q_{h_{\min}} - q >_I . \end{aligned}$$

Furthermore, we can easily show that

$$< \hat{q}_h - q_h, \hat{q}_{h_{\min}} - q_{h_{\min}} >_I = \hat{V}_{h,h_{\min}} + U(h, h_{\min})$$

where

$$\hat{V}_{h,h_{\min}} = \frac{1}{n^2} \int_0^T \int_I w\left(x, h, \frac{X_s - x}{h}\right) w\left(x, h_{\min}, \frac{X_s - x}{h_{\min}}\right) K_h(X_s - x) K_{h_{\min}}(X_s - x) \mathbf{1}_{X_s \in I} dx dN_s,$$

$$U(h, h_{\min}) = \frac{1}{n^2} \int_0^T \int_0^{s^-} \left(G_{h,h_{\min}}(X_u, X_s) + G_{h_{\min},h}(X_s, X_u)\right) dM_u dM_s$$

and

$$G_{h,h'}(X_u, X_s) = \int_I w\left(x, h, \frac{X_u - x}{h}\right) w\left(x, h_{\min}, \frac{X_s - x}{h_{\min}}\right) K_h(X_u - x) \mathbf{1}_{X_u \in I} K_{h_{\min}}(X_s - x) \mathbf{1}_{X_s \in I} dx.$$

Thus,

$$< \hat{q}_h - q, \hat{q}_{h_{\min}} - q >_I = \hat{V}_{h,h_{\min}} + U(h, h_{\min}) + S(h, h_{\min}) + S(h_{\min}, h) + < q_h - q, q_{h_{\min}} - q >_I . \tag{4.20}$$

Control of $U(h, h_{\min})$ The term $U(h, h_{\min})$ can be controlled using Proposition 4.5. Let $x \geq 1$. With conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

$$|U(h, h_{\min})| \leq \tilde{C}\left(C\sqrt{x} + Dx + Bx^{\frac{3}{2}} + Ax^2\right).$$

We have

$$\begin{aligned} A &= \frac{1}{n^2} \sup_{(u,v) \in I^2} |G_{h,h_{\min}}(u,v) + G_{h_{\min},h}(v,u)| \\ &\leq \frac{2}{n^2} \sup_{(u,v) \in I^2} |G_{h,h_{\min}}(u,v)| \\ &\leq \frac{2\|K\|_\infty \|K\|_1 |I|^2}{A_K^2 \nu^2 T^2 n^2 h_{\min}}. \end{aligned}$$

Using Cauchy Schwarz inequality,

$$\frac{1}{n^4}\int_0^s \left(G_{h,h_{\min}}\left(X_u,X_s\right)\right)^2 d\Lambda_u \leq \frac{V_h\|K\|^2|I|^2}{n^2A_K^2\nu^2T^2h_{\min}}.$$

Using the occupation time formula and Young's inequality for convolutions,

$$\begin{aligned}\frac{1}{n^4}\int_0^u \left(G_{h,h_{\min}}\left(X_u,X_s\right)\right)^2 d\Lambda_s &\leq \frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^4\|K_h\left(X_u-\cdot\right)*K_{h_{\min}}\left(\cdot\right)\|^2}{n^3A_K^4\nu^4T^4}\\ &\leq \frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}\|K\|_1^2\|K\|^2|I|^4}{h_{\min}n^3A_K^4\nu^4T^4}.\end{aligned}$$

Thus,

$$\begin{aligned}B^2 &\leq \frac{V_h\|K\|^2|I|^2}{A_K^2n^2\nu^2T^2h_{\min}}+\frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}\|K\|_1^2\|K\|^2|I|^4}{h_{\min}A_K^4n^3\nu^4T^4},\\ B &\leq \sqrt{\frac{V_h\|K\|^2|I|^2}{A_K^2n^2\nu^2T^2h_{\min}}}+\sqrt{\frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^4\|K\|_1^2\|K\|^2}{n^3A_K^4h_{\min}\nu^4T^4}}\end{aligned}$$

and

$$\begin{aligned}Bx^{\frac{3}{2}} &\leq \frac{\theta'}{3}V_h+\frac{6}{\theta'}\frac{\|K\|^2|I|^2x^3}{A_K^2n^2\nu^2T^2h_{\min}}+\frac{\|K\|^2|I|^2x^3}{2h_{\min}A_K^2n^2\nu^2T^2}+\frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2}{2nA_K^2\nu^2T^2}\\ &\leq \frac{\theta'}{3}V_h+\frac{13}{2\theta'}\frac{\|K\|^2|I|^2x^3}{A_K^2n^2\nu^2T^2h_{\min}}+\frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2}{2\theta'nA_K^2\nu^2T^2}.\end{aligned}$$

To dominate C^2 which is bounded by

$$\frac{2}{n^4}\int_0^T\int_0^T \left(G_{h,h_{\min}}\left(X_u,X_s\right)\right)^2 d\Lambda_s d\Lambda_u, \tag{4.21}$$

we use the occupation time formula and Young's inequality for convolutions:

$$\begin{aligned}C^2 &\leq \frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2}{n^3A_K^2\nu^2T^2}\int_0^T \|w\left(\cdot,h,\frac{X_u-\cdot}{h}\right)K_h\left(X_u-\cdot\right)\mathbf{1}_{X_u\in I}\mathbf{1}_{\cdot\in I}*K_{h_{\min}}\left(\cdot\right)\|^2 d\Lambda_u\\ &\leq \frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}\|K\|_1^2|I|^2}{n^3A_K^2\nu^2T^2}\int_0^T \|w\left(\cdot,h,\frac{X_u-\cdot}{h}\right)K_h\left(X_u-\cdot\right)\mathbf{1}_{\cdot\in I}\mathbf{1}_{X_u\in I}\|^2 d\Lambda_u\\ &= \frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}\|K\|_1^2|I|^2}{A_K^2n\nu^2T^2}V_h.\end{aligned}$$

We then have

$$C\sqrt{x}\leq \theta'V_h+\tilde{C}\frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2x}{nA_K^2\nu^2T^2\theta'}.$$

Using twice Cauchy Schwarz inequality, we find that D is bounded by (4.21) and then

$$Dx\leq \theta'V_h+\tilde{C}\frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2x^2}{nA_K^2\nu^2T^2\theta'}.$$

Finally, if $x\geq 1$ and $\theta'\in(0,1)$, with conditional probability given X greater than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D\left(I,\nu\right)$,

$$|U\left(h,h_{\min}\right)|\leq \theta'V_h+\tilde{C}\frac{\|K\|_1^2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|^2x^2}{n\nu^2T^2\theta'}+\tilde{C}\frac{\|K\|_\infty\|K\|_1|I|^2x^3}{\nu^2T^2A_K^2\theta'n^2h_{\min}}. \tag{4.22}$$

Control of S We need to control $S(h, h_{\min})$ and $S(h_{\min}, h)$. Let h, h' in $\mathcal{H}$. We can write

$$S(h, h') = \frac{1}{n}\int_0^T \int_I (q_{h'} - q)\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx dM_s$$

and it is possible to control it with Proposition 4.4. First, using occupation time formula and Proposition 4.2,

$$\begin{aligned}\frac{1}{n}\int_I (q_{h'} - q)(x)\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx &\leq \frac{|I|}{nA_K \nu T}\|K\|_1 \|q - q_{h'}\|_{I,\infty} \\ &\leq \frac{2\|K\|_1^2 \|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{nA_K^2 \nu^2 T^2}.\end{aligned}$$

Then, using the occupation time formula and Young's inequality for convolutions, the term

$$\frac{1}{n^2}\int_0^T \left(\int_I (q_{h'} - q)(x)\, w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x)\, \mathbf{1}_{X_s \in I} dx\right)^2 d\Lambda_s$$

is bounded by

$$\frac{\|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{nA_K^2 \nu^2 T^2}\| (q_{h'} - q)(\cdot)\, \mathbf{1}_{\cdot \in I} * K_h(\cdot) \|^2 \leq \frac{\|q_{h'} - q\|_I^2 \|K\|_1^2 \|l_T\|_{I,\infty}\|q\|_{I,\infty}|I|^2}{nA_K^2 \nu^2 T^2}.$$

With conditional probability given X larger than $1 - \tilde{C}e^{-x}$, on $D(I, \nu)$,

$$S(h, h') \leq \frac{\theta'}{2}\|q_{h'} - q\|_I^2 + \tilde{C}\frac{\|K\|_1^2 \|l_T\|_{I,\infty}|I|^2 \|q\|_{I,\infty} x}{nA_K^2 \nu^2 T^2 \theta'}. \tag{4.23}$$

We apply the previous inequality for $S(h, h_{\min})$ and $S(h_{\min}, h)$.

Control of $< q_h - q, q_{h_{\min}} - q >_I$ We have:

$$| < q_h - q, q_{h_{\min}} - q >_I | \leq \frac{\theta'}{2}\|q_h - q\|_I^2 + \frac{1}{2\theta'}\|q_{h_{\min}} - q\|_I^2. \tag{4.24}$$

At the end, combining (4.20), (4.22), (4.23) and (4.24), we have with conditional probability given X higher than $1 - \tilde{C}|\mathcal{H}|e^{-x}$ for any $h \in \mathcal{H}$, on $D(I, \nu)$,

$$\begin{aligned}| < \hat{q}_h - q, \hat{q}_{h_{\min}} - q >_I - \hat{V}_{h,h_{\min}}| &\leq \theta'\left(\|q_h - q\|_I^2 + V_h\right) + \left(\frac{\theta'}{2} + \frac{1}{2\theta'}\right)\|q_{h_{\min}} - q\|_I^2 \\ &\quad + \frac{\tilde{C}(K)|I|}{\nu^2 T^2 \theta'}\left(\frac{\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|x^2}{n} + \frac{x^3|I|}{n^2 h_{\min}}\right).\end{aligned} \tag{4.25}$$

Furthermore, Proposition 4.9 states that with conditional probability given X higher than $1 - \tilde{C}|\mathcal{H}|e^{-x}$ for any $h \in \mathcal{H}$, on $D(I, \nu)$,

$$\|q - q_h\|_I^2 + V_h \leq 2\|q - \hat{q}_h\|_I^2 + \tilde{C}\frac{|I|\left(1 + 2\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\|K\|_1^2\right)x^2}{n}. \tag{4.26}$$

We combine (4.19), (4.25) and (4.26) and we take $\theta' = \frac{\theta}{4}$ to conclude.

7.3 Proof of Proposition 4.6

Let $\tau = \lambda - 1$, $\epsilon \in (0,1)$ and $\theta \in (0,1)$. Using Proposition 4.10, with conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$(1-\theta)\left\|\hat{q}_{\hat{h}} - q\right\|_I^2 + \tau\hat{V}_{\hat{h}} \leq (1+\theta)\left\|\hat{q}_h - q\right\|_I^2 + \tau\hat{V}_h + \frac{\tilde{C}}{\theta}\|q_{h_{\min}} - q\|_I^2 + \frac{\tilde{C}(K)|I|}{\nu^2T^2\theta}\left(\frac{(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n} + \frac{x^3|I|}{n^2h_{\min}}\right)$$

Equation (4.15) states that with conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any h,

$$|\hat{V}_h - V_h| \leq \frac{\theta}{2}V_h + \tilde{C}\frac{x|I|}{nA_K^2\nu^2T^2\theta}.$$

First, let us consider the case $\tau \geq 0$. We then have, with conditional probability given X larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$(1-\theta)\left\|\hat{q}_{\hat{h}} - q\right\|_I^2 \leq (1+\theta)\left\|\hat{q}_h - q\right\|_I^2 + \tau\left(1+\frac{\theta}{2}\right)V_h + \frac{\tilde{C}}{\theta}\|q_{h_{\min}} - q\|_I^2 + \frac{\tilde{C}(K)|I|}{\nu^2T^2\theta}\left(\frac{(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n} + \frac{|I|x^3}{n^2h_{\min}}\right).$$

Using Proposition (4.9) with $\eta = \frac{\theta}{2+\theta}$, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$\tau V_h \leq \tau\left(1+\frac{\theta}{2+\theta}\right)\|q - \hat{q}_h\|_I^2 + \tau\frac{\tilde{C}(K)|I|(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n\theta^3\nu^2T^2}.$$

As $\left(1+\frac{\theta}{2}\right)\left(1+\frac{\theta}{2+\theta}\right) = 1+\theta$, with probability larger than $1 - \tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$,

$$(1-\theta)\left\|\hat{q}_{\hat{h}} - q\right\|_I^2 \leq (1+\theta+(1+\theta)\tau)\left\|\hat{q}_h - q\right\|_I^2 + \frac{\tilde{C}}{\theta}\|q_{h_{\min}} - q\|_I^2 + \frac{|I|}{\nu^2T^2}\left(\frac{\tilde{C}(K)}{\theta} + \frac{\tilde{C}(K)}{\theta^3}\right)\left(\frac{(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n} + \frac{|I|x^3}{n^2h_{\min}}\right).$$

With $\theta = \frac{\epsilon}{\epsilon+2+2\tau}$, we have

$$\left\|\hat{q}_{\hat{h}} - q\right\|_I^2 \leq (1+\tau+\epsilon)\left\|\hat{q}_h - q\right\|_I^2 + \frac{\tilde{C}(\epsilon+2+2\tau)^2}{(2+2\tau)\epsilon}\|q_{h_{\min}} - q\|_I^2 + \frac{\tilde{C}(K,\epsilon,\tau)|I|}{\nu^2T^2}\left(\frac{(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|)x^2}{n} + \frac{x^3|I|}{n^2h_{\min}}\right).$$

If $-1 < \tau \leq 0$, according to (4.15),

$$|\hat{V}_{\hat{h}} - V_{\hat{h}}| \leq \frac{\theta}{2}V_{\hat{h}} + \tilde{C}\frac{x|I|}{nA_K^2\nu^2T^2\theta}.$$

and according to Proposition 4.9,

$$\tau V_{\hat{h}} \leq \tau \left(1+\frac{\theta}{2+\theta}\right)\|q-\hat{q}_h\|_I^2 + \tau \frac{\tilde{C}(K)|I|\left(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\right)x^2}{n\theta^3\nu^2T^2}.$$

We find

$$\begin{aligned}(1-\theta+\tau(1+\theta))\|\hat{q}_{\hat{h}}-q\|_I^2 &\leq (1+\theta)\|\hat{q}_h-q\|_I^2 + \frac{C_2}{\theta}\|q_{h_{\min}}-q\|_I^2 \\ &+ \frac{|I|}{\nu^2T^2}\left(\frac{\tilde{C}(K)}{\theta}+\frac{\tilde{C}(K)}{\theta^3}\right)\left(\frac{\left(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\right)x^2}{n}+\frac{x^3|I|}{n^2h_{\min}}\right).\end{aligned}$$

Taking $\theta = \frac{\left(\epsilon(\tau+1)^2\right)}{2+\epsilon(1-\tau^2)} < 1$, we obtain, with conditional probability given X larger than $1-\tilde{C}|\mathcal{H}|e^{-x}$, on $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$\begin{aligned}\|\hat{q}_{\hat{h}}-q\|_I^2 &\leq \left(\frac{1}{1+\tau}+\epsilon\right)\|\hat{q}_h-q\|_I^2 + \tilde{C}(\epsilon,\tau)\|q_{h_{\min}}-q\|_I^2 \\ &+ \tilde{C}(K,\epsilon,\tau)|I|\left(\frac{\left(1+\|q\|_{I,\infty}\|l_T\|_{I,\infty}|I|\right)x^2}{n}+\frac{x^3|I|}{n^2h_{\min}}\right).\end{aligned}$$

Inequality (4.12) is then verified.

Now, let us use (4.12) with $x = 5\log(n\vee|\mathcal{H}|)$ and $\epsilon = \tilde{C}(\log(n))^{-1}$. Let $\mathcal{E}$ be the event on which (4.12) is true. Integrating with respect to X and dividing by $\mathbb{P}(D(I,\nu))$, we find

$$\begin{aligned}&\mathbb{E}\left(\|\hat{q}_{\hat{h}}-q\|_I^2\mathbf{1}_{\mathcal{E}}|D(I,\nu)\right) \leq \left(\lambda\vee\frac{1}{\lambda}+\tilde{C}(\log(n))^{-1}\right)\min_{h\in\mathcal{H}}\mathbb{E}\left(\|\hat{q}_h-q\|_I^2|D(I,\nu)\right) \\ &+\tilde{C}(\lambda)\log(n)\,\mathbb{E}\left(\|q_{h_{\min}}-q\|_I^2|D(I,\nu)\right) \\ &+\frac{\tilde{C}(K,\lambda)|I|\log(n)^3}{\nu^2T^2}\left(\frac{25|I|\|q\|_{I,\infty}\mathbb{E}\left(\|l_T\|_{I,\infty}|D(I,\nu)\right)\log(n\vee|\mathcal{H}|)^2}{n}+\frac{125|I|\log(n\vee|\mathcal{H}|)^3}{n^2h_{\min}}\right).\end{aligned} \tag{4.27}$$

Let $N_I = \int_0^T \mathbf{1}_{X_s\in I}dN_s$. On $\mathcal{E}^c\cap D(I,\nu)$, using Cauchy Schwarz,

$$\begin{aligned}\|\hat{q}_{\hat{h}}-q\|_I^2 &\leq 2\|q\|_I^2 + \frac{2|I|^2}{\nu^2T^2A_K^2n^2}N_I\int_I\int_0^T K_h^2(X_s-x)\mathbf{1}_{X_s\in I}dN_sdx \\ &\leq 2\|q\|_I^2 + \frac{2N_I^2\|K\|_\infty\|K\|_1|I|^2}{\nu^2T^2A_K^2n^2h} \\ &\leq 2\|q\|_I^2 + \frac{2N_I^2|I|}{\nu^2T^2A_K^2n}.\end{aligned}$$

Using Cauchy Schwarz,

$$\mathbb{E}\left(N_I^2\mathbf{1}_{\mathcal{E}^c}|X\right) \leq \mathbb{E}\left(N_I^4|X\right)^{\frac{1}{2}}\mathbb{P}\left(\mathcal{E}^c|X\right)^{\frac{1}{2}}.$$

Using Laplace transform formula, we easily show that N_I has the law of a Poisson random variable with parameter $n\int_0^T q(X_s)\mathbf{1}_{X_s\in I}ds$ conditionally on X and

$$\mathbb{E}\left(N_I^4|X\right) \leq \tilde{C}n^4\left(\|q\|_{I,\infty}T+(\|q\|_{I,\infty}T)^2+(\|q\|_{I,\infty}T)^3+(\|q\|_{I,\infty}T)^4\right).$$

We also have

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}^c|X\right) &\leq \tilde{C}\frac{|\mathcal{H}|}{n^5 \vee |\mathcal{H}|^5}\\ &\leq \frac{\tilde{C}}{n^4}.\end{aligned}$$

Integrating with respect to X and dividing by $\mathbb{P}\left(D\left(I,\nu\right)\right)$, we find

$$\mathbb{E}\left(\|\hat{q}_{\hat{h}} - q\|_I^2 \mathbf{1}_{\mathcal{E}^c}|D\left(I,\nu\right)\right) \leq \tilde{C}\frac{\|q\|_I^2}{n^4} + \tilde{C}\left(K\right)\frac{|I|\sqrt{\sum_{i=1}^4 \left(\|q\|_{\infty,I}T\right)^i}}{\nu^2 T^2 n}. \tag{4.28}$$

We obtain (4.9) combining (4.27) and (4.28).

8 Other proofs

8.1 Proof of Proposition 4.1

Under Assumption 4.1 (i), we can define a local time in the sense of the continuous semimartingale L_t^x continuous in t, cadlag in x, see [26, Chapter 6] for more information. [26, Exercise 1.15] states that for every measurable function h on $[0,T]\times\Omega\times\mathbb{R}$.

$$\int_0^t h\left(s,X_s\right)d < X_s, X_s >= \int_{\mathbb{R}} da \int_0^t h\left(s,a\right)dL_s^a.$$

Let f be a measurable function on $\Omega\times\mathbb{R}$. As $\sigma_s \geq \underline{\sigma} > 0$ for every $s\in[0,T]$ almost surely and $d < X_s, X_s >= \sigma_s^2 ds$, we have

$$\int_0^t f\left(X_s\right)ds = \int_{\mathbb{R}} f\left(a\right)da \int_0^t \frac{1}{\sigma_s^2}dL_s^a$$

and (i) is verified with $l_T^x = \int_0^T \frac{1}{\sigma_s^2}dL_s^a$. According to [4, Equation (III)$_\gamma$], as

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|\int_0^t \sigma_s dW_s| + \int_0^T |\mu_s|ds\right) < \infty,$$

$$\mathbb{E}\left(\sup_{x\in\mathbb{R}} L_T^x\right) < \infty$$

and because $\sigma_s \geq \underline{\sigma}$ a.s., we obtain (ii). The continuity of $x\mapsto l_T^x$ follows from [32, Example 2.2.3 (a)] and $\sigma_s^2 > 0$ for all $s\in[0,T]$.

Under Assumption 4.1 (ii), we have $l_t^x = \mathbf{1}_{x\in[0,t]}$ for all x in $\mathbb{R}$ and t in $[0,T]$.

8.2 Proof of Proposition 4.3

Let $\|\cdot\|_k$ denoting the Euclidian norm on $\mathbb{R}^k$ for $k\in\mathbb{N}$. We work on $D\left(I,\nu\right)$. We have

$$\lambda_{\min}\left(B\left(x,h\right)\right) = \inf_{\|v\|_m=1} v^T B\left(x,h\right)v.$$

Let v in $\mathbb{R}^m$ with $\|v\|_m = 1$. Using the occupation time formula, we have

$$\begin{aligned} v^T B(x,h) v &= \int_0^T \left(v^T U\left(\frac{X_s - x}{h}\right)\right)^2 K_h(X_t - x) \mathbf{1}_{X_s \in I} ds \\ &= \int_I \left(v^T U\left(\frac{u - x}{h}\right)\right)^2 K_h(u - x) l_T^u du \\ &\geq \frac{\nu T K_{\min}}{|I| h} \int_I \left(v^T U\left(\frac{u - x}{h}\right)\right)^2 \mathbf{1}_{|u-x|<\Delta h} du \end{aligned}$$

If $x \in [\min I + \Delta h, \max I - \Delta h]$,

$$\frac{1}{h}\int_I \left(v^T U\left(\frac{u - x}{h}\right)\right)^2 \mathbf{1}_{|u-x|<h} du = \int_{\mathbb{R}} \left(v^T U(u)\right)^2 \mathbf{1}_{|u|<\Delta} du,$$

if $x \in [\min I, \min I + \Delta h]$, because $h \leq \frac{2}{3}\Delta|I|$,

$$\frac{1}{h}\int_I \left(v^T U\left(\frac{u - x}{h}\right)\right)^2 \mathbf{1}_{|u-x|<h} du \geq \int_{\mathbb{R}} \left(v^T U(u)\right)^2 \mathbf{1}_{\frac{-\Delta}{2}<u<0} du$$

and if $x \in [\max I - h, \max I]$,

$$\frac{1}{h}\int_I \left(v^T U\left(\frac{u - x}{h}\right)\right)^2 \mathbf{1}_{|u-x|<h} du \geq \int_{\mathbb{R}} \left(v^T U(u)\right)^2 \mathbf{1}_{0<u<\frac{\Delta}{2}} du.$$

Thus, for all $x \in I$, $\lambda_{\min}(B(x,h))$ is greater than

$$\begin{aligned} &\frac{\nu T}{|I|} K_{\min} \min\left(\inf_{\|v\|_m=1}\left(\int_{\mathbb{R}} \left(v^T U(u)\right)^2 \mathbf{1}_{0<u<\frac{\Delta}{2}} du\right), \inf_{\|v\|_m=1}\left(\int_{\mathbb{R}} \left(v^T U(u)\right)^2 \mathbf{1}_{\frac{-\Delta}{2}<u<0} du\right)\right) \\ &= \frac{\nu T}{|I|} K_{\min} \min\left(\lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{0<u<\frac{\Delta}{2}} du\right), \lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{\frac{-\Delta}{2}<u<0} du\right)\right) \end{aligned}$$

and

$$\min\left(\lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{0<u<\frac{\Delta}{2}} du\right), \lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{\frac{-\Delta}{2}<u<0} du\right)\right) > 0$$

applying [30, Lemma 1.4] with $K(u) = \mathbf{1}_{0 \leq u \leq \frac{\Delta}{2}}$ and $K(u) = \mathbf{1}_{-\frac{\Delta}{2} \leq u \leq 0}$.
We have

$$\begin{aligned} |w\left(x, h, \frac{X_s - x}{h}\right) \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1}| &\leq \|B(x,h)^{-1} U\left(\frac{X_s - x}{h}\right)\|_{m+1} \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1} \\ &\leq \|B(x,h)^{-1}\|_{m+1} \| \|U\left(\frac{X_s - x}{h}\right)\|_{m+1} \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1} \\ &\leq \frac{1}{\lambda_{\min}(B(x,h))} \sqrt{1 + \frac{1}{(1!)^2} + \frac{1}{(2!)^2} + ... + \frac{1}{(m!)^2}}. \end{aligned}$$

Then,

$$|w\left(x, h, \frac{X_s - x}{h}\right) \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1}| \leq \frac{|I|}{A_K \nu T} \tag{4.29}$$

with

$$A_K = \frac{K_{\min} \min\left(\lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{0<u<\frac{\Delta}{2}} du\right), \lambda_{\min}\left(\int_{\mathbb{R}} U(u) U^T(u) \mathbf{1}_{-\frac{\Delta}{2}<u<0} du\right)\right)}{2}.$$

8.3 Proof of Proposition 4.7

Lower bound. We suppose for simplicity that $I = [0,1]$. As in [30, Section 2.6.1], we consider a real number $c_0 > 0$ and

$$m = \lfloor c_0 n^{\frac{1}{2\beta+1}} \rfloor + 1,\ h_n = \frac{1}{m},\ x_k = \frac{k-\frac{1}{2}}{m},$$
$$\varphi_k(x) = L h_n^{\beta} \overline{K}\left(\frac{x - x_k}{h_n}\right),\ k = 1, .., m,\ x \in [0,1]$$

and

$$\overline{K}(u) = a \exp\left(-\frac{1}{1-4u^2}\right) \mathbf{1}_{|2u|\leq 1},\ a > 0.$$

For a sufficient small, $\overline{K} \in \Sigma\left(\beta, \frac{1}{2}\right) \cap \mathcal{C}^{\infty}(\mathbb{R})$. According to [30, Equation (2.5)], the functions φ_k belongs to $\Sigma\left(\beta, \frac{L}{2}, I\right)$ and the set of function

$$\mathcal{C} = \{q :\ q(x) = \rho + \sum_{k=1}^{m} w_k \varphi_k(x),\ w_k = 0, 1,\ x \in I\}$$

is included in $\Lambda_{\rho,\beta}$ as ψ_k have disjoint supports.

Let us suppose that $m \geq 8$. According to [30, Lemma 2.9], there exists a subset $\tilde{\mathcal{C}}$ of $\mathcal{C}$ such that for all $f_w = \rho + \sum_{k=1}^{m} w_k \varphi_k(x) \in \tilde{\mathcal{C}}$ and all $f_{w'} = \rho + \sum_{k=1}^{m} w'_k \varphi_k(x) \in \tilde{\mathcal{C}}$, we have

$$\sum_{k=1}^{m} (v_k - v'_k)^2 \geq \frac{m}{8}$$

and with

$$M \geq \frac{2^m}{8}$$

where $M + 1 = |\tilde{\mathcal{C}}|$. Now, if we consider two elements f_w and $f_{w'}$ of $\tilde{\mathcal{C}}$, we have

$$\begin{aligned}
\|f_w - f_{w'}\|_I &= L h^{\beta+\frac{1}{2}} \|\overline{K}\| \sqrt{\sum_{k=1}^{m} (w_k - w'_k)^2} \\
&\geq L h^{\beta+\frac{1}{2}} \|\overline{K}\| \sqrt{\frac{m}{16}} \\
&= \frac{L}{4} \|\overline{K}\| h_n^{\beta} = \frac{L}{4} \|\overline{K}\| m^{-\beta}.
\end{aligned}$$

Thus, if $n \geq n^* = \left(\frac{7}{c_0}\right)^{2\beta+1}$, $m \geq 8$ and $m^{\beta} \leq (2c_0)^{\beta} n^{\frac{\beta}{2\beta+1}}$. Hence,

$$\|f_w - f_{w'}\|_I \geq 2 s_n \tag{4.30}$$

with $s_n = An^{-\frac{\beta}{2\beta+1}}$ and $A = \frac{L}{8}\|\overline{K}\|(2c_0)^{-\beta}$.

The following part of the proof differs from [30]. $\tilde{\mathcal{C}}$ can be written $\{q_0, q_1, ...q_M\}$. Let us denote by $\mathbb{P}_j$ the probability measure associated to the intensity $nq_j(X_s)$, $s \in [0,T]$. We consider the Kullback divergence between $\mathbb{P}_0$ and $\mathbb{P}_j$ denoted by $K(\mathbb{P}_0, \mathbb{P}_j)$.

$$\begin{aligned}
K(\mathbb{P}_0, \mathbb{P}_j) &= \mathbb{E}\left(\int_0^T n\left(q_j(X_s) - q_0(X_s) - q_0(X_s)\log\left(\frac{q_j(X_s)}{q_0(X_s)}\right)\right)ds\right) \\
&= \mathbb{E}\left(\int_0^T nq_0(X_s)\left(\frac{q_j(X_s) - q_0(X_s)}{q_0(X_s)} - \log\left(1 + \frac{q_j(X_s) - q_0(X_s)}{q_0(X_s)}\right)\right)ds\right) \\
&\leq \mathbb{E}\left(\int_0^T nq_0(X_s)\left(\frac{q_j(X_s) - q_0(X_s)}{q_0(X_s)} - \log\left(1 + \frac{q_j(X_s) - q_0(X_s)}{q_0(X_s)}\right)\right)ds\right) \\
&\leq \mathbb{E}\left(\int_0^T n\frac{(q_j(X_s) - q_0(X_s))^2}{q_j(X_s)}ds\right) \qquad (4.31)
\end{aligned}$$

using the fact that for $x > -1$, $\log(1+x) \geq \frac{x}{1+x}$. Continuing from (4.31), we have:

$$\begin{aligned}
K(\mathbb{P}_0, \mathbb{P}_j) &\leq \frac{nh_n^{2\beta}L^2T\|\overline{K}\|_\infty^2}{\rho} \\
&= \frac{nL^2T\|\overline{K}\|_\infty^2 c_0^{-(2\beta+1)}m}{\rho}.
\end{aligned}$$

Let $\alpha \in \left(0, \frac{1}{8}\right)$. As $\log(M) \geq \frac{\log(2)m}{8}$, we choose

$$c_0 = \left(\frac{8L^2\|\overline{K}\|_\infty^2 T}{\alpha\log(2)\rho}\right)^{\frac{1}{2\beta+1}}$$

and we have

$$\frac{1}{M}\sum_{j=1}^M K(\mathbb{P}_0, \mathbb{P}_j) \leq \alpha\log(M). \qquad (4.32)$$

We have according to (4.30) and (4.32)

(i) $\|\lambda_j - \lambda_k\|_I > 2s_n$ for all $j \neq k$ with $s_n = An^{-\frac{\beta}{2\beta+1}}$ and

(ii) $\frac{1}{M}\sum_{j=1}^M K(\mathbb{P}_0, \mathbb{P}_j) \leq \alpha\log(M)$.

We can conclude using [30, Theorem 2.5]:

$$\liminf_{n\to\infty}\inf_{T_n}\mathbf{R}\left(T_n, \Lambda_{\rho,\beta}, n^{\frac{-\beta}{2\beta+1}}\right) > 0. \qquad (4.33)$$

Upper bound. Let us assume that $l \leq m$. Let h in $\mathcal{H}$. If $q \in \Sigma(L, \beta, I)$, the bias part of $\mathbb{E}\left(\|\hat{q}_h - q\|_I^2 | D(I,\nu)\right)$, $\mathbb{E}\left(\|q - q_h\|_I^2 | D(I,\nu)\right)$, is equal to, using Proposition 4.2,

$$\mathbb{E}\left(\int_I \left(\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s \in I} (q(X_s) - q(x)) ds\right)^2 dx | D(I,\nu)\right). \tag{4.34}$$

Using Proposition 4.2 and Taylor's expansion, there exists $(\tau_s)_{0 \leq s \leq T}$ such that the integral inside the expectation in (4.34) is equal to, on the event $D(I,\nu)$,

$$\int_I \left(\int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) \mathbf{1}_{X_s \in I} \left(q^{(l)}(x + \tau_s(X_s - x)) - q^{(l)}(x)\right) \frac{(X_s - x)^l}{l!} ds\right)^2 dx.$$

The bias term is then bounded by

$$\begin{aligned}
&\mathbb{E}\left(\int_I \left(\int_0^T |w\left(x, h, \frac{X_s - x}{h}\right) |K_h(X_s - x) \mathbf{1}_{X_s \in I} \frac{L|X_s - x|^\beta}{l!} ds\right)^2 dx | D(I,\nu)\right) \\
&\leq \mathbb{E}\left(\int_I \left(\int_0^T |w\left(x, h, \frac{X_s - x}{h}\right) |K_h(X_s - x) \mathbf{1}_{X_s \in I} \frac{Lh^\beta}{l!} ds\right)^2 dx | D(I,\nu)\right) \\
&\leq \frac{|I|^2 L^2 h^{2\beta}}{\nu^2 T^2 A_K^2 (l!)^2} \mathbb{E}\left(\|l_T\|_{I,\infty} \int_I \int_I \int_0^T K_h(X_s - x) K_h(z - x) \mathbf{1}_{X_s \in I} ds l_T^z dz dx | D(I,\nu)\right) \\
&\leq \frac{L^2 h^{2\beta} \mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right) \|K\|_1^2 |I|^2}{\nu^2 T A_K^2 (l!)^2}.
\end{aligned}$$

The variance part is equal to

$$\begin{aligned}
\mathbb{E}\left(\|\hat{q}_h - q_h\|_I^2 | D(I,\nu)\right) &= \frac{1}{n} \mathbb{E}\left(\int_I \int_0^T w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) q(X_s) ds dx\right) \\
&\leq \frac{\|q\|_{I,\infty} |I|^2}{A_K^2 \nu^2 T^2 n} \mathbb{E}\left(\int_I \int_0^T K_h^2(X_s - x) ds dx | D(I,\nu)\right) \\
&\leq \frac{\|q\|_{I,\infty} \|K\|^2 |I|^2}{n h A_K^2 \nu^2 T}.
\end{aligned}$$

Hence, as $|\mathcal{H}| \leq \tilde{C}(K, |I|) n$ where $\tilde{C}(K, |I|)$ is a constant depending only on K and $|I|$,

$$\min_{h \in \mathcal{H}} \mathbb{E}\left(\|\hat{q}_h - q\|_I^2 | D(I,\nu)\right) \leq C_1\left(q, \mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right), K, T, \beta, |I|\right) n^{\frac{-2\beta}{2\beta+1}}$$

and (4.9)

$$\mathbb{E}\left(\|\hat{q}_{\hat{h}} - q\|_I^2 | D(I,\nu)\right) \leq C_2\left(q, \mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right), K, T, \beta, |I|\right) n^{\frac{-2\beta}{2\beta+1}}$$

with $C_1\left(q, \mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right), K, T, \beta, |I|\right)$ and $C_2\left(q, \mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right), K, T, \beta, |I|\right)$ constants depending on q, $\mathbb{E}\left(\|l_T\|_{I,\infty} | D(I,\nu)\right)$, K, T, β and $|I|$. Finally,

$$\limsup_{n \to \infty} \mathbf{R}\left(\hat{q}_n, \Lambda_{\rho,\beta}, n^{\frac{-\beta}{2\beta+1}}\right) < \infty. \tag{4.35}$$

In the case where $l > m$, we apply the Taylor expansion formula of the bias up to the order m and we find that the bias is of oder n^{2m} as $q^{(m)}$ is bounded. The convergence rate is then of order $n^{\frac{-m}{2m+1}}$.

8.4 Proof of Proposition 4.8

For simplicity, the proof is done in the case where $\Theta \subset \mathbb{R}$, that is when $d = 1$. The proof is similar for any $d \geq 1$. In the following, let us use the following notation:

$$\tilde{K}(x, h, z) = w(x, h, z) K(z), \text{ for } x \in I, \ h > 0, \ z \in \mathbb{R}.$$

As X is an ancillary statistic, one can work as if X was deterministic. We also work on the event $D(I, \nu)$.

Convergence of $\hat{\theta}_n$ First, let us study the convergence of $M_n(\theta)$ for $\theta \in \Theta$. We can write $M_n(\theta)$ as the sum of

$$\frac{2}{n^2 h_n^2} \int_0^T \int_0^{s^-} \int_I \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) \mathbf{1}_{X_s \in I} \mathbf{1}_{X_u \in I} dx dM_u dM_s, \tag{4.36}$$

$$\frac{1}{h_n^2} \int_I \left(\int_0^T \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \mathbf{1}_{X_s \in I} \left(q(X_s) - g_\theta(X_s)\right) ds \right)^2 dx \tag{4.37}$$

and

$$\frac{-2}{n h_n^2} \int_I \int_0^T \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \mathbf{1}_{X_s \in I} dM_s \int_0^T \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \left(q(X_s) - g_\theta(X_s)\right) \mathbf{1}_{X_s \in I} ds dx. \tag{4.38}$$

The first term (4.36) has expectation 0 and variance equal to

$$\frac{4}{n^4 h_n^4} \int_0^T \int_0^{s^-} \left(\int_I \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) \mathbf{1}_{X_s \in I} \mathbf{1}_{X_u \in I} dx \right)^2 n^2 q(X_u) q(X_s) du ds$$

which is equal to

$$\frac{2}{n^2 h_n^4} \int_0^T \int_0^T \left(\int_I \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) \mathbf{1}_{X_s \in I} \mathbf{1}_{X_u \in I} dx \right)^2 q(X_u) q(X_s) du ds. \tag{4.39}$$

Using the occupation time formula, (4.39) is equal to

$$\frac{2}{n^2 h_n^4} \int_I \int_I \left(\int_I \tilde{K}\left(x, h_n, \frac{r - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{y - x}{h_n}\right) dx \right)^2 q(r) l_T^r q(y) l_T^y dr dy.$$

By writing $I = \left[\underline{I}, \bar{I}\right]$, (4.39) is equal to

$$\frac{2}{h_n n^2} \int_I \int_{\frac{\underline{I} - y}{h_n}}^{\frac{\bar{I} - y}{h_n}} \left(\int_{\frac{y - \bar{I}}{h_n}}^{\frac{y - \underline{I}}{h_n}} \tilde{K}(y - u h_n, h_n, u) \tilde{K}(y - u h_n, h_n, u + p) du \right)^2 q(y + h_n p) l_T^{y + h_n p} q(y) l_T^y dp dy.$$

Using the continuity properties of $x \mapsto l_T^x$ from Proposition 4.1, several times the dominated convergence theorem and the fact that $\inf_{x\in I} l_T^x \geq \nu > 0$ and $\|l_T\|_{I,\infty} < \infty$, we find that (4.39) is equivalent to

$$\frac{2}{h_n n^2}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} w(u)\, w(u+p)\, K(u)\, K(u+p)\, du\right)^2 dp \int_I \frac{(q(y))^2}{(l_T^y)^2} dy$$

with

$$w(u) = U^T(0)\left(\int_{\mathbb{R}} U(z)\, U^T(z)\, K(z)\, dz\right)^{-1} U(u).$$

Thus, (4.36) is of order $O_p\left(\frac{1}{n\sqrt{h_n}}\right)$ and goes to 0 in probability. Concerning the second term (4.37), using the dominated convergence theorem again, we find that it converges to

$$\int_{\mathbb{R}} w(u)\, K(u)\, du \int_I (q(x) - g_\theta(x))^2\, dx$$

and

$$\int_{\mathbb{R}} w(u)\, K(u)\, du = 1$$

as it is the limit of $\int_0^T w\left(x, h_n, \frac{X_s - x}{h_n}\right) K_{h_n}\left(\frac{X_s - x}{h_n}\right) ds$ which is equal to 1 according to Proposition 4.2. The last term (4.38) has mean 0 and variance equal to $O\left(\frac{1}{n}\right)$ and goes to 0 when $n \to \infty$. Thus,

$$M_n(\theta) \xrightarrow{p} M(\theta) = \|q - g_\theta\|_I^2.$$

Under Assumption 4.3, we easily show that

$$|M_n(\theta_1) - M(\theta_1) - (M_n(\theta_2) - M(\theta_2))| \lesssim |\theta_1 - \theta_2|.$$

According to Kolmogorov continuity criterion, the convergence is then uniform in θ. Furthermore, under H_0, the minimum of $M(\theta)$ is achieved for $\theta = \theta_0$. We then have, under H_0,

$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

Using Taylor's formula, on the event $\{M_n^{(2)}(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)}{2} M^{(3)}\left(\tilde{\theta}_n\right) \neq 0\}$,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{-\sqrt{n} M_n'(\theta_0)}{M_n^{(2)}(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)}{2} M^{(3)}\left(\tilde{\theta}_n\right)}$$

with $\tilde{\theta}_n$ between θ_0 and $\hat{\theta}_n$. $M_n'(\theta)$ is equal to

$$-\frac{2}{nh_n^2}\int_0^T f(h_n, \theta, X_s)\,(dN_s - n g_\theta(X_s)\, ds)$$

with

$$f(h_n, \theta, X_s) = \int_I \int_0^T \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) \mathbf{1}_{X_u \in I} \partial_\theta g_\theta(X_u)\, du \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \mathbf{1}_{X_s \in I} dx,$$

$M_n^{(2)}(\theta)$ to

$$-\frac{2}{nh_n^2}\int_0^T \partial_\theta f(h_n,\theta,X_s)(dN_s - ng_\theta(X_s)\,ds)$$
$$+\frac{2}{h_n^2}\int_I\left(\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}\partial_\theta g_\theta(X_s)\,ds\right)^2 dx$$

and $M_n^{(3)}(\theta)$ to

$$-\frac{2}{nh_n^2}\partial_{2,\theta} f(h_n,\theta,X_s)(dN_s - ng_\theta(X_s)\,ds)$$
$$+\frac{6}{h_n^2}\int_I\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}\partial_\theta g_\theta(X_s)\,ds\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}\partial_{2,\theta} g_\theta(X_s)\,dsdx.$$

Under H_0, $M_s = N_s - n\int_0^s g_{\theta_0}(X_s)\,ds$. Thus, $\sqrt{n}M_n'(\theta_0)$ has expectation 0 and variance

$$\frac{4}{h_n^4}\int_0^T\left(\int_I\int_0^T w\left(x,h_n,\frac{X_s-x}{h_n}\right)\partial_\theta g_{\theta_0}(X_s)\,\mathbf{1}_{X_s\in I}ds\tilde{K}\left(x,h_n,\frac{X_u-x}{h_n}\right)\mathbf{1}_{X_u\in I}\right)^2 g_{\theta_0}(X_u)\,du$$

that converges to

$$4\int_I \frac{(\partial_\theta g_{\theta_0}(u))^2 g_{\theta_0}(u)}{l_T^u}du.$$

We can easily show that

$$\frac{8(\sqrt{n})^3}{n^3h_n^6}\int_0^T (f(h_n,\theta,X_s))^3 ng_{\theta_0}(X_u)\,du = O\left(\frac{1}{\sqrt{n}}\right)$$

and converges to 0. Thus, according to [25, Theorem 3], we have the following central limit theorem

$$-\sqrt{n}M_n'(\theta_0)\xrightarrow{\mathcal{L}}\sqrt{4\int_I\frac{(\partial_\theta g_{\theta_0}(u))^2 g_{\theta_0}(u)}{l_T^u}du}\mathcal{N}(0,1). \tag{4.40}$$

Under H_0, the first term of $M_n^{(2)}(\theta_0)$

$$-\frac{2}{nh_n^2}\int_0^T \partial_\theta f(h_n,\theta,X_s)\,dM_s \tag{4.41}$$

has mean 0 and variance

$$\frac{4}{n^2h_n^4}\int_0^T (\partial_\theta f(h_n,\theta,X_s))^2 ng_{\theta_0}(X_s)\,ds = O\left(\frac{1}{n}\right).$$

Then, (4.41) converges to 0 in probability. The second term of $M^{(2)}(\theta_0)$

$$\frac{2}{h_n^2}\int_I\left(\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}\partial_\theta g_\theta(X_s)\,ds\right)^2 dx$$

converges in probability to

$$2\int_I \left(\partial_\theta g_{\theta_0}(x)\right)^2 dx.$$

Then,

$$M_n^{(2)}(\theta_0) \xrightarrow{p} 2\int_I \left(\partial_\theta g_{\theta_0}(u)\right)^2 du. \tag{4.42}$$

The absolute value of $M^{(3)}\left(\tilde{\theta}_n\right)$ has mean equal to $O(1)$. As $\hat{\theta}_n - \theta \xrightarrow{p} 0$,

$$\left(\hat{\theta}_n - \theta\right) M_n^{(3)}\left(\tilde{\theta}_n\right) \xrightarrow{p} 0. \tag{4.43}$$

Using (4.40), (4.42) and (4.43),

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \frac{\sqrt{\int_I \frac{\left(\partial_\theta g_{\theta_0}(u)\right)^2 g_{\theta_0}(u)}{l_T^u} du}}{\int_I \left(\partial_\theta g_{\theta_0}(u)\right)^2 du} \mathcal{N}(0,1),$$

achieving the proof of (i).

Convergence of $M_n\left(\hat{\theta}_n\right)$ under H_0. $M_n\left(\hat{\theta}_n\right)$ is the sum of (4.36), (4.37) and (4.38), replacing θ by $\hat{\theta}_n$. Concerning (4.36), we have

$$\frac{16}{n^4 h_n^6}\int_0^T \int_0^{s^-} \left(\int_I \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) \mathbf{1}_{X_s \in I} \mathbf{1}_{X_u \in I} dx\right)^4 n^2 q(X_u)\, q(X_s)\, du ds$$

which is equivalent to

$$\frac{8}{h_n n^2}\int_{\mathbb{R}} \left(\int_{\mathbb{R}} w(u)\, w(u+p)\, K(u)\, K(u+p)\, du\right)^4 dp \int_I \frac{\left(q(y)\right)^2}{\left(l_T^y\right)^6} dy$$

and converges to 0. According to [25, Theorem 3],

$$\frac{2\sqrt{h_n} n}{n^2 h_n^2}\int_0^T \int_0^{s^-} \int_I \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \tilde{K}\left(x, h_n, \frac{X_u - x}{h_n}\right) dx \mathbf{1}_{X_s \in I} \mathbf{1}_{X_u \in I} dM_u dM_s$$

converges in law to

$$\sqrt{2\int_{\mathbb{R}} \left(\int_{\mathbb{R}} w(u)\, w(u+p)\, K(u)\, K(u+p)\, du\right)^2 dp \int_I \frac{\left(q(y)\right)^2}{\left(l_T^y\right)^2} dy}\, \mathcal{N}(0,1).$$

The term (4.37) is of the same order than $\left(\hat{\theta}_n - \theta_0\right)^2$ which is $O_p\left(\frac{1}{n}\right)$ and

$$\frac{\sqrt{h_n} n}{h_n^2}\int_I \left(\int_0^T \tilde{K}\left(x, h_n, \frac{X_s - x}{h_n}\right) \mathbf{1}_{X_s \in I}\left(q(X_s) - g_{\hat{\theta}_n}(X_s)\right) ds\right)^2 dx = O_p\left(\sqrt{h_n}\right) \xrightarrow{p} 0.$$

The last term corresponding to (4.38) is bounded by

$$M_I|\theta_0-\tilde{\theta}_n|\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}dM_s\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}ds$$

equal to $O_p\left(\frac{1}{n}\right)$. Thus,

$$-\frac{2}{h_n^{\frac{3}{2}}}\int_I\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\mathbf{1}_{X_s\in I}dM_s\int_0^T \tilde{K}\left(x,h_n,\frac{X_s-x}{h_n}\right)\left(q\left(X_s\right)-g_{\hat{\theta}_n}\left(X_s\right)\right)\mathbf{1}_{X_s\in I}ds$$

is equal to $O_p\left(\sqrt{h_n}\right)$ and converges to 0 in probability. Finally,

$$n\sqrt{h_n}M_n\left(\hat{\theta}_n\right)\xrightarrow{\mathcal{L}}\sqrt{2\int_{\mathbb{R}}\left(\int_{\mathbb{R}}w\left(u\right)w\left(u+p\right)K\left(u\right)K\left(u+p\right)du\right)^2dp\int_I\frac{\left(q\left(y\right)\right)^2}{\left(l_T^y\right)^2}dy\mathcal{N}\left(0,1\right)}.$$

Convergence of $M_n\left(\hat{\theta}_n\right)$ under H_1. Under H_1, the term (4.37) is bounded and then

$$|n\sqrt{h_n}M_n\left(\hat{\theta}_n\right)|\to\infty.$$

Bibliography

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[2] Yacine Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial studies*, 9(2):385–426, 1996.

[3] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.

[4] Martin T Barlow and Marc Yor. Semi-martingale inequalities via the garsia-rodemich-rumsey lemma, and applications to local times. *Journal of functional Analysis*, 49(2):198–229, 1982.

[5] Fred Espen Benth, Rüdiger Kiesel, and Anna Nazarova. A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5):1589–1616, 2012.

[6] Fred Espen Benth, Nina Lange, and Tor Age Myklebust. Pricing and hedging quanto options in energy markets. *Journal of Energy Markets*, 2015.

[7] Fred Espen Benth and Jūratė Šaltytė Benth. Weather derivatives and stochastic modelling of temperature. *International Journal of Stochastic Analysis*, 2011, 2011.

[8] Maria Mori Brooks and J Stephen Marron. Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1):157–165, 1991.

[9] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[10] Feng Chen, Paul SF Yip, and KF Lam. On the local polynomial estimators of the counting process intensity function and its derivatives. *Scandinavian Journal of Statistics*, 38(4):631–649, 2011.

[11] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l'institut Henri Poincaré (B)*, volume 47, pages 1171–1196, 2011.

[12] Rama Cont and Peter Tankov. *Financial modelling with jump processes*, volume 2. CRC press, 2003.

[13] David R Cox. Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer, 1992.

[14] Sylvain Delattre, Christian Y Robert, and Mathieu Rosenbaum. Estimating the efficient price from the order flow: a brownian cox process approach. *Stochastic Processes and their Applications*, 123(7):2603–2619, 2013.

[15] Peter Diggle. A kernel method for smoothing point process data. *Applied statistics*, pages 138–147, 1985.

[16] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.

[17] Marc Hoffmann. On estimating the diffusion coefficient: parametric versus nonparametric. In *Annales de l'IHP Probabilités et statistiques*, volume 37, pages 339–372, 2001.

[18] Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer, 2003.

[19] Alan Karr. *Point processes and their statistical inference*, volume 7. CRC press, 1991.

[20] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, pages 1–38, 2016.

[21] Matthieu Lerasle, Nelo Molter Magalhães, and Patricia Reynaud-Bouret. Optimal kernel selection for density estimation. In *High Dimensional Probability VII*, pages 425–460. Springer, 2016.

[22] Thilo Meyer-Brandis and Peter Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(05):503–528, 2008.

[23] Susan Allbritton Murphy and Pranab Kumar Sen. Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1):153–180, 1991.

[24] Yosihiko Ogata and Koichi Katsura. Point-process models with linearly parametrized intensity for application to earthquake data. *Journal of applied probability*, 23(A):291–310, 1986.

[25] Giovanni Peccati and Murad S Taqqu. Central limit theorems for double poisson integrals. *Bernoulli*, 14(3):791–821, 2008.

[26] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[27] Patricia Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126(1):103–153, 2003.

[28] Patricia Reynaud-Bouret. Concentration inequalities, counting processes and adaptive statistics. In *ESAIM: Proceedings*, volume 44, pages 79–98. EDP Sciences, 2014.

[29] Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089, 2005.

[30] Alexandre B Tsybakov. Introduction to nonparametric estimation, 2009.

[31] Klaus J Utikal. Nonparametric inference for a doubly stochastic poisson process. *Stochastic processes and their applications*, 45(2):331–349, 1993.

[32] Ju-Yi Yen and Marc Yor. Local times and excursion theory for brownian motion a tale of wiener and ito measures preface, 2013.

[33] Tingting Zhang and SC Kou. Nonparametric inference of doubly stochastic poisson process data via the kernel method. *The annals of applied statistics*, 4(4):1913, 2010.

Chapter 5

A joint model for electricity spot prices and wind penetration with dependence in the extremes

Abstract

This article analyses the dependence between electricity spot prices and the wind penetration index in the European energy market. The wind penetration index is given by the ratio of the wind energy production divided by the total electricity production. We find that the wind penetration has an impact on the intensity of the spike occurrences in the electricity prices, and we formulate a joint model for electricity prices and wind penetration and calibrate it to recent data. We then use the new joint model in an application where we assess the impact of the modelling assumptions on the potential income of an electricity distributor who buys electricity from a wind farm operator.

1 Introduction

The trend to increase renewable energy production all over the world with the possibility of phasing out conventional energy sources while at the same time ensuring reliability of energy networks constitutes one of the key challenges of modern societies. Due to their high variability renewable sources of energy tend to be more difficult to predict and for instance their precise impact on electricity prices is far from being understood. The recent literature presents various attempts to model and characterise the impact of wind energy production on electricity prices, see for instance [9], [10], [5] and [11] for some recent accounts. In these studies, wind is often considered as an exogenous variable for the electricity price. However, for many applications, one of which will be presented in this article, we in fact need a joint model for electricity prices and wind energy generation (and possibly other sources of electricity as well). Hence the goal of this article is to formulate and estimate a joint model for electricity spot prices and wind energy production, more precisely wind penetration, which is the ratio between wind energy production and total energy production.

Our modelling idea which is guided by our statistical analysis rests on the hypothesis that increasing wind penetration typically results in lower electricity spot prices and that high wind penetration might increase the frequency at which negative price spikes occur. In order to formalise these ideas we draw on new statistical methodology for estimating jump intensities and mean reversion rates in mean-reverting jump processes, which has been developped in Chapter 3 and Chapter 4.

The outline of the remainder of this article is as follows. Section 2 gives a detailed description of the data and data sources used in our statistical analysis. The key contributions are then collected in Section 3, where we introduce and estimate a joint model for hourly electricity spot prices from Germany and Austria and the corresponding wind penetration index. Finally, in Section 4, we take the point of view of an electricity distributor and we ask the question of how his estimated income of a contract which allows him to buy a certain amount of electricity from a wind farm operator depends on the dependence modelling between the electricity spot price and the wind penetration index.

2 Data description and exploratory study

2.1 Data description

In this article we analyse German and Austrian wind energy production data, electricity price data and load data covering a period from the 1st of January 2012 to the 31st of December 2016. The data have been downloaded from the following website
https://data.open-power-system-data.org/. The precise data sets considered are the following ones:

- The German and Austrian hourly electricity spot prices (from the day-ahead auction) from the 1st of January 2012 to the 31st December 2016,
- the German and Austrian hourly load data from the 1st of January 2012 to the 31st December 2016,
- the German and Austrian hourly wind energy production data from the 1st of January 2012 to the 31st December 2016. Note that the data has been aggregated over the four German transmission system operators 50 Hertz Transmission, Amprion, Tennet TSO and EnBW Transportnetze and the Austrian transmission system operator APG.

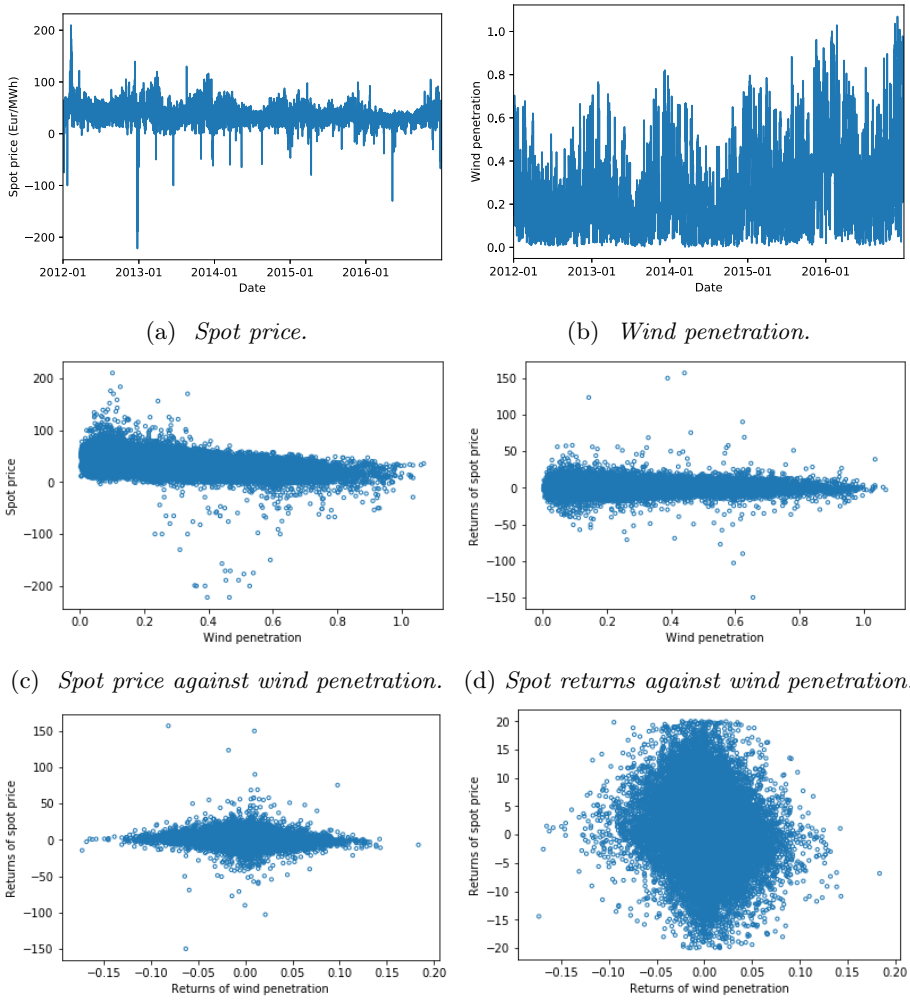
Throughout this article, we will be analysing and modelling the hourly data from the day-ahead market rather than the daily data, which is often done in the literature. The hourly data are revealed in the daily auction at the same time, but it has been found in e.g. [7] that when taking seasonality into account, they can be successfully modelled by classical time series models at an hourly frequency. An alternative approach, which we leave for future research, would be to model the hourly data as a 24-dimensional vector of daily data, as it has for instance been done by e.g. [8] in the discrete-time setting and by [19] in the continuous-time setting.
Motivated by the work by [9] we use our data to compute the German/Austrian wind penetration index, defined as the ratio of the wind energy production and the electricity load. In [9] the authors found a strong association between the wind penetration index and the corresponding electricity prices. Figure 5.1a and Figure 5.1b depict the time series of the German/Austrian spot price and the wind penetration, respectively. We observe that the wind penetration index takes values between zero and one, except for eleven values in 2016 when it exceeds one. This can happen because of the interconnection between the transmission networks of the various European countries. Since the total electricity production can also be used in other countries in Europe, a wind penetration index exceeding one tells us that in these particular hours all of the German/Austrian electricity came from wind energy.

2.2 Exploratory data analysis

Next, we aim to study the relation between the prices and the wind penetration index. To this end, let us look at Figure 5.1c. Here we observe that high negative values of spot prices happen when the wind penetration crosses a certain threshold. More precisely, Figure 5.1d reveals that high negative price returns appear when the wind penetration is over 0.4. Next, Figures 5.1e and 5.1f depict the returns of the spot price against the returns of the wind penetration index, where extreme values have been removed in the latter picture. (The procedure of how the extremes were removed is described below.) The corresponding correlation coefficients for the data in Figures 5.1e and 5.1f are given by -0.125 and -0.148, respectively. We want to investigate now whether or not there is any association between wind penetration and extreme spot returns. We recall from Figure 5.1a that both positive and negative spikes appear in the data. Hence when we are talking about extreme values in the following we mean the absolute value of the corresponding spot returns. Note that as a positive spike (resp. negative) leads to a negative (resp. positive) extreme return due to the high mean reversion, we do not consider to study negative and positive extreme values separately: in this case, the negative (resp. positive) extreme value is caused by the positive (resp. negative) one and not by the wind.

In order to identify the extreme values in the spot returns, we establish the QQ-Plot of the spot returns against a normal law: we obtain Figure 5.2a. Removing the values having an absolute value greater than 20, we obtain the QQ-Plot in Figure 5.2b. This new plot shows a better correspondence with the normal distribution and we consider the returns of the spot having an absolute value greater than 20, corresponding to the quantile of order 99.1%, as extreme values. Figure 5.3 represents the extreme values of the spot return against the wind penetration.
We remark that our procedure of filtering out extreme values can be regarded as a hard thresholding approach, whereas one could also use more sophisticated techniques based on methods from extreme value theory as in [12; 19].

In order to study the impact of wind penetration on spot returns extreme values, one indicator, omitting the time dependency aspect, is the quantity $\mathbb{P}\left(|R| > 20 | WP > x\right)$ for a given x where R



(a) *Spot price.*

(b) *Wind penetration.*

(c) *Spot price against wind penetration.*

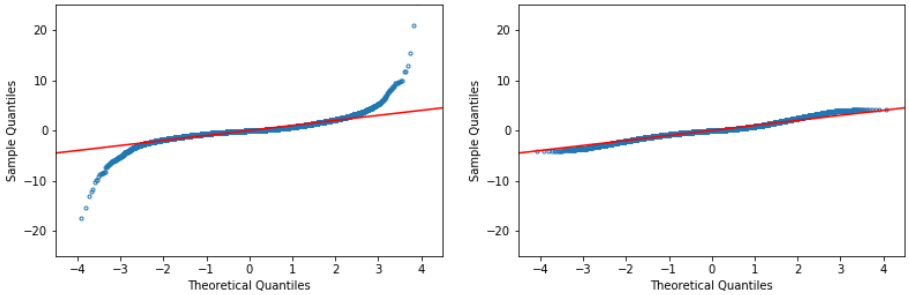
(d) *Spot returns against wind penetration.*

(e) *Spot returns against wind penetration returns.*

(f) *Non extreme spot returns against wind penetration returns.*

Figure 5.1: *Hourly German spot price and wind penetration index.*

corresponds to the return of the spot and WP is the wind penetration. Figure 5.4 corresponds to the function $x \to \mathbb{P}\left(|R| > 20 | WP > x\right)$. For $x = 0$, the function takes value 0.82% where it reaches level of order 2% when x increases: the probability to have an extreme value in the time series of the spot price is higher with higher values of wind penetration; this dependence has to be taken into account if we want to model the joint distribution of the spot price and the wind penetration.



(a) *QQ-Plot with all the values.* (b) *QQ-Plot without extremal returns.*

Figure 5.2: *QQ-Plot of spot returns against a normal law, with and without certain extreme values.*

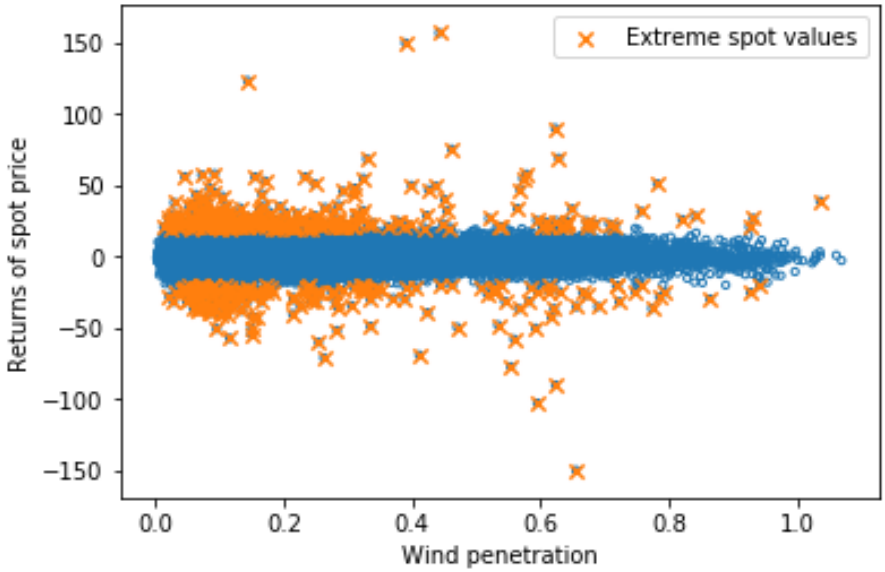


Figure 5.3: *Returns of the spot price against the the wind penetration with extreme spot returns.*

3 A joint model for the electricity spot price and the wind penetration index

The main contribution of this article is that we present and estimate a joint stochastic model for the hourly electricity spot price (denoted by S) and the corresponding wind penetration index (denoted by WP). We will now proceed by explaining step-by-step the stochastic model and the corresponding estimation procedure.

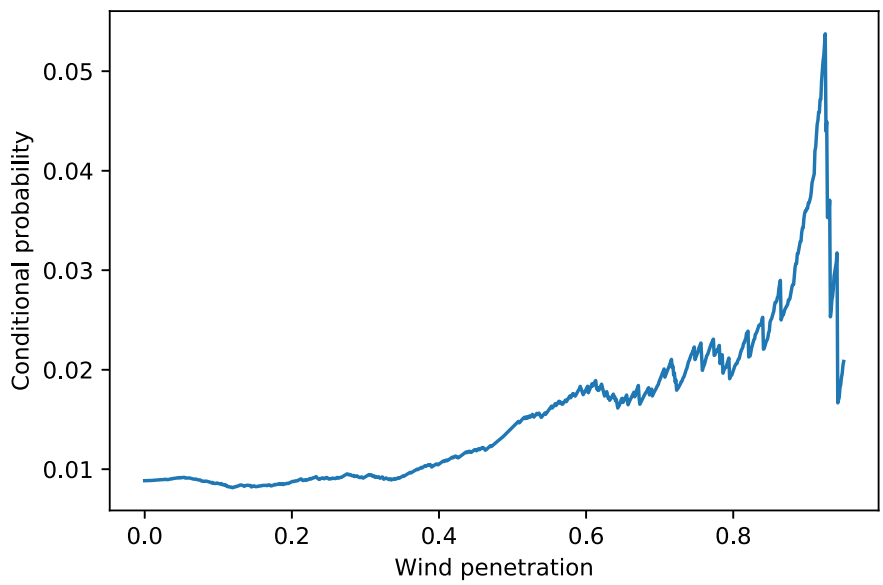


Figure 5.4: *Probability to have an extreme returns conditionally on wind production greater than a certain threshold.*

3.1 Model for the electricity spot price

Let $(S_t)_{0\leq t\leq T}$ denote the electricity spot price. Motivated by our exploratory study, we decompose S in the following way:

$$S_t = \Gamma_{1,t} + X_{1,t} + Y_t$$

where Γ_1 is a seasonality function, X represents a continuous stochastic part of the spot price and Y represents a spike part. The process $\Gamma_1 + X_1$ is assumed to be a continuous semi-martingale.

Modelling the spike part Y

While Section 2.2 gives statistical evidence of a dependence in the extreme values of electricity spot returns and wind penetration, in this section, we add the temporal aspect to our analysis. We note that spikes are characteristic of electricity prices and correspond to extreme values in the time series. We are interested in studying the relation between the intensity of the spike occurrence and the wind penetration index. To this end, we start by modelling the spikes by a doubly stochastic and mean-reverting Poisson-type process. We first need to identify the times of the spikes, and next the intensity of the doubly stochastic Poisson process is studied as a function of the wind penetration. A spike can be modelled as a jump going back to 0 with a strong mean reversion and with a typically stochastic size. Thus, we can model the spike part of the spot Y by

$$Y_t = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} \underline{p}\,(dt, dx)\,, \tag{5.1}$$

where $\underline{p}\,(dt, dx)$ is a Poisson measure with compensator $\underline{q} = \lambda_t dt \otimes \nu\,(dx)$, $\beta > 0$ is the speed of the mean reversion and ν is a finite measure satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} x^2 \nu\,(dx) < \infty$. Y is the

unique solution of the stochastic differential equation

$$dY_t = -\beta Y_t dt + x\underline{p}\,(dt, dx)\,,$$

with initial condition $Y_0 = 0$.

Estimation of the times of the spikes. Chapter 3 establishes a method for estimating the arrival times of the jumps in the above setting. We remark that in Chapter 3, the Poisson measure is not stochastic. However, adding the assumption that λ is bounded below and above, the result can easily be extended to the stochastic case using the same arguments as in the original proof.
Let us briefly recall the key ideas: Suppose the price S is observed over the time interval $[0, T]$ for some $T > 0$ at discrete times $t_i = i\Delta_n$, for $i = 0, 1, \ldots, \lfloor T/\Delta_n \rfloor$. Here $\Delta_n > 0$ and $\Delta_n \to 0$ as $n \to \infty$. Note that the asymptotic framework, contrarily to classical high-frequency statistics, is not limited to $\Delta_n \to 0$; in Chapter 3 an asymptotic is added on β, $\beta = \beta_n \to \infty$ when $n \to \infty$, and on λ, $\lambda_s = \lambda_{s,n} = l_n \tilde{\lambda}_s$ and $l_n \to \infty$ when $n \to \infty$; the case when l_n is bounded is also considered, but we need it to go to ∞ for the estimation procedure and statistical test of the intensity function described in the following. The following assumptions are needed in order to have a consistent framework with the data:

(i) $\beta_n \Delta_n \lesssim 1$, otherwise a jump can happen and go back to 0 in a period Δ_n and is not observed,

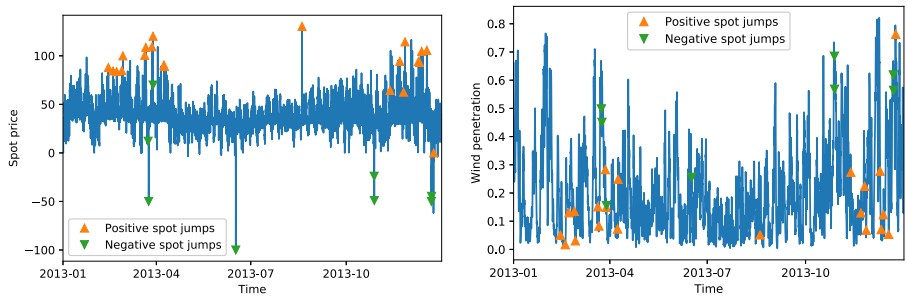
(ii) $\frac{l_n}{\beta_n} \lesssim 1$, insuring the stability of Y_t by taking the frequency of appearance of jumps not large compared to the speed of mean reversion,

(iii) $l_n^2 \Delta_n \to 0$, in order to have at most one jump between two observations.

The method for identifying the jumps is based on the use of a threshold of the form $C\Delta_n^{\varpi}$ for the returns, where $C > 0$ and $\varpi \in (0, 0.5)$ are constants which we specify below. This method is classic for the estimation of jumps in jump diffusion models, see [2, Section 10.4] for instance, and has been adapted to the spike case in Chapter 3. Indeed, as the mean reversion is large and can be of the size of a jump, we need to add the condition that $\Delta_i^n S \Delta_{i+1}^n S < 0$, where $\Delta_i^n S = S_{t_i} - S_{t_{i-1}}$. It indicates that after a jump (in the interval $((i-1)\Delta_n, i\Delta_n]$), the subsequent increment will be of the opposite sign of the jump. We choose a threshold equals to $5\bar{\sigma}\Delta_n^{0.49}$ with $\bar{\sigma}^2$ the multipower variation estimator of the volatility with order 20; the reader can refer to [3] and [17] for more details about multipower variation. The particular choice of our threshold is motivated by the following findings in the literature: First, both [1], see Section 5.3, and [2], see Section 6.2.2, p. 187, recommend using a constant of the form $C = \tilde{C}\bar{\sigma}$, where $\tilde{C}$ is a constant and $\bar{\sigma}$ is an 'average' of the volatility. Moreover they advise choosing $\varpi \in (0, 0.5)$ close to 0.5. In addition, [1] also suggests choosing $\tilde{C}$ between 3 and 5, see also [14] ($\tilde{C} = 3$) and [15] ($\tilde{C} = 4$).
In order to avoid too much change in volatility, the data are segmented in time series of one year in order to identify the jumps. That means that $\bar{\sigma}$ does not stay constant over the entire sample, but just over each of the five years. For the considered data, we find an estimated number of 114 jumps, 30 of which are negative and 84 positive. Figure 5.5a corresponds to the spot price time series with the times of the jumps marked by upward triangles for positive jumps and by downward triangles for negative jumps.

Remark 5.1. *The multipower variation estimator of the volatility is chosen with order 20 which is high compared to the orders typically chosen in the literature. Contrarily to classical jumps, spikes have a strong mean reversion. Using for instance bipower variation in such a scenario is not*

suitable because the impact of the jump is not diminished by the increments that follows. One also needs to compensate the effect of the mean reversion, that can be present in two or three increments following the jumps. Moreover, some spikes have a large amplitude, and then have a strong impact on the value of the volatility estimator if the order is low. Having 8760 observations per year allows us to choose an order 20 with a good estimation quality. Some simulations have been performed on simple models such as an arithmetic or geometric Brownian motions plus spikes and order 20 appears to be a good choice. The choice of 20 corresponds to a thresholds around 20 Euros per MWh per year, corresponding to the one used to select the extreme values in the quantile-quantile plot. Note that we found that the choice of the order of the multipower does not influence the results significantly in any case. Decreasing the order of the estimator leads to higher values for volatility estimator and then also for the threshold. A threshold of 10 leads to a detection of 61 positive spikes and 27 negative spikes. The number of negative spikes, which is our key object of interest, does not change much. Also, the results concerning the dependence with the wind penetration are not affected.



(a) *Spot.* (b) *Wind penetration.*

Figure 5.5: *Jumps in German spot price and wind penetration time series.*

Dependence of the spike intensity on the wind penetration index. Figure 5.5b depicts the positive and negative jump times of the spot price superimposed on the time series plot of the wind penetration index. We observe that negative jumps appear more often when the wind penetration is high, whereas positive jumps also appear for small values of wind penetration. In the following, we will separate the positive and the negative jumps in order to study the impact of the wind penetration on each type of jumps independently.

To this end, let us consider two point processes corresponding to the positive and the negative jumps, respectively. We want to study the dependence between the point processes and the wind penetration. In order to simplify the exposition, we do not use superscripts for the two different doubly stochastic Poisson processes we are considering, but just the generic N for the doubly stochastic Poisson process governing either the positive or the negative spikes. The doubly stochastic Poisson process is entirely defined by its intensity process denoted by $\lambda_n = \left(l_n \tilde{\lambda}_t\right)_{0 \leq t \leq T}$. Since we are interested in studying the dependence between the intensity λ_n and the wind penetration, we assume a functional relationship of the form:

$$\tilde{\lambda}_t = q\left(WP_t\right), \; t \in [0, T],$$

where $q : \mathbb{R} \mapsto (0, \infty)$ is a deterministic function. We remark that conditional on the wind penetration index, N is an inhomogeneous Poisson process. Our goal is to estimate the function q on a given time interval I where wind penetration data is available.

In the following, let us assume that every jump has been identified meaning that we have an estimated sample path of N. In order to estimate the function q, we use a discretised version of the classical Nadaraya-Watson estimator defined, for $h > 0$, by

$$\hat{q}_h(x) = \frac{1}{l_n} \frac{\sum_{i=1}^n K_h\left(WP_{t_{i-1}} - x\right) 1_{WP_{t_{i-1}} \in I} \Delta_i^n N}{\Delta_n \sum_{j=1}^n K_h\left(WP_{t_{j-1}} - x\right) 1_{WP_{t_{j-1}} \in I}},$$

where $\Delta_i^n N$ denote the increments of the estimated path of N and where

$$K_h(x) = h^{-1} K\left(\frac{x}{h}\right),$$

and K a kernel function. In order to estimate the optimal bandwidth, that is the one minimising

$$\mathbb{E}\left(\|\hat{q}_h - q\|_{2,I}^2\right) = \mathbb{E}\left(\int_I \left(\hat{q}_h(x) - q(x)\right)^2 dx\right),$$

Chapter 4 proposes to choose the bandwidth as

$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \|\hat{q}_h - \hat{q}_{h_{min}}\|_{2,I}^2 + pen_\alpha(h),$$

where

$$pen_\alpha(h) = \alpha \hat{V}_h - \hat{V}_h - \hat{V}_{h_{\min}} + 2\hat{V}_{h,h_{\min}}, \ \alpha > 0,$$

$$\hat{V}_h = \frac{1}{l_n^2} \int_I \frac{\sum_{i=1}^n K_h^2\left(WP_{t_{i-1}} - x\right) \mathbf{1}_{WP_{t_{i-1}} \in I} \Delta_i^n N}{\Delta_n^2 \left(\sum_{j=1}^n K_h\left(WP_{t_{j-1}} - x\right) \mathbf{1}_{WP_{t_{j-1}} \in I}\right)^2} dx,$$

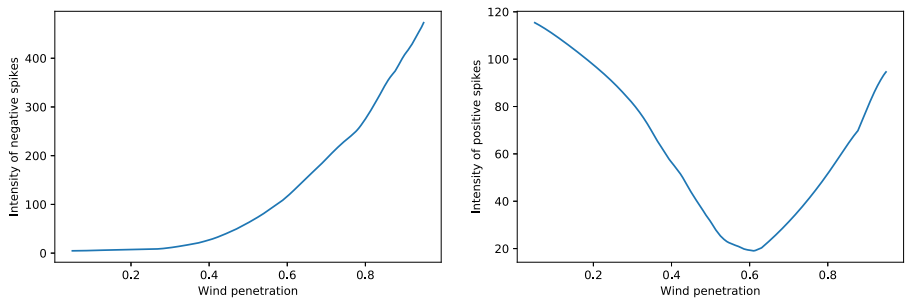
$$\hat{V}_{h,h_{\min}} = \frac{1}{l_n^2} \int_I \frac{\sum_{i=1}^n K_h\left(WP_{t_{i-1}} - x\right) K_{h_{\min}}\left(WP_{t_{i-1}} - x\right) \mathbf{1}_{WP_{t_{i-1}} \in I} \Delta_i^n N}{\Delta_n^2 \sum_{j=1}^n K_h\left(WP_{t_{j-1}} - x\right) \mathbf{1}_{WP_{t_{j-1}} \in I} \sum_{k=1}^n K_{h_{\min}}\left(WP_{t_{k-1}} - x\right) \mathbf{1}_{WP_{t_{k-1}} \in I}} dx,$$

and $\mathcal{H}$ is a finite subset of $(0, \infty)$, $h_{min} = \min \mathcal{H} \geq \frac{\|K\|_\infty \|K\|_1}{l_n}$. Choosing this bandwidth leads to an oracle inequality, justifying this particular choice. In Chapter 4 it is shown that the bandwidth selection procedure can be applied more generally to local polynomial estimators of the intensity function; here we consider the particular case of kernel estimator corresponding to a polynomial of order 0. This method is similar to the one used by [13] for the kernel estimation of a density, which is an extension of the Goldenshluger-Lepski method [6].

We remark that Chapter 4 is casted in a continuous-time framework, whereas our estimation is done in discrete time. However, since $\Delta_n \to 0$ and T is large, the resulting approximation error which is of order Δ_n is asymptotically negligible and the processes can be assumed to be continuously observed. This approximation is justified by the assumption $l_n^2 \Delta_n \to 0$.

In the following, let us consider the Epanechnikov kernel $K(u) = \frac{3}{4}\left(1 - u^2\right) \mathbf{1}_{|u| \leq 1}$. Also, the parameter α is set to 1, leading to an optimal oracle inequality. Choosing $I = [0.05, 0.95]$, $\mathcal{H} = \{h = h_{\min} + 0.01 \times i, \ i \in \mathbb{N}, \ h \leq 0.4\}$, $h_{\min} = 0.0225$ for negative jumps and $h_{\min} = 0.0089$ for positive ones, the optimal bandwidths given by this method are $\hat{h} = 0.25$ for negative jumps and $\hat{h} = 0.30$ for positive ones. The optimal estimator $\hat{q}_{\hat{h}}$ for negative jumps is given in Figure

5.6a and the one for positive jumps in Figure 5.6b. Chapter 4 proposes a method to test if the intensity function belongs to some parametric function. Using this method, we test if the intensity function q is constant with respect to the wind penetration. The test is rejected for negative jumps at level 95% (with p-value equal to 0) but is not for positive jumps (with p-value equal to 0.47). These results confirm our intuition: the wind penetration index has an impact on the probability to have a negative spike but not on the probability to have a negative one. High wind penetration implies that other means of production are not used to satisfy the demand, and renewable energies constitute the cheapest means of energy production. Because of the non-storability of electricity, a producer owning a plant not used to satisfy the demand needs to pause it. Sometimes, it is cheaper for him to pay someone to consume the produced electricity rather than stop and restart his plant. This leads to negative jumps and negative prices for the electricity spot price.



(a) *For negative jumps.* (b) *For positive jumps.*

Figure 5.6: *Kernel estimators of the intensity of the spot spikes as a function the wind penetration.*

Estimating the speed of mean reversion of Y. Note that Chapter 3 presents a method for estimating the speed of mean reversion of the spikes given by $\beta = \beta_n$, assuming that β_n is large enough. The parameter β_n is estimated using the slope of the process after a jump, which is of order $K\left(1 - e^{-\beta_n \Delta_n}\right)$ with K being the size of the jump. We need to introduce some notation before we are able to define the estimator of β_n: Let us denote by

$$\mathcal{I}_n(1) < \mathcal{I}_n(2) < ... < \mathcal{I}_n\left(\hat{\lambda_n}\right)$$

the index in $\{1, ..., n-1\}$ satisfying $|\Delta_i^n S| > 5\hat{\sigma}\Delta_n^{0.49}$ and $\Delta_i^n S \Delta_{i+1}^n S < 0$, corresponding to the estimator of the time index of the jumps, and by $\hat{\lambda}_n$ an estimator of N_T. The estimator of β_n is given by

$$\tilde{\beta}_n = \frac{-1}{\Delta_n} \log\left(\left(1 + \frac{\sum_{q=1}^{\widehat{\lambda}_n} sgn(\Delta_{\mathcal{I}_n(q)}^n S)\left(\Delta_{\mathcal{I}_n(q)+1}^n S + 2\Delta_n \sum_{j=1}^{q-1} \Delta_{\mathcal{I}_n(q)}^n S\right)}{\sum_{q=1}^{\widehat{\lambda}_n} |\Delta_{\mathcal{I}_n(q)}^n S|} \mathbf{1}_{\widehat{\lambda}_n > 0}\right) \vee \Delta_n\right).$$

The term $2\Delta_n \sum_{j=1}^{q-1} \Delta_{\mathcal{I}_n(q)}^n S$ is a bias correction. Under the assumption that $\beta_n \sqrt{\Delta_n l_n} \to \infty$ and some other technical assumptions given in Chapter 3, our estimator is consistent:

$$\frac{\tilde{\beta}_n - \beta_n}{\beta_n} \xrightarrow{p} 0.$$

As β_n is large, we consider the relative error $\frac{\tilde{\beta}_n - \beta_n}{\beta_n}$ instead of $\tilde{\beta}_n - \beta_n$. We find $\tilde{\beta}_n = \frac{0.88}{\Delta_n}$. As for the jump detection, we note that while the results of Chapter 3 consider a constant λ, the proofs of the consistency of the estimator (and possibly also of the central limit theorem) can be generalised to the case of a stochastic intensity, provided the stochastic intensity is bounded below and above. An estimator of the size of the jth jump ΔS_{T_j} is given by $\Delta^n_{\mathcal{I}_n(j)} S$. However, as β_n is large, this estimator is biased: indeed, as noticed in Chapter 3, $\Delta^n_{\mathcal{I}_n(j)} S$ is equivalent to $e^{-\beta_n \left(\lfloor \frac{T_j}{\Delta_n} \rfloor \Delta_n + \Delta_n - T_j \right)} \Delta S_{T_j}$ when Δ_n is small. In the case when $\beta_n \Delta_n$ is not small, the term before the jump size is then not closed to one: as T_j can not be estimated because of the discrete observations, we can not estimate this bias term to deduce the exact jump size ΔS_{T_j}. Nevertheless, as the number of jumps is low, we do not consider a model for the jump sizes and keep the empirical one, knowing that there is a bias.

Modelling the seasonality function Γ_1

The seasonality Γ_1 is assumed to be of the form

$$\Gamma_{1,t} = c_{0,1} + c_{1,1} t + c_{2,1} t^2 + c_{3,1} \cos\left(\frac{\tau_{0,1} + 2\pi t}{365 \times 24}\right) + c_{4,1} \cos\left(\frac{\tau_{1,1} + 2\pi t}{7 \times 24}\right) + c_{5,1} \cos\left(\frac{\tau_{2,1} + 2\pi t}{24}\right). \tag{5.2}$$

In order to estimate the seasonality function, we start by removing spot values corresponding to spikes and jumps. Values such that $|\Delta^n_i S| \geq 5\bar{\sigma}\Delta_n^{0.49}$, which are not always spikes, are removed; they correspond to extreme value returns. We also remove ten values following an extreme value, that insures for a spike to have reverted to low value. After this filtering, a least square minimisation is done. Parameters and standard errors are provided in Table 5.1. The function Γ_1 and the deseasonalised and filtered spot are depicted in Figures 5.7a and 5.7b, respectively.

	$c_{0,1}$	$c_{1,1}$	$c_{2,1}$	$c_{3,1}$	$c_{4,1}$	$c_{5,1}$
Estimate	45.94	-8.2×10^{-4}	1.0×10^{-8}	-3.85	-6.00	-6.10
Standard error	0.16	1.7×10^{-5}	3.8×10^{-10}	7.6×10^{-2}	7.5×10^{-2}	7.6×10^{-2}

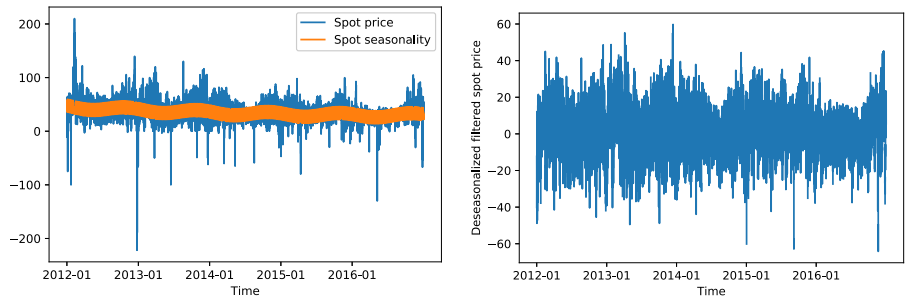
	$\tau_{0,1}$	$\tau_{1,1}$	$\tau_{2,1}$
Estimate	36744.54	1032.51	132.97
Standard error	172.85	2.12	0.30

Table 5.1: *Estimated parameters of* Γ_1.

Modelling the continuous stochastic part X_1

In a discrete time framework, X_1 is often modelled by an ARMA process. As for the deseasonalising, we remove the extreme values and ten values thereafter. The autocorrelation and the partial autocorrelation of the deseasonalised spot process without the extreme values are given in Figure 5.8a and Figure 5.8b. In a first instance, we choose to model X in a discrete-time framework by a autoregressive process of order 24:

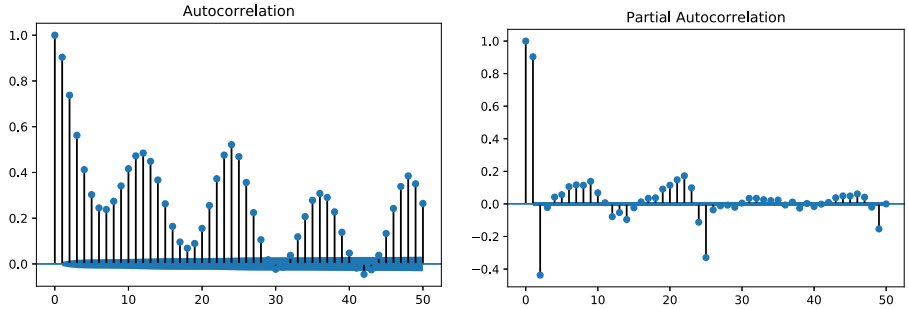
$$X_{1,t} = Q_1(D) X_{1,t} + \epsilon_{1,t},$$



(a) *Seasonality function of the spot price.* (b) *Deseasonalised and filtered spot price.*

Figure 5.7: *Seasonality function and deseasonalised spot price.*

where D is the lag operator $DX_{1,t} = X_{1,t-1}$ and $Q_1(D) = \sum_{i=1}^{24} a_{i,1} D^i$ is a polynomial of degree 24 and $\epsilon_{1,t}$ is a normal random variable with mean 0 and variance σ_1^2.



(a) *Autocorrelation function of the deseasonalized and filtered spot price.* (b) *Partial autocorrelation function of the deseasonalised and filtered spot price.*

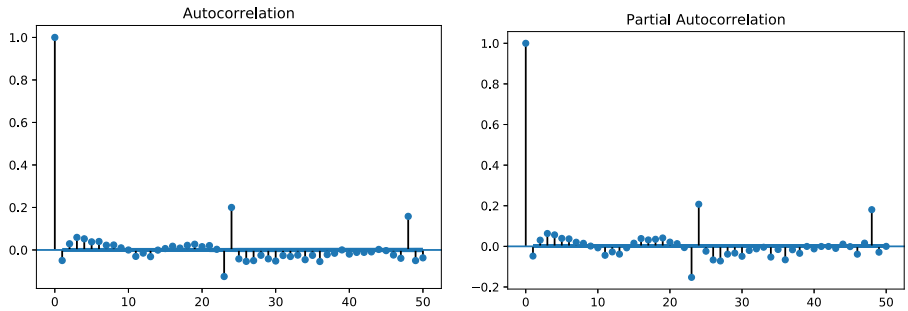
Figure 5.8: *Autocorrelation and partial autocorrelation of the deseasonalised and filtered spot price.*

The parameters estimated by exact likelihood maximisation are given in Table 5.2 together with their standard errors. The autocorrelation function and the partial autocorrelation function of the residuals are given in Figures 5.9a and 5.9b, respectively. The parameters corresponding to the lags 3, 5, 6, 10, 15, 19, 20 and 22 are not significant at level 95% but fixing these parameters to 0 leads to non-convergence of the maximisation of the likelihood. Hence all the parameters are then considered. One observes that there remains a significant autocorrelation at lags 24 and 48, due to a seasonal (diurnal) effect. This seasonal effect is considered in [7] where a seasonal autoregressive process is used to model the spot price and fits better to the data. However, considering a seasonal autoregressive process does not allow to have an equivalent stationary model in continuous time. Furthermore, contrarily to [7], we choose to not consider the fractional integrated part corresponding to a long term memory; indeed, it would consist in considering a fractional Brownian motion in an equivalent continuous-time setting and then to leave the semi-martingale framework.

In a continuous-time framework, an equivalent model is a continuous autoregressive process (CAR),

Parameter	Value	Standard error	CAR parameter ($\alpha_{i,1}$)
$a_{1,1}$	1.19	0.004	1.95
$a_{2,1}$	-0.35	0.006	38.77
$a_{3,1}$	-0.0098	0.007	67.71
$a_{4,1}$	0.027	0.007	631.48
$a_{5,1}$	-0.013	0.008	983.30
$a_{6,1}$	0.0074	0.008	5649.48
$a_{7,1}$	0.035	0.008	7804.14
$a_{8,1}$	-0.040	0.007	30499.38
$a_{9,1}$	0.048	0.007	37143.97
$a_{10,1}$	-0.0013	0.007	102868.84
$a_{11,1}$	0.081	0.007	109599.10
$a_{12,1}$	-0.080	0.007	217258.32
$a_{13,1}$	0.078	0.007	200565.84
$a_{14,1}$	-0.080	0.007	280634.70
$a_{15,1}$	-0.0084	0.007	221774.15
$a_{16,1}$	-0.030	0.008	210406.08
$a_{17,1}$	0.034	0.008	140047.42
$a_{18,1}$	-0.038	0.008	83291.14
$a_{19,1}$	0.0035	0.008	45551.82
$a_{20,1}$	0.0019	0.008	14618.89
$a_{21,1}$	-0.030	0.008	6246.01
$a_{22,1}$	0.0082	0.007	789.80
$a_{23,1}$	0.27	0.007	200.64
$a_{24,1}$	-0.14	0.004	2.71
σ_1^2	13.35	0.057	13.35

Table 5.2: *Estimated parameters of autoregressive process with order 24 and equivalent CAR parameters on the filtered deseasonalised spot price time series.*



(a) *Autocorrelation function of the spot price residuals.*

(b) *Partial autocorrelation function of the spot price residuals.*

Figure 5.9: *Autocorrelation and partial autocorrelation of the spot price residuals.*

see for instance Chapter 4 in [4]. The dynamics of X_1 is modelled by

$$X_1 = \mathbf{b}^T \mathbf{X_1},$$

$$d\mathbf{X_{1,t}} = \mathbf{A_1}\mathbf{X_{1,t}}dt + \mathbf{e}\sigma_1 dW_{1,t},$$

where

$$\mathbf{A_1} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\alpha_{24,1} & -\alpha_{23,1} & \cdots & \cdots & -\alpha_{1,1} \end{pmatrix}, \qquad \mathbf{e} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In order to obtain the parameters of the CAR process, one can infer them from the parameters of the discrete AR process using the exact discretisation of the CAR process. More details about the equivalence between those parameters are given in [4, Chapter 4]. The vector $(\alpha_{i,1})_{1\leq i\leq 24}$ corresponding to the CAR parameters is given in Table 5.2.

3.2 Model for the wind penetration index

As we have seen in Section 2.1, the wind penetration index takes values between 0 and 1, except for a small number of values exceeding 1. As the wind penetration take values between 0 and 1, we write it in the form

$$WP_t = expit\left(\Gamma_{2,t} + X_{2,t}\right),$$

with $expit\left(x\right) = \frac{1}{1+e^{-x}}$, for $x \in \mathbb{R}$. The function Γ_2 is a seasonality function and X_2 a stochastic process with mean 0. The process $X_2 + \Gamma_2$ is equal to $logit\left(WP_t\right)$ with $logit\left(x\right) = \log\left(\frac{x}{1-x}\right)$, for $x \in (0,1)$ and is observable. As for the spot price, the seasonality function is assumed to be of the form

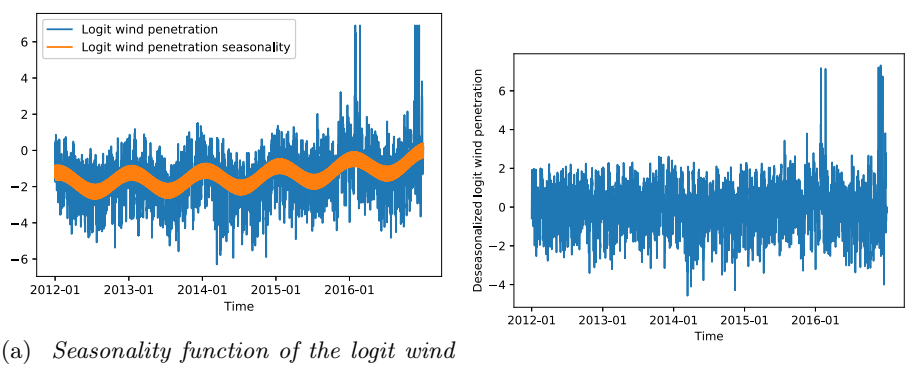
$$\Gamma_{2,t} = c_{0,2} + c_{1,2}t + c_{2,2}t^2 + c_{3,2}\cos\left(\frac{\tau_{0,2} + 2\pi t}{365 \times 24}\right) + c_{4,2}\cos\left(\frac{\tau_{1,2} + 2\pi t}{7 \times 24}\right) + c_{5,2}\cos\left(\frac{\tau_{2,2} + 2\pi t}{24}\right), \tag{5.3}$$

and the parameters are estimated by a least square minimisation. The parameter estimates are given in Table 5.3. The seasonality function and the deseasonalised time series X_2 are shown in Figures 5.10a and 5.10b, respectively.

	$c_{0,2}$	$c_{1,2}$	$c_{2,2}$	$c_{3,2}$	$c_{4,2}$	$c_{5,2}$
Estimate	-1.73	-1.06×10^{-5}	8.82×10^{-10}	0.52	0.10	0.31
Standard error	0.015	1.5×10^{-6}	3.4×10^{-11}	6.9×10^{-3}	6.9×10^{-3}	6.9×10^{-3}

	$\tau_{0,2}$	$\tau_{1,2}$	$\tau_{2,2}$
Estimate	-2203.66	0.18	-94.62
Standard error	118.37	11.18	0.54

Table 5.3: *Estimated parameters of* Γ_2.



(a) *Seasonality function of the logit wind penetration.*

(b) *Deseasonalised logit wind penetration.*

Figure 5.10: *Seasonality function and deseasonalised logit wind penetration.*

As for the spot price, let us first consider a discrete-time modelling setting. The autocorrelation function and the partial autocorrelation function, see Figures 5.11a and 5.11b, respectively, suggest that one suitable model in a discrete-time framework is an autoregressive process of the form:

$$X_{2,t} = Q_2(D) X_{2,t} + \epsilon_{2,t},$$

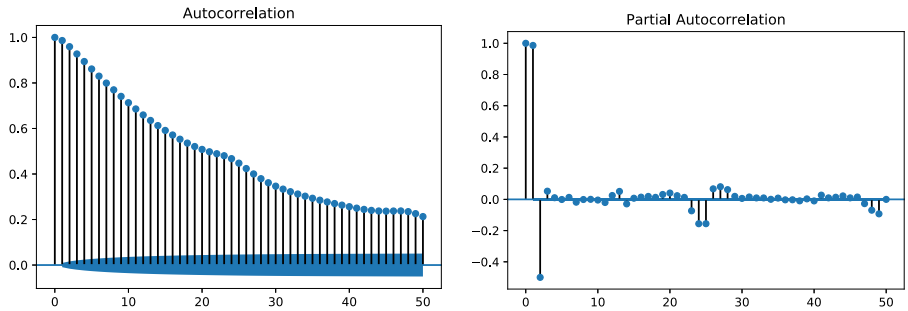
where $Q_2(D) = \sum_{i=1}^{24} a_{i,2} D^i$ is a polynomial of degree 24 and $\epsilon_{2,t}$ a normal random variable with mean 0 and variance σ_2^2. The parameters estimated with exact likelihood maximisation are given in Table 5.4. The parameters corresponding to lags 8, 15, 16, 17 and 18 are not significant, but as for the spot modelling, fixing them to 0 leads to non-convergence of the likelihood maximisation. The autocorrelation function and the partial autocorrelation function of the residuals are given in Figures 5.12a and 5.12b, respectively. We note that a significant autocorrelation at lag 24 remains due to a seasonal component, but we do not take this issue further for the same reasons as in the spot price model.

As for the spot modelling, the deseasonalised logit wind penetration X_2 is modelled by a CAR process of order 24. The dynamics of X_2 is modelled by

$$X_2 = \mathbf{b}^T \mathbf{X_2},$$

Parameter	Value	Standard error	CAR parameter ($\alpha_{i,1}$)
$a_{1,2}$	1.49	0.001	1.86
$a_{2,2}$	-0.54	0.002	38.97
$a_{3,2}$	-0.025	0.004	63.16
$a_{4,2}$	0.023	0.004	636.17
$a_{5,2}$	-0.031	0.003	894.05
$a_{6,2}$	0.036	0.003	5687.46
$a_{7,2}$	-0.017	0.004	6869.22
$a_{8,2}$	0.0028	0.006	30590.17
$a_{9,2}$	-0.017	0.005	31372.93
$a_{10,2}$	0.044	0.002	102444.15
$a_{11,2}$	-0.0162	0.002	87798.34
$a_{12,2}$	-0.084	0.004	213952.50
$a_{13,2}$	0.097	0.005	149955.10
$a_{14,2}$	-0.025	0.006	271808.63
$a_{15,2}$	-0.0090	0.007	151212.31
$a_{16,2}$	-0.0063	0.008	198823.20
$a_{17,2}$	0.034	0.008	84071.24
$a_{18,2}$	-0.010	0.008	75735.70
$a_{19,2}$	-0.018	0.008	22726.15
$a_{20,2}$	0.018	0.008	12434.60
$a_{21,2}$	-0.031	0.007	2319.68
$a_{22,2}$	0.031	0.006	570.60
$a_{23,2}$	0.16	0.004	45.37
$a_{24,2}$	-0.16	0.002	1.33
σ_2^2	0.020	0.000032	0.020

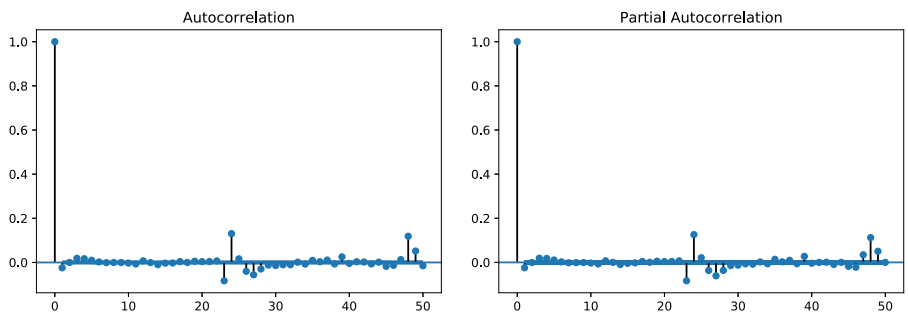
Table 5.4: *Estimated parameters of autoregressive process with order 24 and equivalent CAR parameters on the deseasonalised logit wind penetration time series.*



(a) *Autocorrelation function of the deseasonalised logit wind penetration.*

(b) *Partial autocorrelation function of the deseasonalised logit wind penetration.*

Figure 5.11: *Autocorrelation and partial autocorrelation of the deseasonalised logit wind penetration.*



(a) *Autocorrelation function of the logit wind penetration residuals.*

(b) *Partial autocorrelation function of the logit wind penetration residuals.*

$$d\mathbf{X_{2,t}} = \mathbf{A_2}\mathbf{X_{2,t}}dt + \mathbf{e}\sigma_2 dW_{2,t},$$

where

$$\mathbf{A_2} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\alpha_{24,2} & -\alpha_{23,2} & \cdots & \cdots & -\alpha_{1,2} \end{pmatrix}.$$

The parameters $(\alpha_{i,2})_{1\leq i\leq 24}$ are given in Table 5.4 using the discretisation of the CAR process.

3.3 Dependence modelling

In order to make the joint model between the spot electricity price and the wind penetration index complete, we need to specify the dependence between the two components. Here we proceed by modelling the dependence between the spot price and the wind penetration by four parameters. The first parameter ρ models a linear dependence between the continuous part of the spot and the wind penetration which is (mildly) observed in Section 2.2. The three other parameters are used

to model the intensity of the spike process as a two-valued function. This modelling is motivated by the kernel estimator computed in Section 3.1, where the test investigating whether or not the intensity of negative spikes is a constant function of the wind penetration was rejected. In the following, we will now assume that the spike process Y consists in fact of the sum of two (possibly doubly stochastic) independent Poisson-type processes: one for the negative spikes and one for the positive ones, both having the same mean reversion parameter β. Concerning the negative spikes, we observe that the intensity of negative spikes increases with the wind penetration. One simple way to take this dependence into account is to consider a regime-switching intensity function taking two different values depending on the value of the wind penetration, that is by a function of the form $\lambda : t \mapsto \lambda_{-,min}\mathbf{1}_{WP_t \leq WP_{thre}} + \lambda_{-,max}\mathbf{1}_{WP_t > WP_{thre}}$. There are two states: one state with a low intensity if the wind penetration is under a certain threshold WP_{thre}, a second state where the intensity is higher. This modelling is similar in spirit to the one proposed by [18], where two states were considered for the driven Lévy process: one for low wind penetration values and one for high wind penetration, where the distribution of the driving Lévy process in the regime with high wind penetration featured larger skewness and fatter tails than the one in the low wind penetration setting. Concerning the positive jumps, we choose to model the intensity by a constant function $\lambda : t \mapsto \lambda_+$, corresponding to a simple Poisson process.

The parameters ρ are estimated by the empirical correlation between the spot residuals and the wind residuals. To estimate the parameters $\lambda_{-,min}$, $\lambda_{-,max}$, and WP_{thre}, we minimise the L_2 distance on I between the function $\lambda : t \mapsto \lambda_{-,min}\mathbf{1}_{WP_t \leq WP_{thre}} + \lambda_{-,max}\mathbf{1}_{WP_t > WP_{thre}}$ and the kernel estimator given in Figure 5.13. The parametric intensity function as a function of the wind penetration is given in Figure 5.13. An estimator of the intensity of the positive spikes is given by the number of positive jumps, which is equal to 84. The parameters for the dependence are given in Table 5.5.

Parameters	ρ	$\lambda_{-,min}$	$\lambda_{-,max}$	WP_{thre}	λ_+
Estimate	-0.082	44.10	305.10	0.6773	84

Table 5.5: *Parameters for the dependence between the electricity spot price and the wind penetration index.*

4 Application: Impact of the dependence between electricity spot prices and wind penetration on the income of an electricity distributor

In order to round off this article we consider an application where a joint model for electricity spot prices and wind penetration is needed. To this end, let us take the point of view of an electricity distributor. This distributor settles a contract with a wind farm that produces $Q\%$ of the German and Austrian wind production, assuming that the wind in the location considered resembles the one described by the aggregated wind data which is available to us. The distributor buys the electricity from the wind farm at a fixed price K, say. Its income on this contract over a time period $[0, T]$ is then equal to

$$P = Q \int_0^T (S_t - K)\, WP_t C_t dt,$$

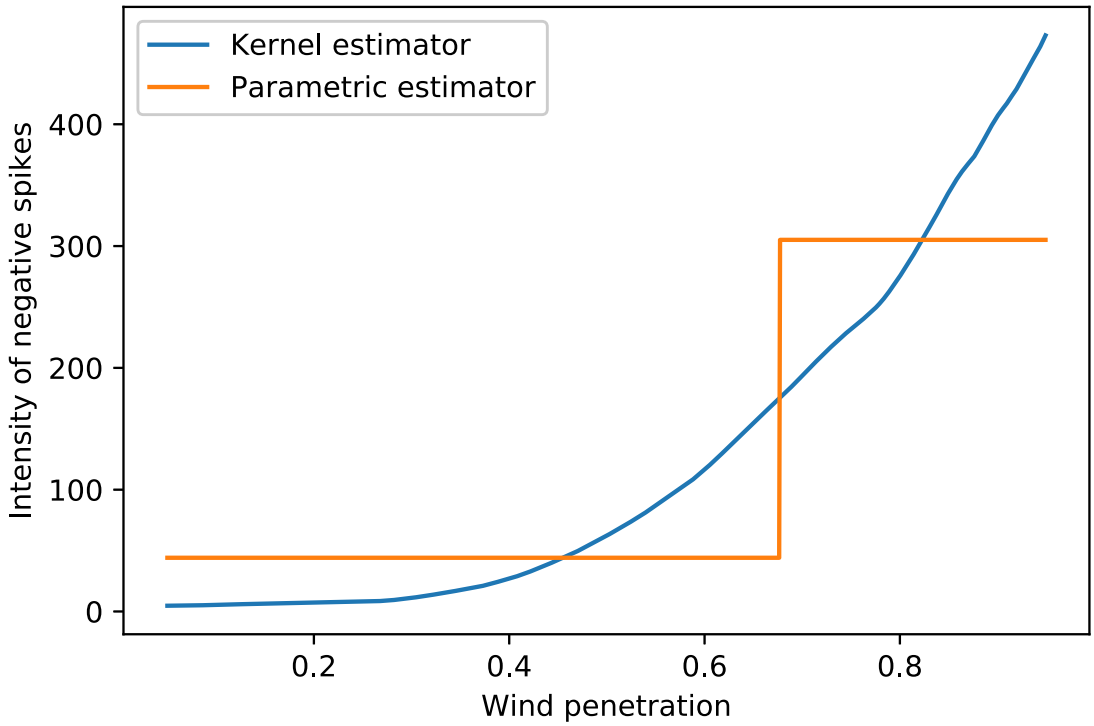


Figure 5.13: *Parametric estimator of the intensity of the negative spot spikes as a function the wind penetration.*

where C_t is the German and Austrian load at each time t. As we want to study the impact of the wind penetration on the spot prices, let us assume that the load C_t is deterministic, knowing this assumption is rather strong. Indeed, the spot price and the load are in fact dependent, but only the dependence arising through the seasonality is considered in the following.

We are interested in the extreme values in the prices and their impact on the distributor's income. We shall consider two risk measures which are widely used in practice: the value at risk (VaR) and the expected shortfall (ES). The value at risk at level $\alpha \in (0,1)$, denoted by $VaR_\alpha(P)$ corresponds to the maximal loss given the confidence level $1-\alpha$, that is the quantile of order α of the distribution of P. The expected shortfall at level $\alpha \in (0,1)$ corresponds to the expected loss in the tail distribution and is defined by $\mathbb{E}(P|P \leq VaR_\alpha(P))$. We choose to work with the levels $\alpha = 95\%$ and $\alpha = 99\%$ that are often considered in finance.

In order to study the impact of the dependencies between electricity spot prices and wind penetration, we consider two models: The first one is the one considered in Section 3 where the intensity of the negative electricity spot spikes can take two values depending on the wind penetration level. It is referred to as the two-state model in the following. The second one differs by considering a constant intensity function for the negative electricity spot prices and is referred to as the one-state model. The value of the estimated intensity in the second model is equal to 30.

Let us assume that the initial time is the 1st of January 2017 and that the maturity T is equal

to the 31st of December 2017. The consumption is chosen as its seasonality during the year 2017: this seasonality is estimated using German hourly load data between the 1st of January 2012 and the 31st December 2016 using the same parametric form as Γ_1 and Γ_2. The spot and the wind are simulated using the two different models. Q is chosen as 1% and the strike to be 30. Results for the different risk measures are given in Table 5.6. These values are computed with the Monte Carlo method using $N = 100,000$ simulated paths. The confidence interval for the different risk measures is given at level 95%. The confidence intervals for the value at risk and the expected shortfall are computed using bootstrap, with $M = 100,000$ new samples. The bootstrap method consists of randomly drawing M new samples of size N from the simulated one, using the empirical distribution function as the probability measure, and to compute the different quantities we are interested in, here the risk measures, for each sample. We then obtain M values for these quantities and the confidence interval for these quantities at level 95% is defined by the values between the quantile of order 2.5% and the quantile of order 97.5% of these M values. The reader can refer to [16], see Chapter 23, p. 326, for more details about the bootstrap method.

Model	One-state	Two-states
Expectation	$[742451.32, 754842.31]$	$[642085.23, 654479.20]$
VaR 95%	$[-906630.39, -880902.99]$	$[-1009660.00, -978858.45]$
VaR 99%	$[-1596797.78, -1553512.84]$	$[-1697907.39, -1647565.64]$
ES 95%	$[-1323982.81, -1293678.85]$	$[-1426258.31, -1395788.60]$
ES 99%	$[-1932779.14, -1876820.83]$	$[-2036072.86, -1979485.25]$

Table 5.6: *Different risk measures for the portfolio with strike equal to 30.*

We observe that modelling the intensity of the negative spikes as a function of wind penetration has an impact on the expectation of P, the values at risk at levels 95% and 99% and the expected shortfalls at levels 95% and 99% of the portfolio: they are lower than in the case when the intensity is constant. Indeed, the portfolio is a function of WP_tS_t, and negative spikes for S_t appear more often when WP_t is high.

Our intuition gained from the above application is that if one were to use a more complex parametric model than the two-state model for the intensity function considered above, then the impact of the dependence modelling could potentially increase which could lead to more extreme values in the corresponding electricity spot prices. It will be worth exploring this aspect in more detail in future research.

Bibliography

[1] Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009.

[2] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[3] Ole Eiler Barndorff-Nielsen, Neil Shephard, and Matthias Winkel. Limit theorems for multipower variation in the presence of jumps. *Stochastic processes and their applications*, 116(5):796–806, 2006.

[4] Fred Espen Benth and Jūratė Šaltytė Benth. *Modeling and pricing in financial markets for weather derivatives*, volume 17. World Scientific, 2012.

[5] Sam Forrest and Iain MacGill. Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian national electricity market. *Energy Policy*, 59:120–132, 2013.

[6] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.

[7] Niels Haldrup and Morten Ørregaard Nielsen. A regime switching long memory model for electricity prices. *Journal of Econometrics*, 135(1):349–376, 2006.

[8] Ronald Huisman, Christian Huurman, and Ronald Mahieu. Hourly electricity prices in day-ahead markets. *Energy Economics*, 29(2):240–248, 2007.

[9] Tryggvi Jónsson, Pierre Pinson, and Henrik Madsen. On the market impact of wind energy forecasts. *Energy Economics*, 32(2):313–320, 2010.

[10] Tryggvi Jónsson, Pierre Pinson, Henrik Aalborg Nielsen, Henrik Madsen, and Torben Skov Nielsen. Forecasting electricity spot prices accounting for wind power predictions. *Sustainable Energy, IEEE Transactions on*, 4(1):210–218, Jan 2013.

[11] Janina C. Ketterer. The impact of wind power generation on the electricity price in Germany. *Energy Economics*, 44:270–280, 2014.

[12] Claudia Klüppelberg, Thilo Meyer-Brandis, and Andrea Schmidt. Electricity spot price modelling with a view towards extreme spike risk. *Quantitative Finance*, 10:963–974, 2010.

[13] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, pages 1–38, 2016.

[14] Jia Li, Viktor Todorov, and George Tauchen. Volatility occupation times. *The Annals of Statistics*, 41(4):1865–1891, 2013.

[15] Jia Li, Viktor Todorov, and George Tauchen. Jump regressions. *Econometrica*, 85(1):173–195, 2017.

[16] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 1998.

[17] Almut Elisabeth Dorothea Veraart. Inference for the jump part of quadratic variation of Itô semimartingales. *Econometric Theory*, 26(2):331–368, 2010.

[18] Almut Elisabeth Dorothea Veraart. Modelling the impact of wind power production on electricity prices by regime-switching Lévy semistationary processes. In Fred Espen Benth and Giulia Di Nunno, editors, *Stochastics of Environmental and Financial Economics*, pages 321–340. Springer, 2016.

[19] Almut Elisabeth Dorothea Veraart and Luitgard Anna Maria Veraart. Modelling electricity day-ahead prices by multivariate Lévy semi-stationary processes. In Fred Espen Benth, Valery Kholodnyi, and Peter Laurence, editors, *Wolfgang Pauli Proceedings*, pages 157–188. Springer, 2014.

5. *A joint model for electricity spot prices and wind penetration with dependence in the extremes*

Résumé

Cette thèse traite de problèmes de dépendance entre processus stochastiques en temps continu. Ces résultats sont appliqués à la modélisation et à la gestion des risques des marchés de l'électricité.

Dans une première partie, de nouvelles copules sont établies pour modéliser la dépendance entre deux mouvements Browniens et contrôler la distribution de leur différence. On montre que la classe des copules admissibles pour les Browniens contient des copules asymétriques. Avec ces copules, la fonction de survie de la différence des deux Browniens est plus élevée dans sa partie positive qu'avec une dépendance gaussienne. Les résultats sont appliqués à la modélisation jointe des prix de l'électricité et d'autres commodités énergétiques.

Dans une seconde partie, nous considérons un processus stochastique observé de manière discrète et défini par la somme d'une semimartingale continue et d'un processus de Poisson composé avec retour à la moyenne. Une procédure d'estimation pour le paramètre de retour à la moyenne est proposée lorsque celui-ci est élevé dans un cadre de statistique haute fréquence en horizon fini. Ces résultats sont utilisés pour la modélisation des pics dans les prix de l'électricité.

Dans une troisième partie, on considère un processus de Poisson doublement stochastique dont l'intensité stochastique est une fonction d'une semimartingale continue. Pour estimer cette fonction, un estimateur à polynômes locaux est utilisé et une méthode de sélection de la fenêtre est proposée menant à une inégalité oracle. Un test est proposé pour déterminer si la fonction d'intensité appartient à une certaine famille paramétrique. Grâce à ces résultats, on modélise la dépendance entre l'intensité des pics de prix de l'électricité et de facteurs exogènes tels que la production éolienne.

Mots Clés

Dépendance, Copule, Mouvement Brownien, Statistique haute fréquence, Semimartingale, Processus de Poisson, Intensité stochastique, Estimation non paramétrique, Estimateur à polynômes locaux, Sélection de fenêtre, Inégalité oracle, Marchés de l'électricité, Pics, Production éolienne, Gestion des risques, Finance mathématique.

Abstract

In this thesis, we study some dependence modeling problems between continuous time stochastic processes. These results are applied to the modeling and risk management of electricity markets.

In a first part, we propose new copulae to model the dependence between two Brownian motions and to control the distribution of their difference. We show that the class of admissible copulae for the Brownian motions contains asymmetric copulae. These copulae allow for the survival function of the difference between two Brownian motions to have higher value in the right tail than in the Gaussian copula case.
Results are applied to the joint modeling of electricity and other energy commodity prices.

In a second part, we consider a stochastic process which is a sum of a continuous semimartingale and a mean reverting compound Poisson process and which is discretely observed. An estimation procedure is proposed for the mean reversion parameter of the Poisson process in a high frequency framework with finite time horizon, assuming this parameter is large. Results are applied to the modeling of the spikes in electricity prices time series.

In a third part, we consider a doubly stochastic Poisson process with stochastic intensity function of a continuous semimartingale. A local polynomial estimator is considered in order to infer the intensity function and a method is given to select the optimal bandwidth. An oracle inequality is derived. Furthermore, a test is proposed in order to determine if the intensity function belongs to some parametrical family. Using these results, we model the dependence between the intensity of electricity spikes and exogenous factors such as the wind production.

Keywords

Dependence, Copula, Brownian motion, High frequency statistics, Semimartingale, Poisson process, Stochastic intensity, Non parametric estimation, Local polynomial estimation, Bandwidth selection, Oracle inequality, Electricity markets, Spikes, Wind production, Risk management, Mathematical finance.