

**NON-PARAMETRIC ESTIMATION OF STOCHASTIC  
DIFFERENTIAL EQUATIONS**

---

A Dissertation Presented to  
the Faculty of the Department of Mathematics  
University of Houston

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

By  
François Alexis Adon Ouegnin

August 2017

# NON-PARAMETRIC ESTIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

---

François Alexis Adon Ouegnin

APPROVED:

---

Dr. Ilya Timofeyev, Committee Chair  
Department of Mathematics, University of Houston

---

Dr. William Ott,  
Department of Mathematics, University of Houston

---

Dr. Andrew Török,  
Department of Mathematics, University of Houston

---

Dr. Gregor Kovacic,  
Department of Mathematical Sciences,  
Rensselaer Polytechnic Institute

---

Dean, College of Natural Sciences and Mathematics

# Acknowledgements

## *It takes a village to raise a child (African Proverb)*

As suggested by this African proverb, this thesis work would not have been possible without support from the UH community, friends and family. The child, last metamorphosis in Nietzsche's philosophy, that I am is thankful to numerous people. First and foremost, as a christian, I would like to thank Almighty God and his son Jesus Christ. Everything is possible with them on your side. Thank you so much for the grace and blessings every single day of the last 5 years in Houston, Texas, USA. This dissertation would not have been possible without the help of my advisor Dr Ilya Timofeyev. I am grateful to him for his insightful suggestions and his patient assistance in the writing of this thesis. I would like to also express my gratitude to the committee members, Dr. Ott, Dr. Török, and Dr. Kovacic. I would like to thank all the people who honored me with their friendship during my stay in Houston. The list is long and I apologize if I do not mention someone's name. I thank all the UH community for the support. In particular, I would like to acknowledge:

- All current UH graduate students of the Math department, in particular: Duong, Kayla, Xi, Peng, Nikolaos, Kazem, Krithika, Kate, Daewa, Dipanwita, Alan, Deep, Mo, Malcolm, Grant, Sergey, Rasoul, Saeed, Amir, Adrian, Homayoon, Zhenhua, Manyang, Sabrine, Megan, Ang, Robert, Worawit, Peixin, Wanli, HG, and others.
- All the faculty and staff of the Math department: in particular Drs Timofeyev, Ott, Török, Blecher, He, Auchmuty, Kao, Pan, Perepelitsa, Climenhaga,

---

Quaini, and Gorb. In the staff, Rachel, Tai, Mariela, Sharon, Chris, Eden, Amy, Lawrencene, Neha, Emilio, Yolanda, Seth, Dave, Jason, and others. Thanks also to Cindy at CASA, and her staff.

- At the College of Natural Sciences and Mathematics(NSM) level: Thanks to Dr Anne Delcour.
- Former UH students: Sibusiso, Tasadduk, Shang-Huang, Brett, Alex, Daniel, Changan, Shujiao, Taoufik, Mauricio, Pankaj, Pooran, Mrinal, Satish, Ben, my former roommate Murad, and many others.
- Friends in USA, Canada and Côte d'Ivoire: Saumitra Saha, Abdoul Saddy my childhood friend, Douwere Grekou (thanks Doc for the weekly call and the constant support during these 5 years), Aicha Coulibaly, Roger, and Helene Adou, Sonia Thalmas, Diane Adou, Jean Noel, and Caroline Ngouan, and others.

This dissertation is first dedicated to my loving and supportive wife Karoll and my children Maëly Gisele and Georges-Armand. To my kids about Doctoral studies always remember: *'Grandpa did it, Dad did it, Aunt Maëva did it; so you can do it'*. It is secondly dedicated to my caring parents, Georges-Armand and Gisele; my brothers and sisters, Christian, Ludovic, Melissa, and Maëva; my sister-in-law Aurelia, and my nieces Jayia and Maya; and my entire family in Côte d'Ivoire. And thirdly, this dissertation is dedicated to the memory of my daughter Nolwenn and my two grandfathers François Maurice Adon Ouegnin and Pascal Alexis Bailly.

# NON-PARAMETRIC ESTIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

---

An Abstract of a Dissertation Presented to  
the Faculty of the Department of Mathematics  
University of Houston

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

By  
François Alexis Adon Ouegnin  
August 2017

# Abstract

Parametric estimation techniques are commonly used, in academia and industry, to estimate the drift and diffusion of Stochastic Differential Equations (SDE). Their major limitation is that they require a functional form for the drift and diffusion terms to be implemented. Such assumptions increase the risk of misspecification. Non-parametric methods on the other hand allow the estimation of both components without a priori assumptions. In this study, we first use a non-parametric technique based on eigenvalues and eigenfunctions to study the reconstruction of the drift and diffusion. In a second part with another non-parametric method, based on conditional expectation, we study the estimation errors generated by the space-time discretization necessary for the estimation. The work on these two non-parametric estimations constitute the two parts of this thesis. In the first part, we motivate the use of non-parametric techniques to model time series data in real applications, explain the spectral reconstruction, and propose a methodology to extend its use to processes commonly used in finance. A real world application on intraday data is presented. In the latter, the components of a SDE driving the crude oil price are reconstructed for the following two periods: 2010-2013 and 2015-2016. A mean reverting process is identified for the first period whereas a random walk hypothesis failed to be rejected for the second period. The reconstruction is sensitive to the sample size and the discretization. Estimating drift and diffusion from discretely sampled data is fraught with the potential for errors from space-time discretization. Therefore, in the second part of this thesis we study in the  $L^2$  sense the impact of the space-time discretization on the estimation errors of conditional expectation

---

based estimators. We concentrate this analysis on the case of Ornstein-Uhlenbeck process. We found that for time series data observed at fixed interval, the choice of an optimal space discretization influences the rate at which the errors decay. When the sample size is known, an upper bound for the errors is obtained analytically and verified using numerical simulations. We also propose upper bounds for the case of unknown sample size.

---

# Contents

---

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Outline . . . . .	4
<b>2</b>	<b>SPECTRAL RECONSTRUCTION OF DRIFT AND DIFFUSION</b>	<b>6</b>
2.1	Motivation . . . . .	7
2.2	Background . . . . .	8
2.2.1	Some processes used in finance . . . . .	10
2.2.2	Affine processes . . . . .	14
2.3	Important concepts in stochastic processes . . . . .	17
2.4	Spectral Reconstruction of one SDE . . . . .	20
2.4.1	Construction of the transition probability matrix $P$ . . . . .	20
2.4.2	Finite Difference (FD) approach . . . . .	22
2.5	Periodic boundary conditions and reconstruction . . . . .	23
2.5.1	Periodic boundary conditions with Fourier filter . . . . .	25

2.5.2	Periodic boundary conditions with Parzen windows . . . . .	29
2.6	Extension to non periodic drift and diffusion functions . . . . .	39
2.6.1	Testing reconstruction using a known generator . . . . .	39
2.6.2	The number of bins or spatial step size . . . . .	51
2.6.3	Reconstruction using simulated data of affine processes . . . . .	56
2.7	Improved reconstruction . . . . .	66
2.7.1	Large data sets . . . . .	66
2.7.2	Errors with subsampling and spatial steps . . . . .	72
2.8	Application: Modeling energy commodity prices dynamics . . . . .	87
2.8.1	Motivation . . . . .	87
2.8.2	Literature review . . . . .	88
2.8.3	Description of crude oil price data . . . . .	89
2.8.4	Model of crude oil price dynamics . . . . .	94
2.8.5	Construction of the Markov chain and parameters selection . . . . .	95
2.8.6	Spectral estimation applied to crude oil price data . . . . .	96
2.9	Conclusion Chapter 2 . . . . .	105
<b>3</b>	<b>CONDITIONAL EXPECTATION BASED RECONSTRUCTION OF DRIFT AND DIFFUSION</b>	<b>107</b>
3.1	Introduction-Chapter 3 . . . . .	108
3.2	Error Analysis . . . . .	111
3.2.1	Conditional expectation with fixed number of points . . . . .	111
3.2.2	Crude error estimates . . . . .	113
3.2.3	Refined error estimates . . . . .	119
3.2.4	Numerical investigation . . . . .	133
3.2.5	Relaxing assumption of the fixed number of points . . . . .	140
3.3	Conditional expectation based estimators of the drift and diffusion . . . . .	154

*CONTENTS*

---

3.4 Conclusion Chapter 3 . . . . .	162
<b>A APPENDIX-CHAPTER 2</b>	<b>164</b>
<b>BIBLIOGRAPHY</b>	<b>169</b>

---

## List of Figures

---

- 2.1 The results of an estimation using only the generator not the adjoint and a Fourier filtering of the eigenfunctions are given in this plot. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve. . . . . 27
- 2.2 The results of an estimation using only the adjoint operator not the generator and a Fourier filtering of the eigenfunctions are given in this figure. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve. . . . . 27

2.3 The results of an estimation using the generator, the adjoint operator and a Fourier filtering of the eigenfunctions are given in these plots. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve. . . . . 28

2.4 The results of an estimation using the entire expression  $E(b, a)$  in (2.29) and a Fourier filtering of the eigenfunctions are given in these plots. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve. . . . . 28

2.5 Reconstruction of the drift and diffusion using only the generator and  $parzen_w(3, 2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift. . . . . 33

2.6 Reconstruction of the drift and diffusion using only the adjoint and  $parzen_w(3, 2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example  $x = 2$  more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some term. . . . . 33

2.7	Reconstruction of the drift and diffusion using the generator and the adjoint as well as $parzen_w(3, 2)$ . The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis provides the values of $b(x)$ and $a(x)$ on the left and right plots respectively. The true values are given by the curve and the reconstructed values by the circles. On the left graph, the reconstructed drift is better than the one in Figure 2.5 but less accurate than in Figure 2.6. On the right side, the diffusion shows similar accuracy than in Figure 2.6. . . . . .	34
2.8	Reconstruction of the drift and diffusion using the entire error function (2.29) and $parzen_w(3, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis provides the values of $b(x)$ and $a(x)$ on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.7 implying than the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figure 2.6 and 2.7. . . . . .	34
2.9	Reconstruction of the drift and diffusion using only the generator and $parzen_w(5, 2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis provides the values of $b(x)$ and $a(x)$ on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift. . . . .	35
2.10	Reconstruction of the drift and diffusion using only the adjoint and $parzen_w(5, 2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis provides the values of $b(x)$ and $a(x)$ on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example $x = 0$ more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some constant term. . . . .	35

2.11 Reconstruction of the drift and diffusion using the generator and the adjoint as well as  $parzen_w(5, 2)$ . The true values are given by the curve and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift is better than the one in Figure 2.9 but less accurate than in Figure 2.10. On the right side, the diffusion shows similar accuracy than in Figure 2.10. . . . . 36

2.12 Reconstruction of the drift and diffusion using the entire error function (2.29) and  $parzen_w(5, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.11 implying than the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figures 2.10 and 2.11. . . . . 36

2.13 Reconstruction of the drift and diffusion using only the generator and  $parzen_w(7, 2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift. . . . . 37

2.14 Reconstruction of the drift and diffusion using only the adjoint and  $parzen_w(7, 2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example  $x = 0$  more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some constant term. . . . . 37

2.15 Reconstruction of the drift and diffusion using the generator and the adjoint as well as  $parzen_w(7, 2)$ . The true values are given by the curve and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift is better than the one in Figure 2.13 but less accurate than in Figure 2.14. On the right side, the diffusion shows similar accuracy than in Figure 2.14. . . . . 38

2.16 Reconstruction of the drift and diffusion using the entire error function (2.29) and  $parzen_w(7, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.15 implying than the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figures 2.14 and 2.15. . . . . 38

2.17 Reconstruction of the drift and diffusion using only the generator not the adjoint of a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively. . . . . 45

2.18 Reconstruction of the drift and diffusion using only the adjoint not the generator of a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively . . . . . 45

2.19 Reconstruction of the drift and diffusion using the error function (2.29) and a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively . . . . . 46

2.20 Reconstruction of the drift and diffusion using the generator, not the adjoint, and a discretized infinitesimal generator without boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively. . . . . 46

2.21	Reconstruction of the drift and diffusion using only the adjoint and not the generator with a discretized infinitesimal generator without boundary conditions. The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis shows on the left and right plots the values of $b(x)$ and $a(x)$ respectively. . . . .	47
2.22	Reconstruction of the drift and diffusion using the error function (2.29) and a discretized infinitesimal generator without boundary conditions. The x-axis gives the values, $x$ , taken by the process $X_t$ . The y-axis shows on the left and right plots the values of $b(x)$ and $a(x)$ respectively. . . . .	47
2.23	Reconstruction of the drift of the OU process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $b(x)$ . The true values are given by the line and estimates by the circles. . . . .	48
2.24	Reconstruction of the diffusion of the OU process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $a(x)$ . The true values are given by the line and estimates by the circles. . . . .	48
2.25	Reconstruction of the drift of a CIR type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $b(x)$ . The true values are given by the line and estimates by the circles. . . . .	49
2.26	Reconstruction of the diffusion of a CIR type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $a(x)$ . The true values are given also by the line and estimates by the circles. . . . .	49
2.27	Reconstruction of the drift of a Jacobi type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $b(x)$ . The true values are given by the line and estimates by the circles. . . . .	50

2.28	Reconstruction of the diffusion of a Jacobi type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values, $xt$ , taken by the process. The y-axis shows the values of $a(x)$ . The true values are given by the line and estimates by the circles. . . . .	50
2.29	Reconstruction of the drift of an OU process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $xt$ , taken by the process whereas on the y-axis values of $b(x)$ are given. The true values are given by the line and estimates by the circles. Large approximation errors appear on the large positive values of the process . . . . .	63
2.30	Reconstruction of the diffusion of an OU process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $xt$ , taken by the process whereas on the y-axis values of $a(x)$ are given. The true values are given by the line and estimates by the circles. Large approximation errors appear at the extreme values of the process . . . . .	63
2.31	Reconstruction of the drift of a CIR process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $xt$ , taken by the process whereas on the y-axis values of $b(x)$ are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process. . . .	64
2.32	Reconstruction of the drift of a CIR process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $xt$ , taken by the process whereas on the y-axis values of $a(x)$ are given. The true values are given by the line and estimates by the circles. Diffusion for small values of the process are better approximated than for large values. . . . .	64
2.33	Reconstruction of the drift of a Jacobi process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $xt$ , taken by the process whereas on the y-axis values of $b(x)$ are given. The true values are given by the line and estimates by the circles. These parameters fail to provide reconstruction. . . . .	65

2.34	Reconstruction of the diffusion of a Jacobi process using simulated data using $T = 10^6$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The x-axis gives different values, $x_t$ , taken by the process whereas on the y-axis values of $a(x)$ are given. The true values are given by the line and estimates by the circles. The shape of the diffusion is recovered but outliers appear when the process takes values around the long-term mean. . . . .	65
2.35	Reconstruction of the drift of an OU process using simulated data using $T = 10^7$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The values taken by the process are given by $x_t$ on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. The parameters provide an overestimated reconstruction. Large errors appear when the process takes extreme values. . . . .	69
2.36	Reconstruction of the diffusion of an OU process using simulated data using $T = 10^7$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 40$ . The values taken by the process are given by $x_t$ on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. The parameters provide an overestimated reconstruction of the diffusion. Large errors appear when the process takes extreme values. . . . .	69
2.37	Reconstruction of the drift of a CIR process using simulated data using $T = 10^7$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 30$ . The values taken by the process are given by $x_t$ on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process. . . . .	70
2.38	Reconstruction of the diffusion of a CIR process using simulated data using $T = 10^7$ , $dt = 10^{-4}$ , $h = 0.01$ , $M = 30$ . The values taken by the process are given by $x_t$ on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. Large estimation errors could be found for large values taken by the process. . . . .	70

2.39 Reconstruction of the drift of a Jacobi process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $x_t$  on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process. . . . . 71

2.40 Reconstruction of the drift of a Jacobi process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $x_t$  on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. Large errors could be found for values close to the long-term mean. . . . . 71

2.41 The number of bins for the OU process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. Number of bins between 20 and 40 provides less errors. When the number of bins is greater than 40, more outliers appear. . . . . 75

2.42 The number of bins for the OU process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set to 10. Observe that 40 bins give the smallest estimation errors for the diffusion. . . . . 75

2.43 The number of bins for the CIR process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. Observe that compared to the OU process, the magnitude of errors made in the drift estimation is at least twice higher. The median is multiplied by at least 2 just by increasing the number of bins from 20 to 30. . . . . 76

2.44 The number of bins for the CIR process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set to 10. The estimation errors are also larger than those of the OU. Small number of bins, between 20 and 40, gives less errors. . . . . 76

2.45 The number of bins for the JAC process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. The error value, when the number of bins is between 20 and 40, is smaller than 20. The error increases fourfold when the number of bins is greater than 40. . . . . 77

2.46 The number of bins for the JAC process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set 10. Same pattern than the drift, a number of bins between 20 and 40 gives a better approximation. . . . . 77

2.47 The  $error_{2,drift}$  for different subsampling times is given for the three affine processes after some values have been truncated. On the x-axis the time interval between consecutive values of the process, the subsampling time,  $h$ , is given. On the y-axis the values of the error  $error_{2,drift}$  are provided. We can see that the error on the drift of the Jacobi is smaller than the one of the other processes. Around  $h = 0.01$  the approximation error of the drift is minimal for the processes. . . . 83

2.48 The  $error_{2,diff}$  for different subsampling times is given for the 3 affine processes after some values have been truncated. On the x-axis the time interval between consecutive values of the process, the subsampling time,  $h$ , is given. On the y-axis the values of the error  $error_{2,diff}$  are provided. We can see that the diffusion is better reconstructed for all 3 process when the subsampling time is close to 0.01. However when the subsampling time,  $h$ , is large, the Jacobi process generates the largest estimation errors for the diffusion. . . . . 83

2.49 Reconstruction of the drift of an OU process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 84

2.50 Reconstruction of the diffusion of an OU process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 84

LIST OF FIGURES

---

2.51 Reconstruction of the drift of a CIR process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 85

2.52 Reconstruction of the diffusion of a CIR process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 85

2.53 Reconstruction of the drift of a Jacobi process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 86

2.54 Reconstruction of the diffusion of a Jacobi process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles. . . . . 86

2.55 WTI 1-minute evolution from November 2010 to December 2016. The value of the benchmark is given by the average of the open and close bid at each minute. Graphically, one can distinguish 3 different regimes: the regime 1 with a value of WTI around \$ 90 from November 2010 to June 2014, regime 2 from June 2014 to January 2015 a fall of the price. And a regime 3 from January 2015 to December 2016 where the WTI oscillates around \$ 40 mark. . . . . 92

2.56 Period 1: High oil price regime, November 2010- to December 2013. This period is dominated by a WTI \$ 95 with a maximum value of \$114.83 on May 2<sup>nd</sup> 2011 after an increase of production by Saudi Arabia. The minimum price in this period was \$ 75.05 in October 04<sup>th</sup> 2011 due to recession fears. . . . . 92

2.57 Period 2: 2014 the transition year. The price oscillates around \$100 from January 2014 until October 07<sup>th</sup> 2014 when it dropped below the \$90 mark. The maximum value of \$ 107.64 has been reached on June 12<sup>th</sup> 2014 while the lowest value \$ 52.51 was on December 31<sup>st</sup>2014 . . . . . 93

*LIST OF FIGURES*

---

2.58 Period 3: Low oil price regime, from January 2015 to December 2016. The maximum value was \$ 62.45 occurred on May 6<sup>th</sup> 2015 and the minimum value was \$26.05 on February 11<sup>th</sup> 2016 . . . . . 93

2.59 Full data histogram of the period 2010-2013. The log price is on the x-axis and the number of observations on y-axis. . . . . 97

2.60 Truncated data histogram for the period 2010-2013. The log price is given on the x-axis and the number of observations on y-axis. We removed values on the both left and right tails such that each bin has at least 2000 observations. . . . . 97

2.61 Drift estimates with  $h = 5dt$ ,  $\Delta x = 0.0092$ , and  $M = 40$  for period from 2010 to 2013. The log price is given on the x axis and the value of the drift on the y axis. The estimates are in circles and the linear fit in line. The estimates are centered around a linear line with decreasing slope. This supports a mean reversion assumption for this period. . . 100

2.62 Diffusion estimates with  $h = 5dt$ ,  $\Delta x = 0.0092$ ,  $M = 40$  for period from 2010 to 2013. The log price is given on the x axis and the value of the diffusion on the y axis. The estimates are in circles and the linear fit in line. The estimates are distributed around a quadratic curve. This supports a non constant diffusion assumption for this period. . . 100

2.63 Full histogram for the period 2015-2016. The left tail contains bins with less than 2000 observations. Similarly on the right tail when the log price is close to 4, bins have a small number of observations. . . . 102

2.64 Truncated histogram for the period 2015-2016. We removed the bins in both the left and right tails that contain less than 2000 observations. 102

2.65 Drift estimate with  $h = 5dt$ , and  $M = 60$ . The log price is given on the x axis and the value of the drift on the y axis. The estimates are in circles and the linear fit in line. The estimates are centered around the zero line what supports the random walk assumption. . . . . 104

2.66 Diff estimate with  $h = 5dt$  and  $M = 60$ . The log price is given on the x axis and the value of the diffusion on the y axis. The estimates are in circles and the linear fit in line. A quadratic fit well approximates the general distribution of the data, what supports the rejection of a constant drift assumption. . . . . 104

LIST OF FIGURES

---

3.1 This graph shows the autocorrelation of order 1,2, and 3 of the data recorded in the first bin formed by the different partition of the interval  $[\mu_X - \sigma_X, \mu_X + \sigma_X]$  into bins. After dividing the interval by 4 we have no autocorrelation of order 2 and 3. When the interval is divided by 7 the data does not show autocorrelation. . . . . 123

3.2 The QQplot shows that the data inside the first bin are normal around the mean but differ greatly in the tails. . . . . 123

3.3 Error and bounds for  $\hat{m}_1$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . The number of points per bin,  $M_{c_j}$ , is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The errors are closely bounded by the refined bound whereas the crude one is about five times higher than the error and refined bound. . . . . 135

3.4 Error and bounds for  $\hat{m}_2$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . The number of points per bin,  $M_{c_j}$ , is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The errors are relatively smaller than both crude and refined bounds. . . . . 135

3.5 Log error vs  $M_{c_j}$ ,  $\hat{m}_1$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . In the log log scale, the number of bins is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The  $\frac{1}{\sqrt{M}}$  decay rate is confirmed for  $\hat{m}_1$ . . . . . 136

3.6 Log error vs  $M_{c_j}$ ,  $\hat{m}_2$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . In the log log scale, the number of bins is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The  $\frac{1}{\sqrt{M}}$  decay rate is confirmed for  $\hat{m}_2$ . . . . . 136

3.7 The errors for  $\hat{m}_1$  is plotted with the two bounds. The subsampling time  $\Delta t$  is fixed at 0.01 and the spatial step size is  $\Delta x = \frac{1}{M_{c_j}}$ . Both bounds are above the error values. . . . . 138

3.8 The errors for  $\hat{m}_2$  is plotted with the two bounds. The subsampling time  $\Delta t$  is fixed at 0.01 and the spatial step size is  $\Delta x = \frac{1}{M_{c_j}}$ . Again both bounds are close to the value of the time step  $\Delta t$ . The errors are smaller than the bounds. . . . . 138

3.9 The error of  $\hat{m}_1$  in the log log scale when  $\Delta t = 0.01$  and  $\Delta x = \frac{1}{M_{c_j}}$ . The slope of the error is  $-0.649 \pm 0.043$ . The slope of the log errors of  $\hat{m}_1$  being smaller than  $-0.5$  implies that the estimation errors decay a little bit slower than expected. . . . . 139

*LIST OF FIGURES*

---

- 3.10 The error of  $\hat{m}_2$  in the log log scale when  $\Delta t = 0.01$  and  $\Delta x = \frac{1}{M_{c_j}}$ . The slope of the error is equal to  $-0.692 \pm 0.068$ . The slope of the log errors of  $\hat{m}_2$  being smaller than  $-0.5$  implies that the estimation errors decay a little bit slower than expected. . . . . 139

---

## List of Tables

---

2.1	Comparison of the eigenvalues obtained by simulating ensemble of trajectories with the one in [18] . . . . .	25
2.2	Comparison before and after Fourier filtering of the difference of eigenpairs and their derivatives with those of the discretized generator . . .	26
2.3	Comparison of the eigenvalues using a transition probability computed with $parzen_w$ with those [18] . . . . .	31
2.4	Comparison of the difference of eigenpairs and their derivatives obtained using $parzen_w(3, 2)$ with those of a discretized generator. . . .	32
2.5	Comparison of the difference of eigenpairs and their derivatives obtained using $parzen_w(5, 2)$ with those of a discretized generator. . . .	32
2.6	Comparison of the difference of eigenpairs and their derivatives obtained using $parzen_w(7, 2)$ with those of a discretized generator. . . .	32
2.7	The eigenvalues obtained from the discretized generator for the three affine processes is compared to their real values. . . . .	44
2.8	The eigenvalues obtained from data-based Markov chain for the three affine processes are compared to their real value. . . . .	60
2.9	Comparison the difference of eigenpairs and their derivatives obtained for the OU process with those of a discretized generator. . . . .	60

*LIST OF TABLES*

---

2.10	Comparison the difference of eigenpairs and their derivatives obtained from the CIR process with those of a discretized generator. . . . .	60
2.11	Comparison the difference of eigenpairs and their derivatives obtained for the Jacobi process with those of a discretized generator. . . . .	60
2.12	Estimation errors obtained by varying the number of bins in the case of OU process . . . . .	73
2.13	The eigenvalues after truncation of the extreme values the three affine processes are compared to their real value. . . . .	79
2.14	This table gives the error corresponding to different subsampling time after applying a truncation of some values of the process histogram.	80
2.15	Description of WTI 1-minute data set. Note that not everyday is reported and within a day the price for some minutes is missing. In this data set, a regular year appears to have between 310 and 311 days of quotations. . . . .	91
2.16	Descriptive statistics of the WTI 1-minute dataset. The mean of each period shows three possible regimes in the last seven years. Period 1, 2010-2013, is a high oil price regime. Period 2, 2014, is a transition year during which the price of a barrel dropped. It was also a year of high uncertainty as shown by the standard deviation twice higher than during the two other periods. The Period 3, 2015-2016, is a low price regime with the average price per barrel less than half what it was in period 1. . . . .	91
2.17	Least squares fit for reconstruction during the period 2010-2013. The coefficients of a linear fit for the drift are provided for different subsampling times $h$ . A quadratic fit was used for the diffusion and its coefficients are given for different values of $h$ . The mean reverting assumption is validated for the drift whereas a non constant diffusion is identified . . . . .	98
2.18	Half-life and RMSE for period 1 2010-2013 . . . . .	98
2.19	Least Squares fit for reconstruction during the period 2015-2016. The coefficients of a linear fit for the drift are provided for different subsampling times $h$ . A quadratic fit is used for the diffusion and its coefficients are given for different values of $h$ . The random walk assumption for the drift failed to be rejected whereas a non-constant diffusion is identified. . . . .	103

*LIST OF TABLES*

---

2.20 Half-life and RMSE for period 3 from 2015 to 2016 . . . . . 103

3.1 This table gives us the parameters used to test optimal scaling when  
 $\Delta t$  is fixed and  $\Delta x = \frac{1}{M_{c_j}}$ . . . . . 137

# CHAPTER 1

---

## INTRODUCTION

---

### 1.1 Motivation

The increasing availability of financial time series data has motivated the use of various stochastic modeling techniques to understand the dynamics of financial markets. A good example of this growing accessibility is the fact that popular web browsers such as Yahoo and Google now offer, in addition to corporate news headlines, free historical data to model various asset prices, back testing investments strategies, or trigger entry and exit signals in trading systems. In quantitative finance, Stochastic Differential equations (henceforth SDE) has been introduced as a continuous time

model of financial observables; such as assets and commodities prices, rates [8, 9, 17, 28–30, 59, 71]; as well as unobservable quantities such as derivative securities prices, risk, creditworthiness, and others [14, 16, 23, 36, 37, 39, 44, 45, 48]. The identification from time series data of the basic components a SDE, its drift and diffusion terms, is a critical and necessary step toward practical financial applications. In the literature, one could find numerous studies on this topic, [1, 4–7, 18–20, 42, 51, 52, 74, 75]. In statistics, there are two standard ways to estimate coefficients or functions: using a parametric or a non-parametric approach [3, 11, 38]. Both of these statistical methods can be challenging for different reasons. One of them could be related to the sample size available for the estimation. On one hand, parametric methods can be implemented with less data but require some assumptions on the process [11, 38, 40]. On the other hand, non-parametric techniques generally do not impose conditions on the process but work best with a large number of data points [3]. The sample size could influence the choice of the type of estimation used by researchers. Another reason is the sampling frequency of the data. Financial data could be nowadays obtained at different time frequencies such as annually, monthly, weekly, daily, minute, second, milliseconds, and even microseconds. The high frequency data are currently the hardest to find and the more expensive. The existence of such data is sometimes associated to the belief that in financial data analysis the more data we have the better. However, some studies have clearly shown that using more data in an estimation doesn't imply the best goodness of fit [2, 4, 6]. Indeed discrepancies in the statistical behavior between the proposed SDE and the observational data might lead to large errors in the parameter or functional estimation, [2, 78]. Such issue is well known in

the finance literature where high frequency time series are known to be composed of market microstructure noise that if unaccounted can impact the quality of the estimators. In that case, subsampling which means reducing the sampling frequency produces better results, [2, 4–7]. A third reason, that is also the central theme of this dissertation, is the model selection. In other terms, which functional assumptions to make on the drift and diffusion of a proposed SDE to obtain the greatest goodness of fit. When dealing with observational datasets in financial markets, the appropriateness of the model can be measured by its ability to generate profits and losses [25, 26, 55, 69, 73]. For this reason, model selection is critical for market participants and remains a very active research area in academia. Imposing parametric restrictions on the drift and diffusion is for many practitioners the starting point. After a preliminary work on a 'toy' model, more sophisticated diffusion processes are considered. A good example is in derivatives pricing where to model an asset whose value  $V_t$  follows a certain stochastic process, [8, 9, 48, 59] the building block equations are generally Arithmetic or Geometric Brownian Motion, respectively ABM and GBM with the following known form:

$$\text{ABM : } dV_t = \mu dt + \sigma dW_t \quad \text{and} \quad \text{GBM : } dV_t = \mu V_t dt + \sigma V_t dW_t,$$

where  $\mu$  and  $\sigma$  are constant parameters and  $W_t$  is a standard Brownian motion. The literature contains a great amount of references on parametric estimation of SDEs, [1, 3, 5, 7, 11, 38]. A survey paper on different parametric approaches could be found in [72]. But we chose in the following to avoid any assumption on the SDE. For this purpose, we selected non-parametric methods and explore them in a free model environment. Consequently, the first objective of this dissertation is to identify

from discrete data the drift and diffusion *without parametric restrictions*. Different approaches in the literature have been proposed to implement such estimation, [1, 18–20, 33, 42, 51, 52, 74, 75]. We chose to follow the approach developed by [18] that makes use of the eigenspectrum, i.e eigenvalues and eigenfunctions, of the discretized generator of the process to reconstruct the drift and diffusion. The sensitivity of the chosen non-parametric technique to space and time discretization motivates an error analysis. This study uses conditional expectation based estimators and constitutes the second part of this work.

## 1.2 Outline

This thesis is divided into 2 main chapters.

In **Chapter 2**, we review the methodology of reconstruction of the drift and diffusion using spectral information. The procedure is implemented on a SDE whose drift and diffusion have periodic boundary conditions. We propose for this type of SDE an alternative matrix to generate the eigenspectrum. We then propose an extension to stochastic processes commonly used in financial applications, known as affine processes. An application to the modeling of crude oil price is realized using real 1-minute data for a period covering November 2010 to December 2016.

In **Chapter 3**, we propose to study the impact of the time and space discretization on the estimation of the drift and diffusion of a SDE. We restrict our attention to the study of an Ornstein-Uhlenbeck (OU) process and to conditional expectation

## 1.2. *OUTLINE*

---

based estimators. We propose two estimators, one for the drift and one for the diffusion. The chapter begins by assuming that the number of data points is fixed in each interval, or bin, where estimation is performed, and study in this context how the spatial and time step sizes impact the behavior of errors made when computing the conditional expectation of an OU process using our estimators. The following sections relax the assumption of a fixed number of points and then propose a scaling of the errors made when estimating drift and diffusion.

## CHAPTER 2

---

# SPECTRAL RECONSTRUCTION OF DRIFT AND DIFFUSION

---

**Goal of this chapter:** to study the non parametric estimation of the drift and diffusion of SDEs using the spectral information from the discrete time-series.

## 2.1 Motivation

One of the greatest achievements of modern financial theory is undoubtedly the design and widespread use of derivatives securities to mitigate risk. Since the Black Scholes Merton equation [9], countless applications all over the financial spectrum have been done and the interest for this mathematical model has barely bulged among researchers and investors. The recent 2008's financial crisis seems to have shifted the practitioners' focus from derivative pricing to risk management issues (including volatility modeling) but also academics' interest from traditional inference methods such as least square or maximum likelihood to more data driven and machine learning based techniques. Despite such apparent paradigm shift, stochastic analysis continues to provide a rigorous and powerful framework to understand the asset price dynamics. Diffusion type processes described by SDEs, remain the mathematical building blocks commonly used by the financial community to price instruments such as equity, interest rate, credit derivatives or construct standard risk measures such as Value at Risk (VaR) [48]. In order to use this tool, financial modelers have to provide a model specification, or in other terms identification of the parameters or functions of the mathematical model that explains the drivers of the stochastic behavior of a financial asset. Inappropriate specification and use of theoretical models for the purpose of investing have received a considerable amount of attention from regulators and academics since they have been blamed for the last financial crisis [69, 73]. Referred in the industry as Model risk, the hazard of working with a not well suited model can result in major losses, that could lead to bankruptcies such as LTCM in the nineties or Lehman Brothers in 2008, and even jeopardize the entire

economic system. Regulators, worldwide, have set forth guidelines to mitigate model risk even if more needs to be done to understand the complexity of financial systems [26]. One key element of the regulations is the crucial role played by SDEs in domain such as scenario generations. The estimation of these equations is at the center of this thesis.

## 2.2 Background

In the literature, there are some studies that mathematically justify the use of spectral data to estimate the drift and diffusion of SDE, [18–20, 42, 52]. Even though applications on physical data exist (atmospheric data for example [76]) works on financial time series are difficult to find. Our main reference has been the estimation technique introduced by Crommelin and Vanden-Eijnden [18]. In this work, the authors used estimates of eigenvalues and eigenfunctions to reconstruct the drift and diffusion of SDEs with periodic boundary conditions. One key contribution of our work is to explore ways to implement this inference technique on SDE commonly used in financial modeling, e.g., with non periodic drift and diffusion components. In addition to [18], we will also briefly refer subsequent papers by the same authors [19, 20]. We start by discussing the object of the proposed estimation. Its goal is to use the eigenvalues and eigenfunctions of the transition probability matrix generated by the data to infer the drift and diffusion of a Stochastic Differential Equation. This inference is performed via an optimization problem on the eigenvalue equation, or

## 2.2. BACKGROUND

---

eigenequation, solved using standard quadratic programming techniques. The particular interesting feature of this reconstruction is that no functional assumption on the drift and diffusion of the stochastic process is needed. In [18, 19] a Finite Differences (FD) scheme is used to derive the eigenequation. Due to a certain number of limitations of FD such as discretization errors of the derivatives, another method using Finite Element (FE) approximation of the diffusion operator and its adjoint was introduced in [20]. Although the FE approach described below is the most recent study, our work is centered around the FD approach that we attempt to extend to processes with non periodic drift and diffusion with applications in the field of finance. One part of this work is to apply this methodology to financial market data in order to reach a pure data-driven and free-model environment estimation of these equations.

We start laying the background of this method by recalling that the simplest SDE has two major components, the drift denoted  $b(X_t, t)$  that indicates the general trend, followed by the process and the diffusion term,  $a(X_t, t)$ , also referred as the volatility term, that models the degree of randomness of the system. No functional assumptions on  $b(X_t, t)$  and  $a(X_t, t)$  are used in this work. We do not consider jumps in this thesis and will only concentrate on diffusions. We assume time homogeneity, what allows us to omit the time dependency and to focus only on the relationship between the components and the state of the process. The mathematical formulation of our SDE of interest is quite standard [24, 64]:

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t. \quad (2.1)$$

with an initial condition,  $X(0) = x_0$ , and  $W_t$ , the standard Brownian motion. In

order for a SDE to be usable in practical applications,  $b(X_t)$  and  $a(X_t)$ , are calibrated to market data or estimated. A variety of parametric and non-parametric methods have been applied to time series data to find robust estimates of these terms, [1, 10, 42, 51, 52]. Most of these studies on non-parametric estimations rely on the statistical definition of the drift and diffusion in terms of conditional expectations [24, 27, 64]:

$$\begin{aligned} b(X_t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E [X_{t+\Delta t} - x | X_t = x], \\ a(X_t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E [(X_{t+\Delta t} - x)^2 | X_t = x]. \end{aligned} \tag{2.2}$$

These definitions that seem at first glance intuitive and easy to handle, provide undeniably to the modeler a simple way to think of the estimation of the SDE's components. After a careful look, many issues could arise from these definitions, as noted by [18]. For example, what is the appropriate time resolution to model the data, i.e. do we need seconds, minutes, hourly, daily, or more coarse grained data to compute these expectations? Secondly, note that according to the formulas (2.2), the data subsampling needs to be sufficiently small for the above limit to converge, otherwise, we can have a bias due to a finite-time effect. A third type of issue could be the lack of flexibility of the above estimators where prior information on the components is hardly taken into consideration.

### 2.2.1 Some processes used in finance

The choice of functional forms of the drift and diffusion of asset-price processes is more an art than a science. For simplicity, many modelers start by making easy to handle assumptions such as the components are either both constant, or both

## 2.2. BACKGROUND

---

linear. The case where they are both linear is a key assumption of the Black Scholes Merton model [9, 59]. This SDE is known as a Geometric Brownian motion. Another example of an assumed functional form is the square root diffusion of the asset value introduced by Cox, Ingersoll, and Ross in the context of interest rate models in 1985 [17]. In option pricing theory, despite the success of the Black Scholes framework, assumptions underpinning the model have long been questioned. The first one was the assumption of constant volatility of the stock return during the entire life of a derivative. Practitioners and academics both agree that the very dynamic nature of the financial market prevents the volatility from remaining constant and that a varying variance term should enable a more accurate description of financial assets. One way to take into account this issue is to model the volatility as a stochastic process itself. This led to the famous Heston models [44, 45]. For illustration, consider an asset price value modeled by processes,  $V_t$  and  $X_t$ , with constant parameters  $\{\mu, \sigma, \kappa, \theta, \zeta\}$  and Brownian motions  $W_t, W_{1,t}$ , and  $W_{2,t}$ . Here are some of the most commonly used SDE in finance that can be constructed with these variables:

$$\text{Arithmetic Brownian Motion (ABM) : } dV_t = \mu dt + \sigma dW_t,$$

$$\text{Geometric Brownian Motion (GBM) : } dV_t = \mu V_t dt + \sigma V_t dW_t,$$

$$\text{OrnsteinUhlenbeck (OU) : } dV_t = \kappa(\theta - V_t)dt + \sigma dW_t,$$

$$\text{Cox Ingersoll Ross (CIR) : } dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t,$$

It's also possible to use more than one SDE. A second family of models could consider an asset price process,  $X_t$ , and a volatility process,  $V_t$ . Assume two correlated Brownian motion,  $W_{1,t}$  and  $W_{2,t}$  such that  $corr(W_{1,t}, W_{2,t}) = \rho$  with  $|\rho| < 1$ . This type of model is known as a standard stochastic volatility model and examples of

such model are :

- Heston model [44]

$$\begin{aligned}dX_t &= \mu X_t dt + \sqrt{V_t} X_t dW_{1,t}, \\dV_t &= \kappa(\theta - V_t) dt + \zeta \sqrt{V_t} dW_{2,t}.\end{aligned}$$

- 3/2 model [45]

$$\begin{aligned}dX_t &= \mu X_t dt + \sqrt{V_t} X_t dW_{1,t}, \\dV_t &= \kappa(\theta - V_t) dt + \zeta \sqrt[3]{V_t} dW_{2,t}.\end{aligned}$$

- 4/2 model [39]

It is the latest stochastic volatility model introduced by Grasseili [39] that combines the Heston and the 3/2 model into a single tractable specification.

$$\begin{aligned}dX_t &= \mu X_t dt + \left( a\sqrt{V_t} + \frac{b}{\sqrt{V_t}} \right) X_t dW_{1,t} \\dV_t &= \kappa(\theta - V_t) dt + \zeta \sqrt{V_t} dW_{2,t}.\end{aligned}$$

A third category of models that has been introduced as a way to relax the constant volatility assumption in the Black Scholes model takes into account a nonlinear relationship between the risk of a financial asset and its value. This is the Constant Elasticity of Variance model, henceforth CEV [14, 16, 23] that offers a compact expression of a family of volatility models. Introduced in the 1970's to take into account the leverage effect, in other words the intertwined relationship between asset value and risk, the CEV model with mean reversion with an appropriate choice of parameters provide a generalization of the above models. A standard formulation of CEV

## 2.2. BACKGROUND

---

is:

$$dV_t = \kappa(\theta - V_t)dt + \sigma V_t^{\frac{\beta}{2}}dW_t.$$

Notice that an OU specification could be obtained if  $\beta = 0$ , when  $\beta = 1$  we have the CIR model with mean reversion [44], and  $\beta = 3$  gives us a mean reverting 3/2 model, [45]. The GBM can be obtained with additional assumptions on the drift parameters whereas the ABM could be derived from CEV by choosing the right parameters and also removing the state dependency of the components of the SDE. The transition probability distribution of some of the above models can be obtained using the CEV formulation above:

- Case of  $\beta = 2$

The OU process,  $\theta = 0$  and  $\beta = 2$  has a log-normal transition probability density so we define  $y = \log(V_{t+\Delta t})$  and  $x = \log(V_t)$ :

$$Prob(y, t + \Delta t | x, t) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} e^{-\frac{1}{2}\left(\frac{y - (x - \kappa - \frac{\sigma^2}{2})}{\sigma^2\Delta t}\right)^2}$$

- Case of  $\beta \neq 2$

When  $\beta \neq 2$ , the transition probability distribution generally follows a non central Chi squared. In this case, no log variables are needed. Thus define  $y = V_{t+\Delta t}$  and  $x = V_t$ . In addition to these, denote  $\mathcal{I}_q(\cdot)$  a modified Bessel function of order  $q$  and define the following parameters:

$$k = \frac{-2\kappa}{\sigma^2(2 - \beta)(e^{(2-\beta)(-\kappa\Delta t)} - 1)},$$

$$x = kx^{2-\beta}e^{(2-\beta)(-\kappa\Delta t)},$$

$$z = ky^{2-\beta}.$$

1. For  $\beta < 2$  [14]

A special case is the Heston model. The transition probability density has the form:

$$Prob(y, t + \Delta t | x, t) = (2 - \beta) k^{\frac{1}{2-\beta}} (xz^{1-2\beta})^{\frac{1}{4-2\beta}} e^{-x-z} \mathcal{I}_{\frac{1}{2-\beta}} [2\sqrt{xz}].$$

2. For  $\beta > 2$  [23]

A popular special case is the 3/2 stochastic volatility model and its transition probability density:

$$Prob(y, t + \Delta t | x, t) = (\beta - 2) k^{\frac{1}{2-\beta}} (xz^{1-2\beta})^{\frac{1}{4-2\beta}} e^{-x-z} \mathcal{I}_{\frac{1}{\beta-2}} [2\sqrt{xz}].$$

### 2.2.2 Affine processes

Our objective is to test spectral information extracted from time series data to identify the drift and diffusion of SDEs such as the ones presented above. For this purpose, it appears natural to start by processes with known spectral characteristics and useful applications in financial modeling. Good processes to start with are the affine ones. Indeed Affine Diffusion processes have well-known spectral characteristics, transition probability densities, and are the starting points of most financial models, [35]. In the following, we will focus on three of such processes: The Ornstein-Uhlenbeck (OU) process, the Cox-Ingersoll-Ross (CIR) process, also referred as the square root process, and the Jacobi Process (JAC). The characteristics of these processes are now presented. More details are found in [35]. Assume, we have a stochastic process,  $X_t$ , and parameters,  $\{\kappa, \theta, \gamma, a, b\}$ , and a standard Brownian motion,  $W_t$ . We define:

1. **The OU process**

The SDE for the OU process is:

$$dX_t = \kappa(\theta - X_t)dt + \gamma dW_t. \quad (2.3)$$

The eigenvalues are given by  $\lambda_n = -\kappa n$ .

The eigenfunctions are the Hermite polynomials given by:

$$He_n(X_t) = (n!)^{\frac{1}{2}} \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^m \frac{1}{m! 2^m (n-2m)!} \left(\frac{-2b}{c_0}\right)^{\frac{n-2m}{2}} (X_t - \beta)^{n-2m}. \quad (2.4)$$

Note that the eigenfunctions are standardized with respect to Gaussian distribution  $\mathcal{N}(\theta, -\frac{\gamma^2}{2\kappa})$ . The recurrence relation for the Hermite polynomials is given by:

$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0. \quad (2.5)$$

The stationary distribution is

$$f(x) = \sqrt{\frac{\kappa}{\pi\gamma^2}} \exp\left(\frac{-\kappa(x - \theta)^2}{\gamma^2}\right). \quad (2.6)$$

2. **The CIR process**

The standard SDE for the CIR process is:

$$dX_t = \kappa(\theta - X_t)dt + \gamma\sqrt{X_t}dW_t. \quad (2.7)$$

The eigenvalues are given by  $\lambda_n = -\kappa n$ .

The eigenfunctions are given by the Generalized Laguerre polynomials:

$$L_n^\alpha(-2\kappa X_t) = \binom{n+\alpha}{n}^{-\frac{1}{2}} \sum_{m=0}^n \left[ \binom{n+\alpha}{n-m} (2\kappa)^{\frac{(X_t)^m}{m!}} \right], \quad (2.8)$$

## 2.2. BACKGROUND

---

where  $\alpha = -2\kappa\theta - 1$ . The eigenfunctions are standardized with respect to the marginal distribution of the process which is an affine transformation of the gamma distribution  $\Gamma(-2\kappa\theta)$ . The recurrence relation for the Laguerre polynomials is given by:

$$(n+1)L_{n+1}^\alpha(x) - (2n + \alpha + 1 - x)L_n^\alpha(x) + (n + \alpha)L_{n-1}^\alpha(x) = 0. \quad (2.9)$$

The stationary distribution of the CIR process is

$$f(x) = \frac{(2\kappa)^{2\kappa\theta}}{\Gamma(2\kappa\theta)} \exp(-2\kappa\theta)x^{2\kappa\theta-1} \mathbb{1}_{(0,+\infty)}(x). \quad (2.10)$$

### 3. The JAC process

This process is defined in a bounded interval  $[a, b]$  with  $a < b$ . The standard SDE for the Jacobi process is:

$$dX_t = \kappa(\theta - X_t)dt + \gamma\sqrt{(X_t - a)(b - X_t)}dW_t. \quad (2.11)$$

The eigenvalues are given by  $\lambda_n = -\kappa n - \frac{n(n-1)}{2}$ .

The eigenfunctions are given by the Jacobi polynomials:

$$P_n^{\alpha,\beta}(X_t) = \left[ \frac{\Gamma(\alpha + n + 1)(2n + \alpha + \beta + 1)\Gamma(\alpha + 1)\Gamma(\beta + 1)}{n!\Gamma(\alpha + \beta + n + 1)\Gamma(\alpha + \beta + 2)\Gamma(\beta + n + 1)} \right]^{\frac{1}{2}} \times \sum_{m=0}^n \left[ \binom{n}{m} \frac{\Gamma(\alpha + \beta + n + m + 1)}{\Gamma(\alpha + m + 1)} \left( \frac{X_t - b}{b - a} \right)^m \right], \quad (2.12)$$

where  $\alpha = 2\theta \left( \frac{b-\theta}{b-a} \right)$  and  $\beta = 2\theta \left( \frac{\theta-a}{b-a} \right) - 1$ . The eigenfunctions are standardized with respect to the marginal distribution of the process which is an affine transformation of the Beta distribution,  $Beta(\beta + 1, \alpha + 1)$ . The recurrence

relation for the Jacobi polynomials can be expressed as:

$$\begin{aligned}
 & 2n(n + \alpha + \beta)(2n + \alpha + \beta - 2)P_n^{\alpha, \beta}(x) \\
 &= (2n + \alpha + \beta - 1)\{(2n + \alpha + \beta)(2n + \alpha + \beta - 2)x + \alpha^2 - \beta^2\}P_{n-1}^{\alpha, \beta}(x) \\
 &\quad - 2(n + \alpha - 1)(n + \beta - 1)(2n + \alpha + \beta)P_{n-2}^{\alpha, \beta}(x).
 \end{aligned} \tag{2.13}$$

The Jacobi process has a stationary distribution:

$$f(x) = \frac{(x - a)^{2\kappa \frac{\beta - a}{b - 1} - 1} (b - x)^{2\kappa \frac{b - \beta}{b - 1} - 1}}{(b - a)^{2\kappa - 1} \text{Beta}(2\kappa \frac{\beta - a}{b - a}, 2\kappa \frac{b - \beta}{b - a})} \mathbb{1}_{(a, b)}(x), \tag{2.14}$$

where  $\mathbb{1}_{(a, b)}(x)$  is an indicator function.

## 2.3 Important concepts in stochastic processes

Before presenting an alternative inference methodology to the ones commonly used in the econometric literature, key concepts in stochastic process are recalled in this section. Consider  $\{X_t, t \geq 0\}$  an ergodic diffusion process. The dynamics of  $X_t$  is described by the equation (2.1). One key element used throughout this thesis is the duality between Stochastic Differential Equations(SDE) and Partial Differential Equations (henceforth PDE). Indeed, if a SDE provides a powerful probabilistic description of the process, a PDE's perspective offers a more physical interpretation in terms of diffusion. In this latter case, (differential) Chapman- Kolmogorov equations provides a time evolution of the process in its state space. The change, backward or forward in time, of the transition probability density, denoted  $P(x, t | y, t')$  with  $t > t'$  for given states,  $x$  and  $y$ , generates the following differential equations (see [27]):

- Backward differential Chapman-Kolmogorov equation

$$\frac{\partial P(x, t | y, t')}{\partial t'} = -b(y) \frac{\partial P(x, t | y, t')}{\partial y} - \frac{1}{2} a(y) \frac{\partial^2 P(x, t | y, t')}{\partial y^2}. \quad (2.15)$$

- Forward differential Chapman-Kolmogorov equation

$$\frac{\partial P(x, t | y, t')}{\partial t} = -\frac{\partial}{\partial x} [b(x)P(x, t | y, t')] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [a(x)P(x, t | y, t')]. \quad (2.16)$$

The Backward Chapman Kolmogorov equation is represented by a infinitesimal generator operator,  $L$ , and the Forward Chapman Kolmogorov equation by its adjoint operator,  $L^*$ . They both are defined as followed:

$$\begin{aligned} Lf &= b(x, t) \frac{\partial f(x, t)}{\partial x} + \frac{1}{2} a(x, t) \frac{\partial^2 f(x, t)}{\partial x^2}, \\ L^* f &= -\frac{\partial}{\partial x} [b(x, t)f(x, t)] + \frac{1}{2} a(x, t) \frac{\partial^2}{\partial x^2} [a(x, t)f(x, t)]. \end{aligned} \quad (2.17)$$

Assuming a sampling process in equilibrium allows the removal of the time dependency of the infinitesimal generator,  $L$ , and similarly for the adjoint,  $L^*$ , that can now be expressed as follows:

$$\begin{aligned} Lf &= b(x) \frac{\partial f(x)}{\partial x} + \frac{1}{2} a(x) \frac{\partial^2 f(x)}{\partial x^2}, \\ L^* f &= -\frac{\partial}{\partial x} [b(x)f(x)] + \frac{1}{2} a(x) \frac{\partial^2}{\partial x^2} [a(x)f(x)]. \end{aligned} \quad (2.18)$$

Assume  $L$  has a discrete spectrum,  $\{\psi_k(x), \phi_k(x), \lambda_k\}$  for  $k \in \mathbb{N}$  where  $\{\psi_k(x), \phi_k(x), \lambda_k\}$  is a set of left, right eigenfunctions and eigenvalues, respectively such that

$$L\phi_k(x) = \lambda_k \phi_k(x), \quad (2.19)$$

$$L^*\psi_k(x) = \bar{\lambda}_k \psi_k(x). \quad (2.20)$$

Note that in the case of a stationary process, the leading eigenvalue,  $\lambda_1$ , would be equal to zero, while all others would be strictly negative. Furthermore, the left

eigenfunction,  $\psi_1$ , corresponding to the leading eigenvalue of the adjoint operator would be the stationary distribution of the process.

**Connection between transition probability, infinitesimal generator and their spectrum**

Consider the diffusion process,  $X_t$ , introduced in the previous section, with initial condition,  $X_0 = x$ , and a suitable function,  $f(x)$  and  $t \geq 0$ . Define the following operator  $P_t$  by

$$(P_t f)(x) = E[f(X_t)|X_0 = x]. \tag{2.21}$$

The infinitesimal generator  $L$  associated to  $P_t$  is [27]

$$L f(x) = \lim_{t \downarrow 0} \frac{(P_t f)(x) - f(x)}{t}. \tag{2.22}$$

Using the above definition and semi-group theory, one can show that for a given time,  $t$ , the operators,  $P_t$  and  $L$ , are related by the equation.

$$P_t = e^{tL}. \tag{2.23}$$

The same relationship is satisfied by the adjoint operators,  $P_t^*$  and  $L^*$ . This relationship between  $P_t$  and  $L$  implies a direct relationship between the underlying eigenstructures of both operators. More precisely, note that both operators share the same left and right eigenfunctions denoted  $\psi_k, \phi_k$ , respectively and that the

eigenvalues of  $P_t$  are just an exponential transformation of those of  $L$ : for  $k \in \mathbb{N}$ :

$$P_t \phi_k(x) = \Lambda_k \phi_k(x), \quad (2.24)$$

$$P_t^* \psi_k(x) = \bar{\Lambda}_k \psi_k(x), \quad (2.25)$$

$$\lambda_k = \frac{1}{t} \log(\Lambda_k). \quad (2.26)$$

where  $\{\Lambda_k\}$  are eigenvalues of  $P_t$  and  $\{\lambda_k\}$  those of  $L$ . As mentioned in [18], these relationships are valid regardless of the value of  $t$ . By using a finite time value, one is not making any time discretization error since by the above relations time has no effect on the estimation of the infinitesimal generator.

## 2.4 Spectral Reconstruction of one SDE

Here, we summarized briefly [18]. This paper tries to answer the following question: given some data  $\{U_n, n = 1, \dots, N\}$  can we recover the components  $b(x)$  and  $a(x)$  of the data generating SDE? Different approaches to this problem already exist in the literature but one key contribution of [18] has been to develop an efficient method to find the "closest", measured by the spectrum, drift, and diffusion of the underlying infinitesimal generator of the data. Different stages of the reconstruction are discussed below.

### 2.4.1 Construction of the transition probability matrix $P$

A discrete time Markov chain obtained by discretization of the state space of the process could be constructed as a finite state representation of  $P_t$ . We call it  $P_{m,m'}$ .

## 2.4. SPECTRAL RECONSTRUCTION OF ONE SDE

---

It is a  $M \times M$  stochastic matrix where  $M$  is the number of states,  $m, m'$  are integers,  $\{1, \dots, M\}$ . All the entries are between zero and 1 and the rows sum to one. The construction of this stochastic matrix straight from the data doesn't guarantee that the relationship (2.23) holds. In fact, this issue is related to an open question on the theory of Markov chains called the embedding problem [41]. Not every construction of a discrete time Markov chain such as  $P_{m,m'}$  is associated with an infinitesimal generator. If such generator could be found,  $P_{m,m'}$  is said to be embeddable. The authors in [18] get around the embedding problem by finding the drift and diffusion of not the exact generator but those of the generator whose spectrum is the closest to the eigenvalues and eigenfunctions generated from the data. The procedure omits the embedding issue and starts with the construction of  $P_{m,m'}$  and the calculation of the eigenpairs.  $P_{m,m'}$  is the  $(m, m')$  transition probability defined by:

$$P_{m,m'} = \text{Prob} (X_{(j+1)\Delta t} \in \text{state } m' | X_{j\Delta t} \in \text{state } m) .$$

where  $j \in \mathbb{N}$ ,  $\Delta t$  is time step, and  $m, m' \in \{1, \dots, M\}$ . States are defined by discretizing the state space with a step size  $\Delta x$ , i.e.  $\{X_{j\Delta t} \in \text{state } m\}$  is equivalent to  $X_{j\Delta t} \in \left[ m - \frac{\Delta x}{2}, m + \frac{\Delta x}{2} \right]$  The standard approach to construct these probabilities to use the maximum likelihood estimator (henceforth MLE ) of this Markov chain. The likelihood function can be expressed as:

$$\mathcal{L}(P_{m,m'}) = \prod_{m,m' \in \mathcal{M}} P_{m,m'}^{\delta_{m,m'}} .$$

where  $\mathcal{M} = \{1, \dots, M\}$ ,  $\delta_{m,m'}$  is equal to 1 when there is a transition between  $m$  and  $m'$  and zero otherwise. A MLE estimator of the transition probabilities is obtained by maximizing the log likelihood function,  $\mathfrak{L}(P_{m,m'})$ , subject to the constraint that

all the rows of the matrix  $P_{m,m'}$  sum to one. Using Lagrange multipliers,  $(\mu_m)_{m \in K}$ , the estimator is obtained by maximizing the following objective function:

$$\mathfrak{L}(P_{m,m'}) - \sum_{m' \in \mathcal{M}} \mu_m \left( \sum_{m' \in \mathcal{M}} P_{m,m'} - 1 \right).$$

The solution of the above problem provides the following formula for each transition probability:

$$P_{m,m'} = \frac{\delta_{m,m'}}{\sum_{m' \in \mathcal{M}} \delta_{m,m'}}. \quad (2.27)$$

Following authors in [18], we define a spectral estimation using the empirical counterpart of the above formula that can be written as follows:

$$P_{m,m'} = \frac{\sum_{j=1}^{N_t} \mathbb{1}(X_{j\Delta t} = m) \mathbb{1}(X_{(j+1)\Delta t} = m')}{\sum_{j=1}^{N_t} \mathbb{1}(X_{j\Delta t} = m)}. \quad (2.28)$$

### 2.4.2 Finite Difference (FD) approach

As can be seen from equation (2.18), the infinitesimal generator is an operator which depends on the drift and diffusion,  $b(x)$  and  $a(x)$ , so the goal of the method is to find the drift and diffusion that would generate a spectrum which is as close as possible to the one generated by the data. Practically, this reduces to an optimization problem whose objective function minimizes the error between the data-based spectrum and the one from the infinitesimal generator,  $L$ , from its adjoint and from the eigenequation of the diffusion process. Let  $K \in \mathbb{N}$  be the highest mode that is possible to reliably estimate. The objective function used has the following form:

$$E(b, a) = \sum_{k=1}^K \alpha_k \|L^* \psi_k - \lambda_k \psi_k\|^2 + \beta_k \|L \phi_k - \lambda_k \phi_k\|^2 + \gamma_k |\langle \psi_k, L \phi_k \rangle - \lambda_k|^2, \quad (2.29)$$

where  $\alpha_k = \frac{1}{|\lambda_k \psi_k|^2}$ ,  $\beta_k = \frac{1}{|\lambda_k \phi_k|^2}$ , and  $\gamma_k = \frac{1}{|\lambda_k|^2}$ . Note that the discrete counterpart of the above error function is obtained using a finite difference approximation of the infinitesimal generator where  $D_i$  and  $D_i^2$  stand for the first and second finite differences operators, respectively, defined as follows:

$$D_i f = \frac{f_{i+1} - f_{i-1}}{2\Delta x}, \quad D_i^2 f = \frac{f_{i+1} + f_{i-1} - 2f_i}{\Delta x^2}. \quad (2.30)$$

Using these operators, the discrete analog of (2.29):

$$\begin{aligned} \tilde{E}(b, a) = & \sum_{k=1}^K \alpha_k \sum_{i=1}^N \left| -D_i(b\tilde{\psi}_k) + \frac{1}{2}D_i^2(a\tilde{\psi}_k) - \tilde{\lambda}_k \tilde{\psi}_{k,i} \right|^2 \\ & + \sum_{k=1}^K \beta_k \sum_{i=1}^N \left| b_i D_i(\tilde{\phi}_k) + \frac{1}{2}a_i D_i^2(\tilde{\phi}_k) - \tilde{\lambda}_k \tilde{\phi}_{k,i} \right|^2 \\ & + \sum_{k=1}^K \gamma_k \left| \sum_{i=1}^N \tilde{\psi}_{k,i} \left( b_i D_i(\tilde{\phi}_k) + \frac{1}{2}a_i D_i^2(\tilde{\phi}_k) \right) - \tilde{\lambda}_k \right|^2 \end{aligned} \quad (2.31)$$

This expression can be reduced to a quadratic optimization problem of the form

$$\tilde{E}(b, a) = \langle v, H v \rangle + \langle v, F \rangle + \tilde{E}_0 \quad (2.32)$$

where  $\tilde{E}_0$  is a constant term,  $F$  is vector,  $H$  is a positive definite symmetric matrix and  $v$  is the  $2N$  dimensional vector containing the discretized drift and diffusion term  $v = (b_1, b_2, \dots, b_N, a_1, a_2, \dots, a_N)$ .

## 2.5 Periodic boundary conditions and reconstruction

The first numerical part of this work has been to reproduce, using MATLAB, the result from [18] for 1-dimensional SDE with periodic drift and diffusion. The goal is

to generate synthetic data from a known SDE, then assume that we ignore the data-generating process and apply the reconstruction procedure to identify the potential SDE that have generated the data. The true drift and diffusion being known, we can compare them with the reconstructed ones. Consider a bounded domain  $\Omega = [-\pi, \pi]$  and a SDE,  $dx = b(x)dt + a(x)dW$ , with drift and diffusion terms defined as follows:

$$b(x) = 1 + \cos(x), \quad a(x) = 1 + \frac{1}{2} \sin(x). \quad (2.33)$$

In the original work [18],  $10^6$  points were generated using a Euler scheme with a time-step of  $10^{-4}$ , a subsampling time of 0.1 between consecutive points and a number of bins and of discretization points both equal to 60. The parameters used are  $T = 10^6$ ,  $dt = 10^4$ ,  $h = 0.1$ , and  $M = N = 60$ . Only the first three eigenvalues and eigenfunctions were used, so  $K$ , the number of eigenpairs in the reconstruction is such that  $K = 3$ . We used a slightly different approach. Instead of generating only one time series, we generated 200 paths by numerical integration of the process. The idea is to avoid any randomness in the conclusion of this estimation. The trajectories are generated with two nested loops. The first one computes  $\frac{T}{h}$  points where the solution would be evaluated, and the second loop inside the first sub-samples by integrating  $\frac{h}{dt}$  points of the SDE between each time step  $h$ . This subsampling is very computationally expensive, i.e. more than  $\frac{T}{h}$  operations for one trajectory. From these 200 paths, we constructed 200 transition probability matrices and averaged their entries to have only one stochastic matrix. The matrix rows sum to 1 and we recover eigenvalues and eigenfunctions close to the one in [18]. The eigenvalues are shown in the Table 2.1 below. Through the reconstruction process, the computed eigenfunctions exhibit oscillations mostly due to a finite sample size. The authors

	Our work	Eigenvalues in [18]
$\lambda_1$	$-2.2204 \text{ e-15} + 0i$	0
$\lambda_2$	$-0.6408 + 0.8988i$	$-0.6508 + 0.9086i$
$\lambda_3$	$-0.6408 - 0.8988i$	$-0.6508 - 0.9086i$

**Table 2.1:** Comparison of the eigenvalues obtained by simulating ensemble of trajectories with the one in [18]

in [18] Fourier filtered the computed eigenfunctions in order to reduce the small scale errors that distort the final reconstructed functions. We use a slightly different Fourier filter in our reconstruction process and also try a different approach based on the idea of a sliding, Parzen, windows computation .

### 2.5.1 Periodic boundary conditions with Fourier filter

In order to have an idea of the magnitude of the errors made on eigenfunctions, we discretized the infinitesimal generator associated to (2.33) and constructed a finite matrix  $A$  representing this operator. Then we used the eigenpairs of  $A$ ,  $\{\phi^A, \psi^A\}$  as reference of the true one. We compute the  $L^2$  norm of the difference between the eigenfunctions from the Markov chain  $\{\phi, \psi\}$  and the eigenfunctions from the generator  $A$ . The norms before applying a Fourier filter are given in Table 2.1. The numerical second derivatives appear to be the major source of errors. In order to smooth out the derivatives of the eigenfunctions, the authors in [18] used a Fourier filter whose goal was to project each eigenfunction into the phase space, identify the leading wave-numbers and discarding the higher ones. They arbitrarily chose to discard all eigenmodes with wavelengths greater than 6. We attempted the same

2.5. PERIODIC BOUNDARY CONDITIONS AND RECONSTRUCTION

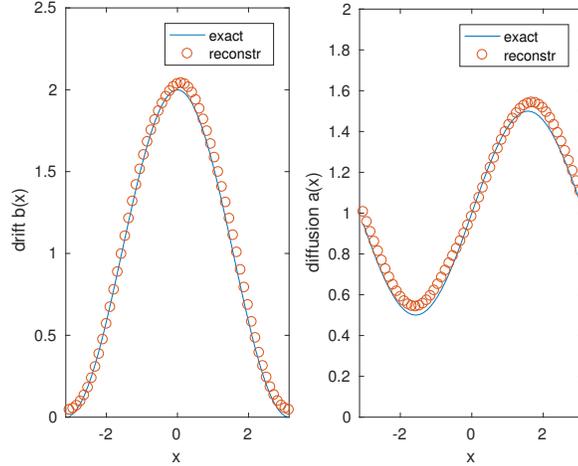
---

Before Fourier filtering						
$K$	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	0.00423	0.0370	0.01470	0.12198	0.4507	4.6005
2	0.11921	0.2771	0.16117	0.32288	1.6991	4.6687
3	0.11921	0.2771	0.16117	0.32288	1.6991	4.6687
After Fourier filtering						
$K$	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	6.013e-14	0.03246	8.05e-14	0.04305	4.58e-13	0.09866
2	0.11893	0.27616	0.15535	0.28786	0.33009	0.35506
3	0.11893	0.27622	0.15517	0.28919	0.32499	0.37616

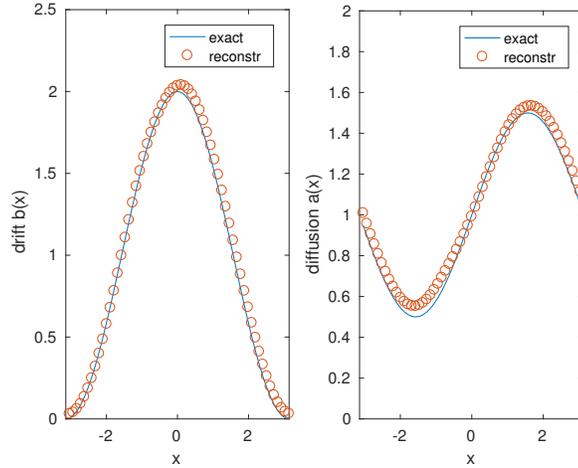
**Table 2.2:** Comparison before and after Fourier filtering of the difference of eigenpairs and their derivatives with those of the discretized generator

approach but found it unreliable in our case. We instead used the fast Fourier transform command *fft* of MATLAB, identify and set to zero all the Fourier coefficients less than  $10^{-3}$ , and apply the inverse transform using the command *ifft*. After using this Fourier filter, we recompute the 2-norm of the difference of these eigenfunctions and the reference ones, see Table 2.1. As expected the Fourier filter significantly smoothed out the second derivatives that show less discrepancy with the reference ones. We also carry over the reconstruction whose results are presented in Figures 2.1, 2.2, 2.3, and 2.4.

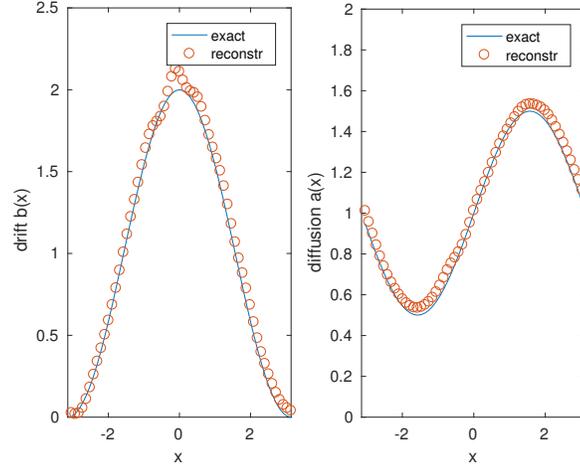
**Conclusion of this subsection:** From an ensemble of trajectories we have constructed a Markov chain whose spectrum has been used to reconstruct the drift and diffusion of the data-generating process. In the procedure, we found that the leading eigenvalues and eigenfunctions are well approximated and that the major source of errors appears to be the numerical differentiation of the eigenfunctions using finite differences.



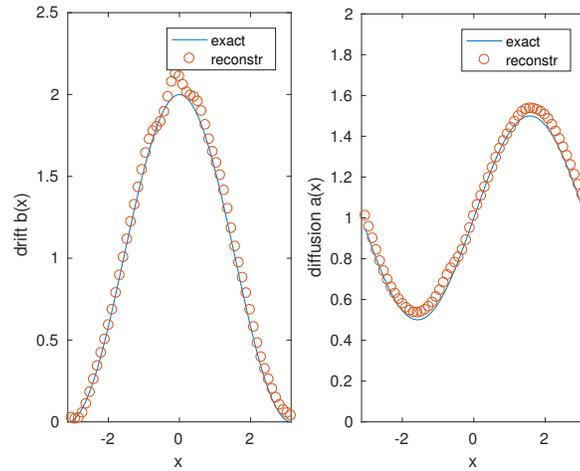
**Figure 2.1:** The results of an estimation using only the generator not the adjoint and a Fourier filtering of the eigenfunctions are given in this plot. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve.



**Figure 2.2:** The results of an estimation using only the adjoint operator not the generator and a Fourier filtering of the eigenfunctions are given in this figure. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve.



**Figure 2.3:** The results of an estimation using the generator, the adjoint operator and a Fourier filtering of the eigenfunctions are given in these plots. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve.



**Figure 2.4:** The results of an estimation using the entire expression  $E(b, a)$  in (2.29) and a Fourier filtering of the eigenfunctions are given in these plots. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left plot, the reconstructed drift is shown in circle while their true value is given by the line curve. On the right plot, the estimated diffusion is shown in circle and the corresponding true values are given by the curve.

### 2.5.2 Periodic boundary conditions with Parzen windows

The filtering approach in [18] which consists in discarding all the Fourier modes greater than a certain value appears rather excessive and some relevant information might be lost in the process. We tried a different approach based on modifying entries of the transition probability matrix to smooth out the eigenvalues. We refer to it as a Parzen Windows smoothing. The key idea is to generalize the definition (2.28) of the entries transition probabilities  $P_{m,m'}$ . This is done by replacing the indicator function that accounts for a transition between two states  $m$  and  $m'$  by a more general kernel function. We modified the recording of transitions by adding one value at not only the relevant entry  $(m, m')$  where transition occurs but also at the immediate neighboring entries for example  $(m - m', m')$  and  $(m + 1, m')$ . The matrix,  $P_{m,m'}$ , has therefore more nonzero entries than in the Fourier Filter case. More formally, we generalize the formula (2.28) as follows. Denote  $\kappa(m, m')$  a function from which a transition probabilities can be constructed as:

$$P_{m,m'} = \frac{\kappa(m, m')}{\sum_{m' \in \mathcal{M}} \kappa(m, m')}, \quad (2.34)$$

where  $\kappa_{m,m'}$  stores the transitions. We tried a triangular (or tent) function which consists on giving a value less than 1 to the entry where transition occurs and also a nonzero value to neighboring entries such that the total amount input is 1. This scheme takes into account cases where a transition to  $m'$  from  $m$  is observed but the probability of a transition to neighbors of  $m'$  should be non-zero. We refer to this scheme as  $parzen_w(hc, sl)$  where  $hc$  is the number of entries to input, including neighbors and target entry, and  $sl$  is a positive constant that gives the relationship

between the values in one entry and its neighbors. For example, if we have a slope of 2 it would mean that the entries  $(m, m')$  in the matrix is assigned a certain value and the 2 first neighboring entries, 1 transition apart from the left and right, have a value 2 times smaller than that of  $(m, m')$ , the entries 2 transitions apart are 4 times smaller, etc. The sum of all these values are set to 1. The idea of the parameter  $hc$  is to identify a certain number of entries to be assigned. For example, let  $hc = 2$  means that we consider  $(m, m')$  and 2 entries  $(m, m' \pm 1)$ , when  $hc = 4$  involves 4 entries around  $(m, m')$ , i.e.  $(m, m' \pm 1)$  and,  $(m, m' \pm 2)$ .

Different number of neighboring entries have been used. The values used for each are as follows:

1. For  $parzen_w(3, 2)$ , the entries when a transition  $m, m'$  occurs are:

$$((m-1, m'), (m, m'), (m+1, m')) = (0.25, 0.5, 0.25).$$

2. For  $parzen_w(5, 2)$ , the entries when a transition  $m, m'$  occurs are:

$$((m-2, m'), (m-1, m'), (m, m'), (m+1, m'), (m+2, m')) = (0.1, 0.2, 0.4, 0.2, 0.1).$$

3. For  $parzen_w(7, 2)$ , the entries when a transition  $m, m'$  occurs are:

$$((m-3, m'), (m-2, m'), (m-1, m'), (m, m'), (m+1, m'), (m+2, m'), (m+3, m')) \\ = (0.045455, 0.090909, 0.18182, 0.36364, 0.18182, 0.090909, 0.045455).$$

Recall that the eigenvalues obtained from the Markov chain based on MLE estimation in [18], were  $(\lambda_1, \lambda_2, \lambda_3) = (0, -0.6508 + 0.9086i, -0.6508 - 0.9086i)$ . Using  $parzen_w(hc, sl)$ , we obtain larger eigenvalues as shown in Table 2.3. These values seems to increase with the number of neighbors  $hc$ . Like in the Fourier filtering case,

## 2.5. PERIODIC BOUNDARY CONDITIONS AND RECONSTRUCTION

---

we also calculated the 2-norm of the difference between the eigenfunctions, their derivatives and those of discretized generator  $A$ . Comparison is presented in the Tables 2.4, 2.5, and 2.6. The best filtered numerical derivatives with  $parzen_w(hc, sl)$  appear when  $hc = 5$ . We implement the reconstruction by generating with a Euler

	$parzen_w(3, 2)$	$parzen_w(5, 2)$	$parzen_w(7, 2)$
$\lambda_1$	$3.1086\text{e-}14 + 0i$	$2.7311\text{e-}13 + 0i$	$-3.0864\text{e-}13 + 0i$
$\lambda_2$	$-0.6745 + 0.9064i$	$-0.7007 + 0.9152i$	$-0.7329 + 0.9218i$
$\lambda_3$	$-0.6745 - 0.9064i$	$-0.7007 - 0.9152i$	$-0.7329 - 0.9218i$

**Table 2.3:** Comparison of the eigenvalues using a transition probability computed with  $parzen_w$  with those [18]

scheme of 200 trajectories of  $T = 10^6$  sample points, a time-step of  $dt = 10^{-4}$ , a sub-sampling time of  $h = 0.1$  and a number of bins and discretization points to  $M = 60$ . Results of this Parzen windows-reconstruction approach are presented in the Figures 2.5 to 2.16. Note that the reconstructed diffusion component via this technique is higher than the true one. This could be explained by the fact that in each row of the transition probability matrix, a window of values that sum to one is used to record a transition instead of a single entry. Each entry with one nonzero transition will impact the neighboring entries and the diffusion spreads out a little more with this window reaching more states than it would have without Parzen windows. This results in a greater inferred diffusion coefficient at each spatial point as shown below.

2.5. PERIODIC BOUNDARY CONDITIONS AND RECONSTRUCTION

---

<i>parzen<sub>w</sub>(3, 2)</i>						
<i>K</i>	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	6.1e-14	0.0183	8.35e-14	0.02111	4.77e-13	0.1228
2	1.9972	0.13613	2.2132	0.1437	3.6638	0.2294
3	1.9972	0.13613	2.2132	0.1437	3.6638	0.2294

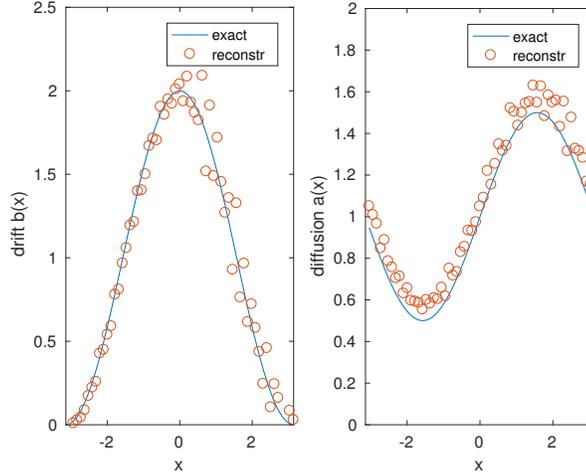
**Table 2.4:** Comparison of the difference of eigenpairs and their derivatives obtained using *parzen<sub>w</sub>(3, 2)* with those of a discretized generator.

<i>parzen<sub>w</sub>(5, 2)</i>						
<i>K</i>	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	2.0000	1.9999	2.88e-13	0.8809	8.57e-13	1.0366
2	0.0996	0.0211	0.1805	0.0411	0.5027	0.2320
3	0.0996	0.0211	0.1805	0.0411	0.5027	0.2320

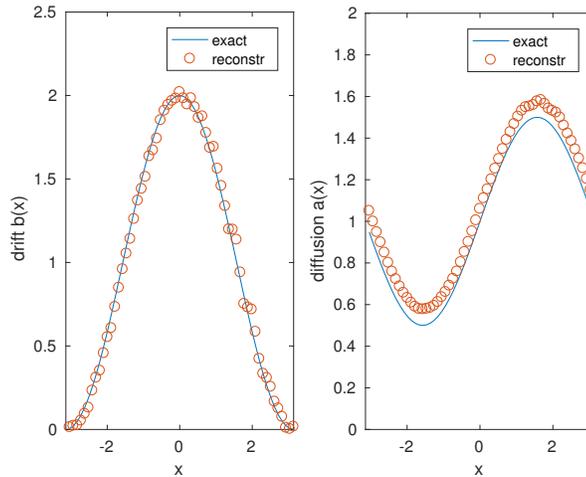
**Table 2.5:** Comparison of the difference of eigenpairs and their derivatives obtained using *parzen<sub>w</sub>(5, 2)* with those of a discretized generator.

<i>parzen<sub>w</sub>(7, 2)</i>						
<i>K</i>	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	2.0000	1.9998	4.84e-13	0.8712	1.09e-12	1.0230
2	0.2054	0.0341	0.3051	0.0658	0.7534	0.2288
3	0.2054	0.0341	0.3051	0.0658	0.7534	0.2288

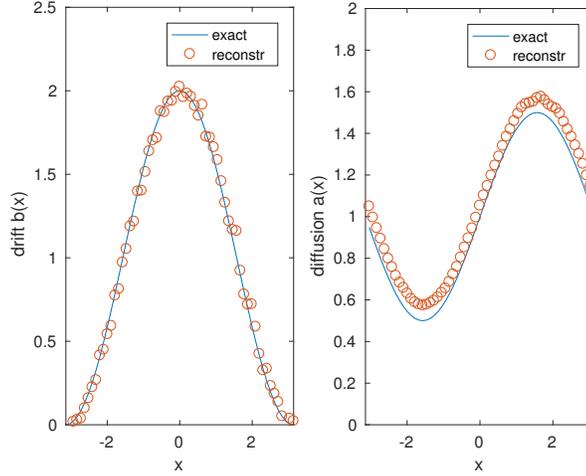
**Table 2.6:** Comparison of the difference of eigenpairs and their derivatives obtained using *parzen<sub>w</sub>(7, 2)* with those of a discretized generator.



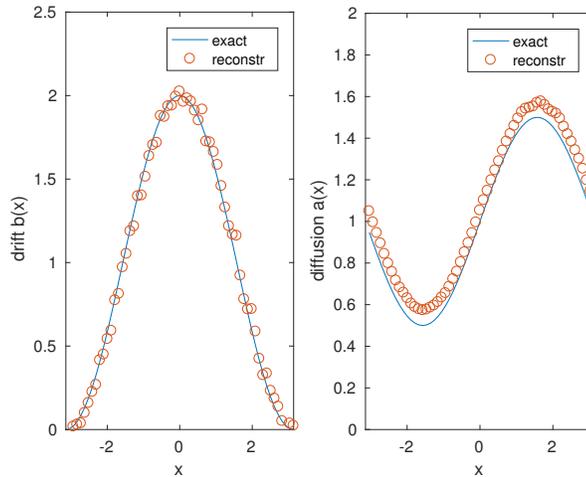
**Figure 2.5:** Reconstruction of the drift and diffusion using only the generator and  $parzen_w(3,2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift.



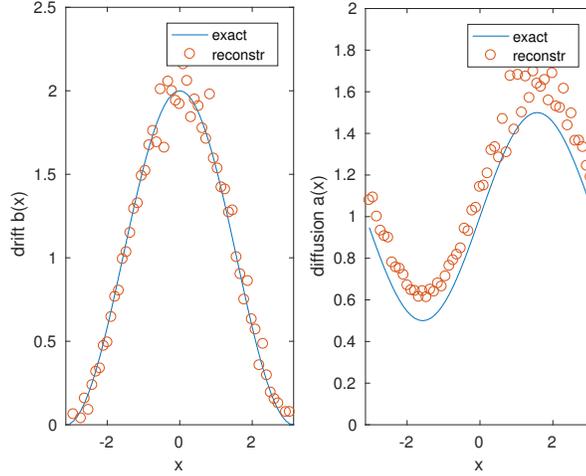
**Figure 2.6:** Reconstruction of the drift and diffusion using only the adjoint and  $parzen_w(3,2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example  $x = 2$  more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some term.



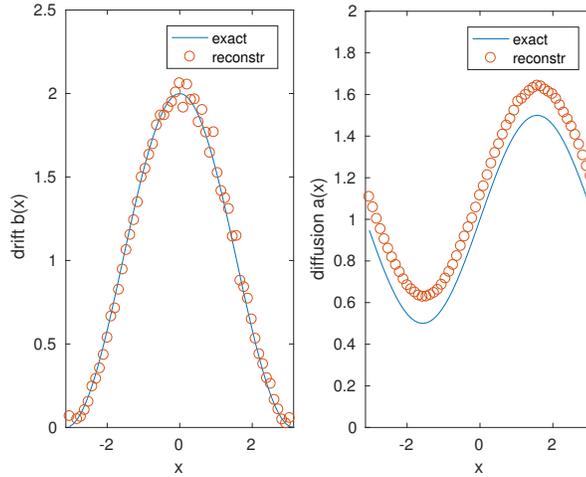
**Figure 2.7:** Reconstruction of the drift and diffusion using the generator and the adjoint as well as  $parzen_w(3, 2)$ . The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. The true values are given by the curve and the reconstructed values by the circles. On the left graph, the reconstructed drift is better than the one in Figure 2.5 but less accurate than in Figure 2.6. On the right side, the diffusion shows similar accuracy than in Figure 2.6.



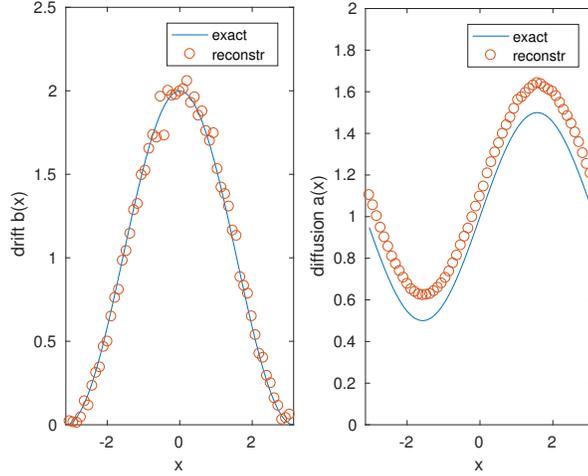
**Figure 2.8:** Reconstruction of the drift and diffusion using the entire error function (2.29) and  $parzen_w(3, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.7 implying than the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figure 2.6 and 2.7.



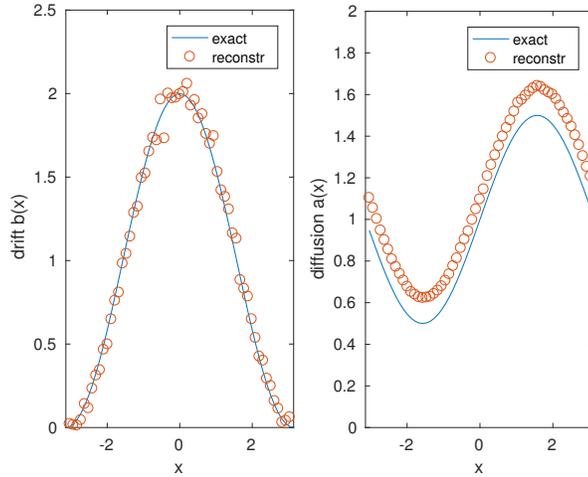
**Figure 2.9:** Reconstruction of the drift and diffusion using only the generator and  $parzen_w(5, 2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift.



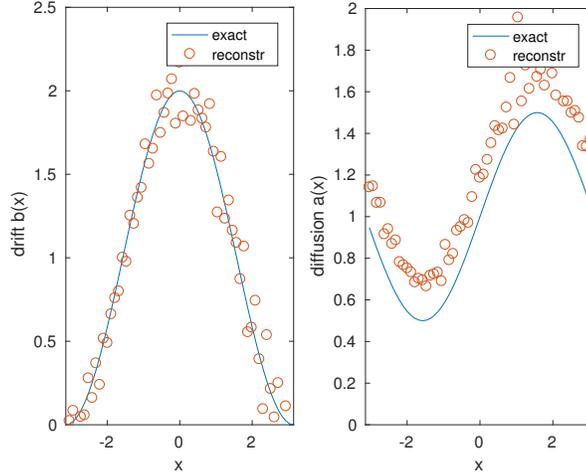
**Figure 2.10:** Reconstruction of the drift and diffusion using only the adjoint and  $parzen_w(5, 2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example  $x = 0$  more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some constant term.



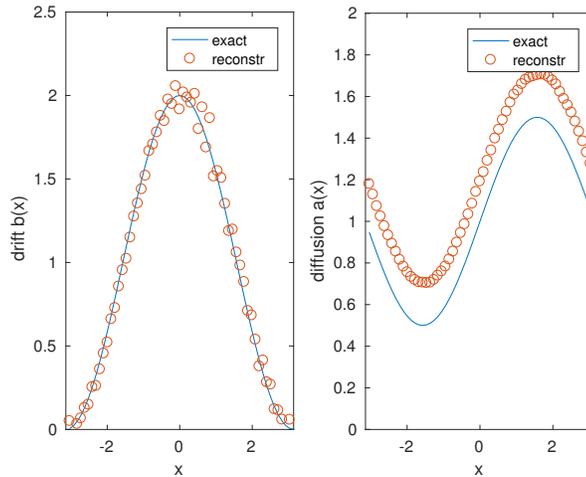
**Figure 2.11:** Reconstruction of the drift and diffusion using the generator and the adjoint as well as  $parzen_w(5, 2)$ . The true values are given by the curve and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift is better than the one in Figure 2.9 but less accurate than in Figure 2.10. On the right side, the diffusion shows similar accuracy than in Figure 2.10.



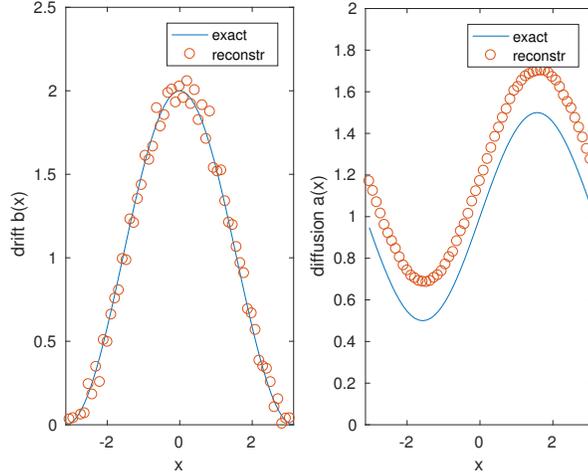
**Figure 2.12:** Reconstruction of the drift and diffusion using the entire error function (2.29) and  $parzen_w(5, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.11 implying that the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figures 2.10 and 2.11.



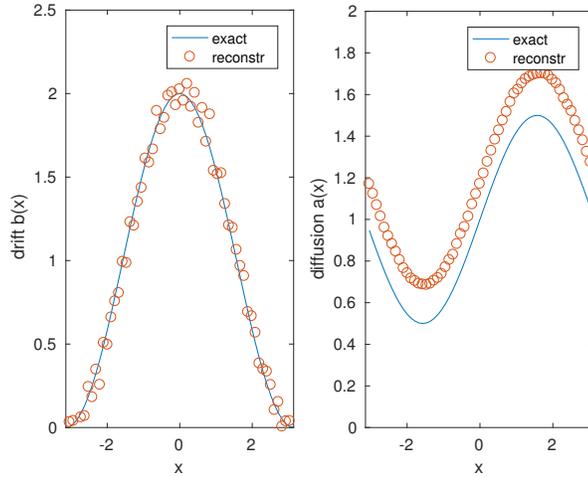
**Figure 2.13:** Reconstruction of the drift and diffusion using only the generator and  $parzen_w(7, 2)$ . The exact value of the drift is given by the line and the reconstructed value in circle. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is overall reconstructed with some oscillations that take the form of cluster of points. On the right plot, we can see that the reconstructed diffusion is close to the true value with less oscillations compared to the drift.



**Figure 2.14:** Reconstruction of the drift and diffusion using only the adjoint and  $parzen_w(7, 2)$ . The true values are given by the lines and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the drift is close to the true value even if at some space, for example  $x = 0$  more errors could be seen. On the right plot, we can see that the reconstructed diffusion matches the true diffusion minus some constant term.



**Figure 2.15:** Reconstruction of the drift and diffusion using the generator and the adjoint as well as  $parzen_w(7, 2)$ . The true values are given by the curve and the reconstructed values by the circles. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift is better than the one in Figure 2.13 but less accurate than in Figure 2.14. On the right side, the diffusion shows similar accuracy than in Figure 2.14.



**Figure 2.16:** Reconstruction of the drift and diffusion using the entire error function (2.29) and  $parzen_w(7, 2)$ . The estimates are in circles the true values are given by the curve. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis provides the values of  $b(x)$  and  $a(x)$  on the left and right plots respectively. On the left graph, the reconstructed drift has the same precision than in Figure 2.15 implying that the third term in the formula has a relative small contribution. On the right side, the diffusion, except a vertical translation, shows again similar accuracy than in previous Figures 2.14 and 2.15.

**Conclusion of this subsection:** The reconstruction method given by [18] has been applied on an ensemble of trajectories. We found that the eigenpairs are very close but the filtering of the eigenfunctions by discarding all eigenmodes of a certain values has to be adjusted to reconstruct the drift and diffusion. We modified the Fourier filter and obtain similar results seen in [18]. We also proposed a slightly different scheme using an alternative definition of the transition probabilities that we refer as Parzen windows smoothing. This latter approach systematically overestimates diffusion even if it can recover the overall shape of both drift and diffusion.

## 2.6 Extension to non periodic drift and diffusion functions

### 2.6.1 Testing reconstruction using a known generator

The goal of this subsection is to test how the non periodic boundary conditions affect the spectral reconstruction process. As a way to test our reconstruction process, the Kolmogorov Backward operator is discretized for each process using finite difference, the spectrum and eigenfunctions are computed and the reconstruction procedure is implemented. We emphasize that no data are used. In our testing approach, the generator is discretized over an interval and expressed in matrix form, and this matrix now contains the information that would permit reconstruction via its eigenvalues and eigenfunctions. Note that we restrict our applications to polynomial eigenfunctions processes such as OU, CIR, and JAC, because an analytic expression of their

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

---

eigenvalues and eigenfunctions exist which was given in the previous section. The affine processes under investigation are:

$$\begin{aligned}
 dX_t &= \kappa(\theta - X_t)dt + \gamma dW_t, \\
 dX_t &= \kappa(\theta - X_t)dt + \gamma\sqrt{X_t}dW_t, \\
 dX_t &= \kappa(\theta - X_t)dt + \gamma\sqrt{(X_t - a)(b - X_t)}dW_t.
 \end{aligned}
 \tag{2.35}$$

We chose  $(\kappa, \theta, \gamma) = (1, 1, 1)$  in the above processes. Note that using these coefficients the Feller condition ( $2\kappa\theta > \sigma_2$ ) is satisfied. The interval  $[a, b]$  is chosen to be  $[a, b] = [\theta - 1, \theta + 2] = [0, 3]$ .

### Construction of a finite representation of the generator

Using these SDEs a discretized version of generators corresponding to each of the above processes can be constructed and written in matrix form. Consider an interval  $I$ . Discretize  $I$  into  $M$  points,  $\{x_i\}$  with  $i = 1, \dots, M$ . Now define for  $i = 1..M$  the values of a drift  $b(x_i) = b_i$ , a diffusion  $a(x_i) = a_i$  and a smooth function  $f(x_i) = f_i$  representing the solution of the Kolmogorov equation. Using the finite difference operators (2.30), the discretized Backward differential equation is

$$Lf_i = b_i D_i f + \frac{1}{2} a_i D_i^2 f.
 \tag{2.36}$$

Note, that to compute the derivative at each point  $x_i$  the value of the function at  $x_{i-1}$  and  $x_{i+1}$  are needed. Boundary conditions need to be supplied to implement this scheme. For diffusions, we generally encounter three types of boundary conditions: absorbing, reflecting, and periodic. The idea is that the value of a stochastic process at an absorbing barrier or boundary is imposed or a priori known, it could be zero or another constant. A reflecting barrier would mean that the value of the process at

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

---

the barrier is the same than the value right before the boundary. Finally, a periodic boundary means that the values of the process at the extreme are the same. More details could be found in [27]. In mathematical terms, these boundary conditions can be formulated as follows. Assume that we found the solution under some boundary conditions of a equation of the type (2.36) at each discretization point  $x_l = x_1 < x_2 < \dots < x_M = x_u$ . Periodic boundary conditions in this context would imply that  $f(x_l) = f(x_u)$ , absorbing boundary conditions  $f(x_l) = f(x_u) = 0$  and reflecting boundaries  $f(x_l) = f(x_{l+1})$  and  $f(x_{u-1}) = f(x_u)$ . Each condition would modify the structure of the matrix representing the generator. In our work, since no functional form is assumed for the drift and diffusion, the boundary conditions are difficult to impose. In that case, we chose to omit the boundary points.

The discretized generator is given generally by a matrix of the form:

$$\begin{bmatrix} \frac{a_1}{2(\Delta x)^2} - \frac{b_1}{2(\Delta x)} & -\frac{a_1}{(\Delta x)^2} & \frac{a_1}{2(\Delta x)^2} + \frac{b_1}{2(\Delta x)} & \dots & 0 \\ 0 & \frac{a_2}{2(\Delta x)^2} - \frac{b_2}{2(\Delta x)} & -\frac{a_2}{(\Delta x)^2} & \dots & \vdots \\ \vdots & \ddots & \dots & \dots & \vdots \\ \vdots & \dots & \frac{-a_{M-1}}{(\Delta x)^2} & \frac{a_{M-1}}{2(\Delta x)^2} + \frac{b_{M-1}}{2(\Delta x)} & 0 \\ 0 & \dots & \dots & -\frac{a_M}{(\Delta x)^2} & \frac{a_M}{2(\Delta x)^2} + \frac{b_M}{2(\Delta x)} \end{bmatrix}$$

Note that the rows should sum to zero when boundary conditions are imposed. Since we opted for no condition in the boundaries, the sum of the entries of the first and last rows of our discretized generators would be different than zero. The first and

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

last columns are associated to the boundary conditions and removing them gives:

$$\begin{bmatrix} -\frac{a_1}{(\Delta x)^2} & \frac{a_1}{2(\Delta x)^2} + \frac{b_1}{2(\Delta x)} & 0 & \cdots & 0 \\ \frac{a_2}{2(\Delta x)^2} - \frac{b_2}{2(\Delta x)} & -\frac{a_2}{(\Delta x)^2} & \frac{a_2}{2(\Delta x)^2} + \frac{b_2}{2(\Delta x)} & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \cdots & \frac{a_{M-1}}{2(\Delta x)^2} - \frac{b_{M-1}}{2(\Delta x)} & -\frac{a_{M-1}}{(\Delta x)^2} & \frac{a_{M-1}}{2(\Delta x)^2} + \frac{b_{M-1}}{2(\Delta x)} \\ 0 & \cdots & 0 & \frac{a_M}{2(\Delta x)^2} - \frac{b_M}{2(\Delta x)} & -\frac{a_M}{(\Delta x)^2} \end{bmatrix}$$

The eigenvalues and eigenvectors of the matrix are computed and the drift and diffusion estimators are obtained by minimizing the error function (2.31).

### About the Optimization

Regarding the minimization, two approaches are possible. The first one, suggested by [18], consisted in writing the above error in terms of a constraint quadratic programming problem whose objective function is:

$$X^T H_{\lambda_k} X + F_{\lambda_k}^T X, \quad (2.37)$$

where the matrix,  $H_{\lambda_k}$ , and the vector,  $F_{\lambda_k}$ , represent the discretization of the operators and the column vector  $X$ , the coefficients written as  $X = (b_1, \dots, b_M, a_1, \dots, a_M)$ . This approach is a little tedious because the adjoint term in the error contains the derivatives of the product  $b(x)f(x)$  and  $a(x)f(x)$  that need to be discretized and then squared, this complicates the expansion. A second, and equivalent way, to approach this estimation is to consider that the summation over the number of data  $i$  is the error term of a least squares problem, and the overall sum over the number of eigenvalues  $K$  used represent a weighted sum of least squares problems. We therefore separately construct for each  $\lambda_k$  a matrix,  $H_{\lambda_k}$ , representing the discretized

operator, multiply each of the matrix by the weights, and combine the matrices. As an example, take the expression for the generator:  $K$  matrices are constructed, multiplied by a weight, and then pooled to created a big matrix  $H$  where the first  $N$  block represents the first  $H_{\lambda_1}$ , second block  $H_{\lambda_2}$ , etc. We then used the built in MATLAB function for constraint least squares *lsqlin*. This way to program the optimization problem appears more flexible since different inference techniques can be used, ridge regression for example. Both approaches have been tested and on the generator, they give similar results. We chose to use the second approach to perform inference in the rest of this chapter.

### Estimation using different boundary conditions

In order to determine what needs to be done for the condition at the boundary, we tested the two following alternatives, (i) imposing no conditions or (ii) reflecting boundary condition. We do not consider the absorbing boundary condition since it is a rather strong assumption on the process. Using 40 bins for a CIR process, we present in the Figures 2.17, 2.18, and 2.19 the reconstruction when reflective boundary conditions are imposed to the discretized generators. Figures 2.20, 2.21, and 2.22 show the reconstruction when no boundary conditions are imposed. These figures show that the boundary conditions affect mostly the reconstruction using error function including the adjoint and the bi-orthonormality constraint i.e. the error function (2.31). The same type of behavior has been observed when we consider other processes (OU and JAC). This leads to the conclusion than the error function involving only eigenequation of the generator appears more robust to change of conditions at the boundary than the more general error function (2.31). This is consistent with

[18] where they decided to keep only the generator term in their subsequent paper [20].

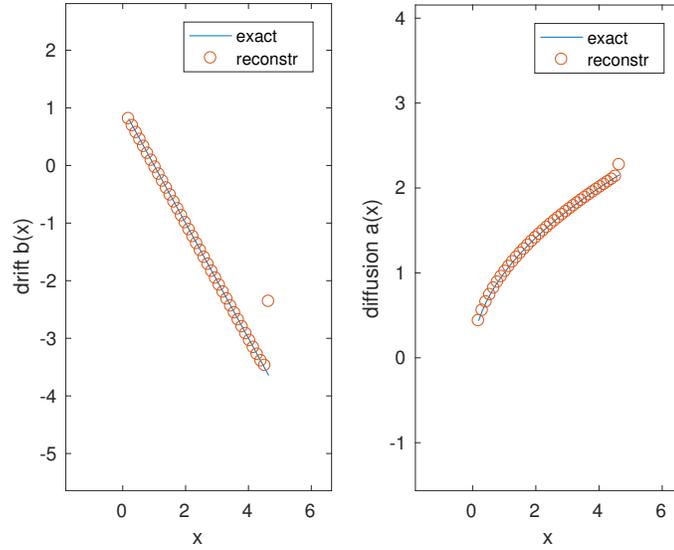
**Estimation using discretized generator of drift and diffusion of affine processes**

No data was used. The goal is to test how the boundary conditions affect the reconstruction. We focused on the reconstruction based on a discretized generator with no boundary conditions and an error function that only involves the eigenequation of the generator. The number of bins is set to 40 and only the first 3 eigenvalues are used. The eigenvalues of the discretized generator are very close to their theoretical counterparts. Table 2.7 shows the first 3 eigenvalues for each process. The method perfectly reconstructs the drift and diffusion of the targeted processes as shown in Figures 2.23 to 2.28.

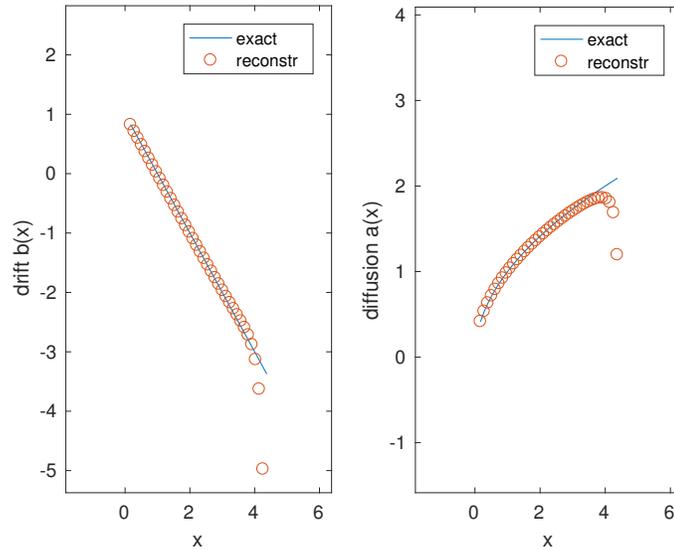
	True value	OU	CIR	JAC
$\lambda_1$	0	-0.0109	0.0639	-0.0062
$\lambda_2$	-1	-1.0795	-0.9988	-1.0188
$\lambda_3$	-2 (-3 for JAC)	-2.2701	-2.4444	-3.0315

**Table 2.7:** The eigenvalues obtained from the discretized generator for the three affine processes is compared to their real values.

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

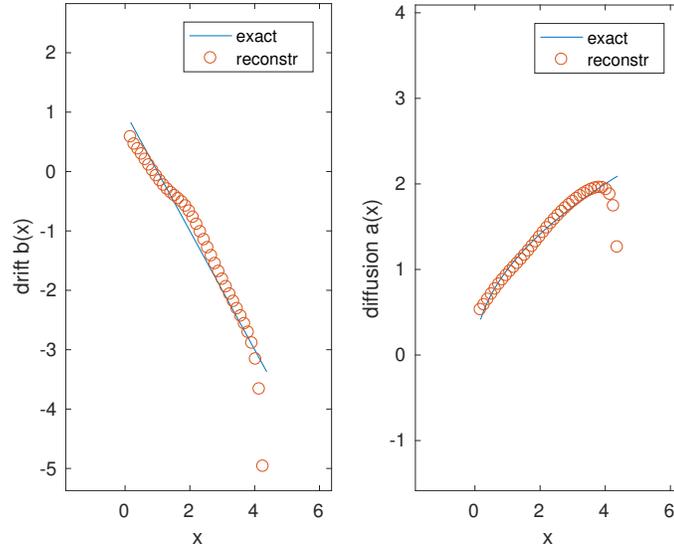


**Figure 2.17:** Reconstruction of the drift and diffusion using only the generator not the adjoint of a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively.

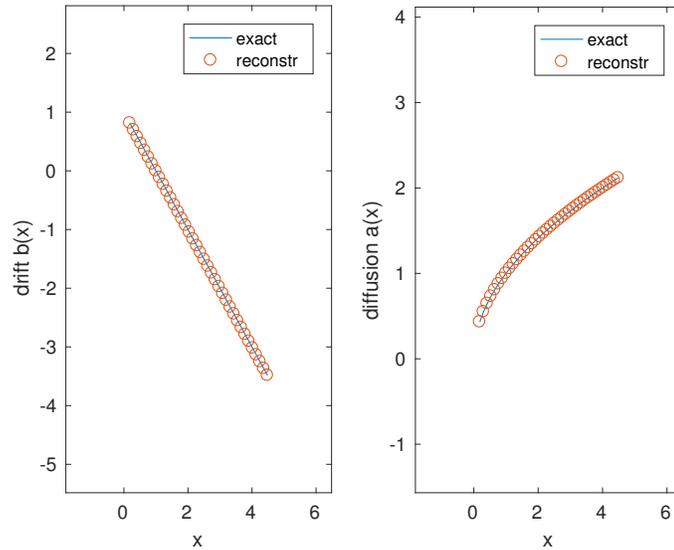


**Figure 2.18:** Reconstruction of the drift and diffusion using only the adjoint not the generator of a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively .

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

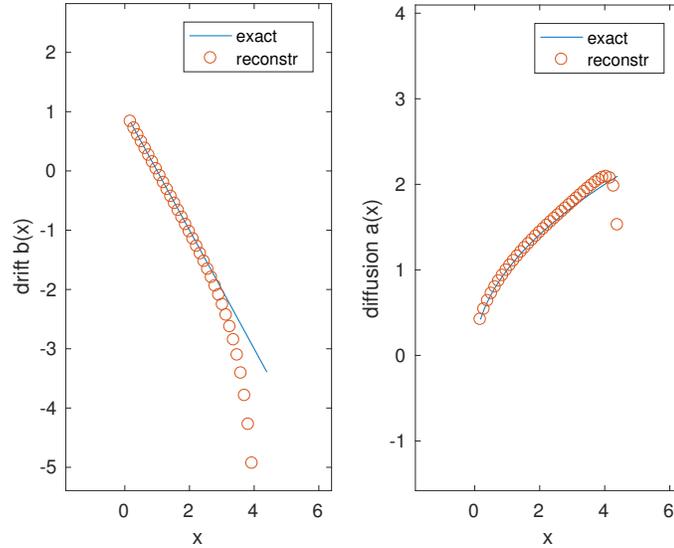


**Figure 2.19:** Reconstruction of the drift and diffusion using the error function (2.29) and a discretized infinitesimal generator with reflective boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively .

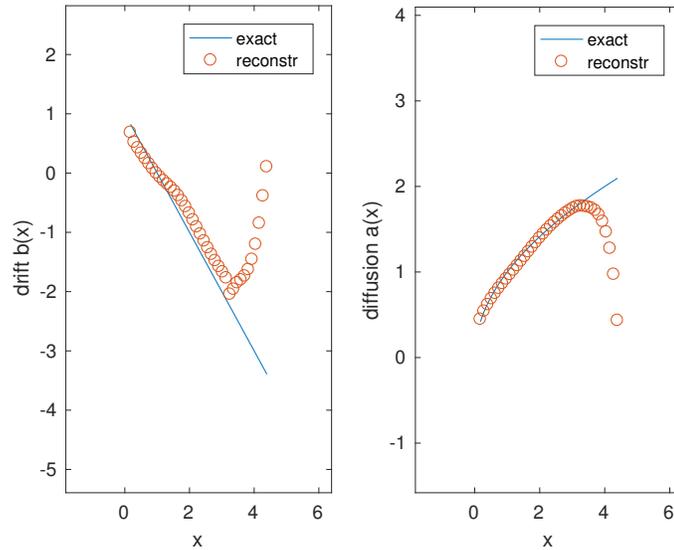


**Figure 2.20:** Reconstruction of the drift and diffusion using the generator, not the adjoint, and a discretized infinitesimal generator without boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively.

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

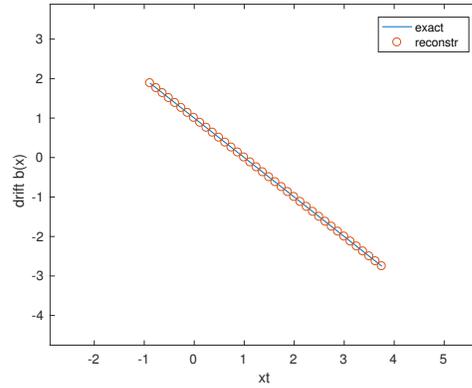


**Figure 2.21:** Reconstruction of the drift and diffusion using only the adjoint and not the generator with a discretized infinitesimal generator without boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively.

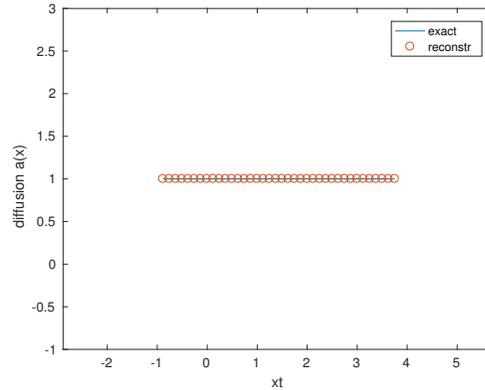


**Figure 2.22:** Reconstruction of the drift and diffusion using the error function (2.29) and a discretized infinitesimal generator without boundary conditions. The x-axis gives the values,  $x$ , taken by the process  $X_t$ . The y-axis shows on the left and right plots the values of  $b(x)$  and  $a(x)$  respectively.

### Reconstruction of the OU process

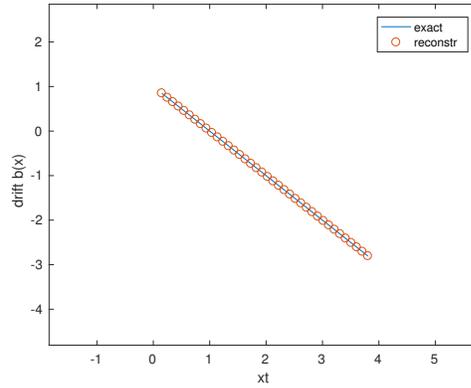


**Figure 2.23:** Reconstruction of the drift of the OU process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $b(x)$ . The true values are given by the line and estimates by the circles.

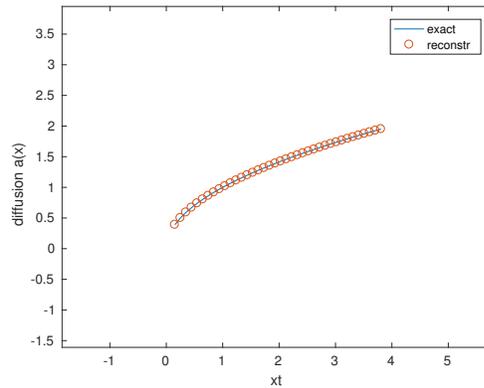


**Figure 2.24:** Reconstruction of the diffusion of the OU process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $a(x)$ . The true values are given by the line and estimates by the circles.

### Reconstruction of the CIR process

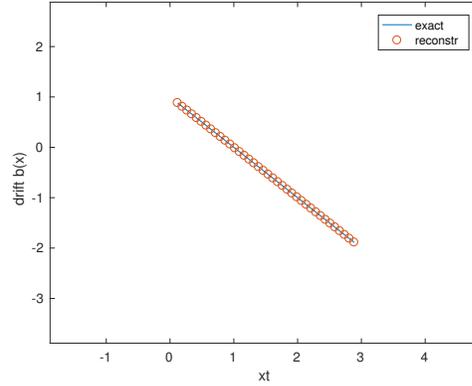


**Figure 2.25:** Reconstruction of the drift of a CIR type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $b(x)$ . The true values are given by the line and estimates by the circles.

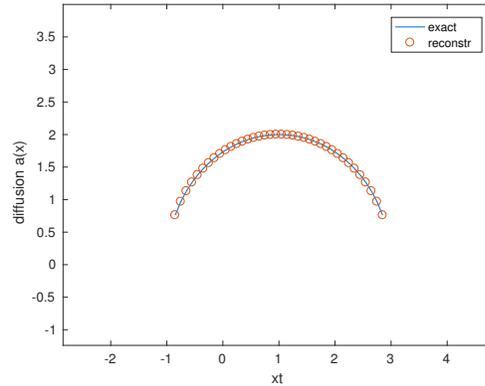


**Figure 2.26:** Reconstruction of the diffusion of a CIR type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $a(x)$ . The true values are given also by the line and estimates by the circles.

### Reconstruction of the JAC process



**Figure 2.27:** Reconstruction of the drift of a Jacobi type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $b(x)$ . The true values are given by the line and estimates by the circles.



**Figure 2.28:** Reconstruction of the diffusion of a Jacobi type process using the discretization of its known generator and no boundary conditions. The x-axis gives different values,  $xt$ , taken by the process. The y-axis shows the values of  $a(x)$ . The true values are given by the line and estimates by the circles.

**Conclusion of this subsection:** the objective of this section was to test how boundary conditions affect the spectral reconstruction. We do not use discrete data, instead we use a discretized generator to address this issue. We found that the adjoint and last terms in (2.29) could be source of large errors since they appear sensitive to boundary conditions. Using the error function that includes only the generator eigenequation, we have been able to reconstruct the drift and diffusion of each affine process using spectral data. Therefore, we will in the next section use only a generator-based error function.

### 2.6.2 The number of bins or spatial step size

The choice of the time step and spatial size are crucial in the reconstruction. The stability of the finite difference scheme is used to understand the spatial discretization. Recall that given a suitable function,  $f$ , the infinitesimal generator is

$$Lf = b(x)\frac{\partial f(x)}{\partial x} + \frac{1}{2}a(x)\frac{\partial^2 f(x)}{\partial x^2}.$$

Using central finite differences, the infinitesimal generator for every point  $x_i$  can be written as:

$$\begin{aligned} Lf(x) &= \sum_{i=1}^N b(x_i)D_i f(x_i) + \frac{1}{2}a(x_i)D_i^2 f(x_i), \\ &= \sum_{i=1}^N b_i D_i f_i + \frac{1}{2}a_i D_i^2 f_i, \\ &= \sum_{i=1}^N b_i \left( \frac{f_{i+1} - f_i}{2\Delta x} \right) + \frac{1}{2}a_i \left( \frac{f_{i-1} - 2f_i + f_{i+1}}{\Delta x^2} \right), \end{aligned}$$

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

---

where  $b_i = b(x_i)$ ,  $a_i = a(x_i)$ , and  $f_i = f(x_i)$ . This forward Euler discretization is similar to the scheme of a one dimensional advection-diffusion equation with space-dependent parameters:

$$\frac{\partial f(t, x)}{\partial t} = b(x) \frac{\partial f(t, x)}{\partial x} + \frac{1}{2} a(x) \frac{\partial^2 f(t, x)}{\partial x^2}.$$

Discretizing advection and diffusion terms using second order central differences gives for a point  $x_j$ :

$$\frac{\partial f_j}{\partial t} = b_j \left( \frac{f_{j+1} - f_j}{2\Delta x} \right) + \frac{a_j}{2} \left( \frac{f_{j-1} - 2f_j + f_{j+1}}{\Delta x^2} \right). \quad (2.38)$$

If we assume a constant drift and diffusion, then the stability of the finite difference approximation can be obtained using the Courant-Friedrichs-Lewy (henceforth CFL) condition [66]. The advection equation is known to be unstable when using a forward Euler scheme with a constant or variable diffusion is :

$$\left| \frac{a\Delta t}{\Delta x^2} \right| \leq 1 \quad \text{or} \quad \left| \frac{\sup_x(a(x))\Delta t}{\Delta x^2} \right| \leq 1.$$

The diffusion equation stabilizes the advection and a stability condition can be obtained by using a Von Neumann Analysis [66] as follows. Substitute in (2.38)

$$f_n = \rho e^{ikx_n}, \quad x_{n+1} = x_n + \Delta x, \quad \text{and} \quad x_{n-1} = x_n - \Delta x.$$

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

We then take the derivative of  $f_n$  with respect to  $t$  :

$$\begin{aligned}
\frac{\partial f_n}{\partial t} &= b_n \left( \frac{f_{n+1} - f_n}{2\Delta x} \right) + \frac{a_n}{2} \left( \frac{f_{n-1} - 2f_n + f_{n+1}}{\Delta x^2} \right), \\
&= b_n \left( \frac{\rho e^{ikx_{n+1}} - \rho e^{ikx_{n-1}}}{2\Delta x} \right) + \frac{a_n}{2} \left( \frac{\rho e^{ikx_{n-1}} - 2\rho e^{ikx_n} + \rho e^{ikx_{n+1}}}{\Delta x^2} \right), \\
&= b_n \frac{\rho e^{ikx_n}}{\Delta x} \left( \frac{e^{ik\Delta x} - e^{-ik\Delta x}}{2} \right) + a_n \frac{\rho e^{ikx_n}}{\Delta x^2} \left( \frac{e^{-ik\Delta x} - 2 + e^{ik\Delta x}}{2} \right), \\
&= b_n \frac{\rho e^{ikx_n}}{\Delta x} \left( \frac{e^{ik\Delta x} - e^{-ik\Delta x}}{2} \right) + a_n \frac{\rho e^{ikx_n}}{\Delta x^2} \left( \frac{-2 + e^{ik\Delta x} + e^{-ik\Delta x}}{2} \right), \\
&= b_n \frac{\rho e^{ikx_n}}{\Delta x} (i \sin(k\Delta x)) + a_n \frac{\rho e^{ikx_n}}{\Delta x^2} (-1 + \cos(k\Delta x)), \\
&= i b_n f_n \frac{(\sin(k\Delta x))}{\Delta x} - a_n f_n \frac{(1 - \cos(k\Delta x))}{\Delta x^2}.
\end{aligned}$$

The usual CFL conditions for the diffusion and advection equations involve the time step. Then let's introduce the following variables that are in fact the CFL numbers of the advection, denoted by  $C_A$ , and the diffusion,  $C_D$ . Note that they depend on  $n$ . In practice, when the velocity and diffusion parameters are variable, the supremum of these functions is taken and the two constant can be written as:

$$C_A = \frac{\tilde{b}\Delta t}{\Delta x} \quad \text{and} \quad C_D = \frac{\tilde{a}\Delta t}{\Delta x^2}.$$

with  $\tilde{b} = \sup b_n$  and  $\tilde{a} = \sup a_n$ .

$$\begin{aligned}
\frac{\partial f_n}{\partial t} &= i b_n f_n \frac{(\sin(k\Delta x))}{\Delta x} - a_n f_n \frac{(1 - \cos(k\Delta x))}{\Delta x^2}, \\
&= i C_A f_n \frac{(\sin(k\Delta x))}{\Delta t} - C_D f_n \frac{(1 - \cos(k\Delta x))}{\Delta t}, \\
&= \left( i C_A \frac{(\sin(k\Delta x))}{\Delta t} - C_D \frac{(1 - \cos(k\Delta x))}{\Delta t} \right) f_n.
\end{aligned}$$

If we consider the time discretization again using the forward Euler scheme for the left hand side, it yields the following:

$$f_n^{t+1} = (1 - C_D (1 - \cos(k\Delta x)) + i C_A (\sin(k\Delta x))) f_n^t.$$

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

This scheme is stable if

$$|1 - C_D (1 - \cos(k\Delta x)) + i C_A (\sin(k\Delta x))| < 1.$$

This constraint is equivalent to the following

$$(1 - C_D (1 - \cos(k\Delta x)))^2 + (C_A (\sin(k\Delta x)))^2 < 1.$$

Expanding the constraint

$$\begin{aligned} (1 - C_D (1 - \cos(k\Delta x)))^2 + (C_A (\sin(k\Delta x)))^2 &< 1, \\ 1 - 2C_D(1 - \cos(k\Delta x)) + C_D^2(1 - \cos(k\Delta x))^2 + C_A^2 \sin^2(k\Delta x) &< 1, \\ -2C_D(1 - \cos(k\Delta x)) + C_D^2(1 - \cos(k\Delta x))^2 + C_A^2(1 - \cos^2(k\Delta x)) &< 0, \\ (1 - \cos(k\Delta x)) (-2C_D + C_D^2(1 - \cos(k\Delta x)) + C_A^2(1 + \cos(k\Delta x))) &< 0. \end{aligned}$$

This reduces to

$$\begin{aligned} -2C_D + C_D^2(1 - \cos(k\Delta x)) + C_A^2(1 + \cos(k\Delta x)) &< 0, \\ C_A^2 - 2C_D + C_D^2 &< (C_D^2 - C_A^2) \cos(k\Delta x). \end{aligned}$$

The value of the right hand side defines different cases

- case 1:  $\cos(k\Delta x) = 0$

$$C_A^2 - 2C_D + C_D^2 < 0 \Rightarrow C_A^2 < 2C_D - C_D^2 \Rightarrow C_A^2 < 2C_D(1 - \frac{C_D}{2}),$$

- case 2:  $\cos(k\Delta x) = 1$

$$C_A^2 - 2C_D + C_D^2 < (C_D^2 - C_A^2) \Rightarrow C_A^2 < C_D,$$

- case 3:  $\cos(k\Delta x) = -1$

$$C_A^2 - 2C_D + C_D^2 < (C_A^2 - C_D^2) \Rightarrow C^D(C^D - 1) < 0.$$

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

Case 3 is only possible if  $0 < C_D < 1$ . Using now case 2, the condition becomes  $C_A^2 < C_D < 1$ . From conditions 1 and 3, one can see that  $C_D < 2C_D(1 - \frac{C_D}{2})$ . In general, the first condition is neglected and only the second and third ones are used. Therefore, our condition is

$$C_A^2 < C_D < 1. \quad (2.39)$$

The first inequality gives a bound for the time step that is already met by the small time step used. We therefore focused on the second inequality. Using the definitions of the spatial step and of the CFL of the diffusion,

$$\begin{aligned} C_D < 1 &\iff \frac{\tilde{a}\Delta t}{\Delta x^2} < 1, \\ &\Rightarrow \tilde{a}\Delta t < (\Delta x)^2 = \left(\frac{R}{M}\right)^2, \\ &\Rightarrow M^2 < \frac{R^2}{\tilde{a}\Delta t} \iff M < \frac{R}{\sqrt{\tilde{a}\Delta t}} \end{aligned}$$

with  $R$  standing for the range of the sample path and  $M$  for the number of bins. Recall that the diffusion in the generator denoted,  $a(X_t)$ , is the square of the diffusion component of the SDE,  $\sigma(X_t)$ . Using an estimate of the diffusion of the SDE, denoted  $\tilde{\sigma}$ , and such that  $\tilde{a} = (\tilde{\sigma})^2$ , the bound for  $M$  or the spatial step size could be equivalently written as

$$M < \frac{R}{\tilde{\sigma}\sqrt{\Delta t}} \quad \text{or} \quad \Delta x > \tilde{\sigma}\sqrt{\Delta t}. \quad (2.40)$$

For practitioners, these inequalities provide a criterion to select the number of bins or the spatial step size. More precisely, to pick  $M$  for some data observed at a fixed time step, one can calculate its upper bound using the range, standard deviation, and the square root of the sampling frequency of the observations. Equivalently, if a

spatial step size has to be selected, it should be greater than the square root of the sampling time multiplied by the standard deviation of the data.

**Conclusion of this subsection:** Using stability analysis of a diffusion equation with constant parameters, we derived an upper bound on the number of bins, and equivalently the spatial step size. From this result, one could conclude that the spatial step size should be larger than the time step. For practical purposes, the inequalities in (2.40) provide a criterion for the spatial discretization in the reconstruction process.

### 2.6.3 Reconstruction using simulated data of affine processes

Before exploring extension to non-periodic drift and diffusion, we discuss in this paragraph the optimization step. As mentioned in [18], it's possible for one to perform reconstructions using only the generator and not the adjoint or the adjoint and not the generator. Our tests on the discretized generator of known affine processes suggest that the error function involving only the generator eigenequation provides a more robust estimation. We will then for the rest of this work concentrate on this type of error function to reconstruct the drift and diffusion. Our goal is to minimize an objective function of the form:

$$\tilde{E}(b, a) = \sum_{k=1}^K \beta_k \sum_{i=1}^N \left| b_i D_i(\tilde{\phi}_k) + \frac{1}{2} a_i D_i^2(\tilde{\phi}_k) - \tilde{\lambda}_k \tilde{\phi}_{k,i} \right|^2 \quad (2.41)$$

This is equivalent to setting  $\alpha_k = \gamma_k = 0$  in the error function (2.31). The next step of testing the spectral reconstruction procedure consists of using synthetic data generated by different affine processes.

## Simulation of affine processes

### 1. Simulation of OU process

The simulation of such process can be done using the Euler Maruyama scheme, [32]. To simulate  $T$  sample points of (2.3), with time step  $h = t_{i+1} - t_i$  and initial condition  $X_0 = x_0$ , we used the discretized version of the SDE which is  $X_{t_{i+1}} - X_{t_i} = \kappa(\theta - X_{t_i})h + \gamma \sqrt{h} z_{t_i}$ , with  $z_{t_i} \sim \mathcal{N}(0, 1)$  to generate each new value.

### 2. Simulation of CIR process

Euler scheme also could be used for this process but it could produce complex values. One way to simulate the CIR process avoiding negative and complex values makes use of the transition probability density of the process [32]. Suppose that we want to simulate  $T$  sample points of (2.7) with time step  $h = t_{i+1} - t_i$  and initial condition  $X_0 = x_0$ . The transition density is, for time  $u < t$ , a non central Chi squared of the form:

$$X_t | X_u \sim \frac{1 - e^{-(t-u)}}{4} \chi_d^2 \left( \frac{4e^{-(t-u)}}{1 - e^{-(t-u)}} X_u \right), \quad (2.42)$$

where  $\chi_d^2(\lambda)$  is a non central Chi squared with  $d$  degrees of freedom and a non-centrality parameter,  $\lambda$ . Each new value of the CIR process  $X_t$  could be obtained using the following loop. For  $i = 1 \dots T - 1$  The following steps were done with  $d = 4$ :

- step 1: generate  $c \leftarrow \frac{1 - e^{-(t_{i+1} - t_i)}}{4}$ ,
- step 2: generate  $\lambda \leftarrow \frac{e^{-(t_{i+1} - t_i)}}{c} X_{t_i}$ ,
- step 3: generate  $z_{t_i} \sim \mathcal{N}(0, 1)$ ,

step 4: generate  $R \sim \chi_{d-1}^2$ ,

step 5: compute  $X_{t_{i+1}} = c \left[ (z + \sqrt{\lambda})^2 + R \right]$ ,

step 6: if  $i < T$  do  $i \leftarrow i + 1$  and go back to step 1. If  $i = T$ , stop.

### 3. Simulation of JAC process

The simulation of such process can be done using a truncated version of the Euler Maruyama scheme as shown in [34]. Suppose that we want to simulate  $T$  sample points of (2.11), at each time point,  $t_i$ , with  $i = 1 \dots T$ , with time step,  $h = t_{i+1} - t_i$ , initial condition,  $X_0 = x_0$ , and in an interval  $[a, b]$ . The discretized version of the SDE (2.11)  $X_{t_{i+1}} - X_{t_i} = \kappa(\theta - X_{t_i})dt + \gamma\sqrt{(X_{t_i} - a)(b - X_{t_i})}\sqrt{h} z_{t_i}$ . with  $z_{t_i} \sim \mathcal{N}(0, 1)$  is used to simulate new points. However, to guarantee that the process remains inside the interval  $[a, b]$ , we can't keep all the  $X_{t_{i+1}}$  generated with the Euler scheme. Hence, we only validate as a solution of the Jacobi process, values  $\hat{X}_{t_{i+1}}$  defined as:

$$\hat{X}_{t_{i+1}} = \begin{cases} b - 0.01 & , \text{ if } X_{t_{i+1}} > b \\ X_{t_{i+1}} & , \text{ if } X_{t_{i+1}} \in [a, b] \\ a + 0.01 & , \text{ if } X_{t_{i+1}} < a. \end{cases} \quad (2.43)$$

We generate only one trajectory and not 200 paths seen in the previous subsection. The drift and diffusion are no longer a continuous periodic function, so we can't guarantee that all the bins will be visited. If bins are not visited at least once, we would have rows in our stochastic matrix that do not sum to 1. It is computationally expensive to randomly generate 200 paths to visit every bin. To simplify our study, we generate only one path and ensure that all the rows sum to 1. Another adjustment

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

---

is better results are obtained with a subsampling time,  $h$ , smaller than the spatial step size,  $\Delta x$ . This is also a conclusion of the upper bound (2.40) that we derived for the space discretization. We used  $h = 0.01$  instead of  $h = 0.1$  in [18] chosen with 60 bins such a way that the subsampling time is closed to the spatial step, i.e.  $\Delta x = 0.105$ . When dealing with non-periodic functions, we would like  $h < \Delta x$ . This means that our bins are a little bit larger and have higher probability of being visited at least once. Therefore, we use the following parameters to generate the trajectories and test the impact of the boundaries:  $T = 10^6$ ,  $h = 0.01$ ,  $dt = 10^{-4}$ , and the number of bins is equal to 40 such that it satisfies the upper bound (2.40). The number of eigenvalues and eigenfunctions used for reconstruction is set at 3. We also tried to use a higher number for eigenvalues and eigenfunctions to improve the estimation. In particular we tried  $K = 5, 7, 10$  eigenpairs. But more eigenpairs didn't significantly affect the outcome of the reconstruction. The parameters for the affine processes are the same as before  $(\kappa, \theta, \gamma) = (1, 1, 1)$ . The eigenvalues and eigenfunctions of the discretized generator are compared to the eigenfunctions obtained by the data-based Markov chains. From Table 2.8, we can see that the estimated eigenvalues are close to their theoretical values except for the OU process where they overestimated. The eigenfunctions, and their derivatives are compared to the ones obtained from a discretized generator in Tables 2.9, 2.10, and 2.11. The largest deviations from the reference values appear in the computation of the numerical second derivatives. We proceed with the estimation for each process of the drift and diffusion.

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

	True value	OU	CIR	JAC
$\lambda_1$	0	1.55e-13	-1.22e-13	-2.11e-13
$\lambda_2$	-1	-0.8100	-1.0004	-0.9882
$\lambda_3$	-2 (-3 for JAC)	-1.6544	-2.8196	-3.0718

**Table 2.8:** The eigenvalues obtained from data-based Markov chain for the three affine processes are compared to their real value.

OU Process						
$K$	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	0.1724	0.1024	0.4348	0.4049	1.9479	5.0605
2	0.5055	0.7406	0.9892	1.7846	7.6413	17.8950
3	1.8511	1.6370	2.3809	3.5300	12.5390	13.3690

**Table 2.9:** Comparison the difference of eigenpairs and their derivatives obtained for the OU process with those of a discretized generator.

CIR Process						
$K$	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	1.9785	1.9962	0.4782	2.3482	3.1038	11.5840
2	0.3993	0.5575	0.7576	1.6360	2.8979	7.3186
3	1.9244	1.7860	1.4645	4.6457	3.8957	23.7660

**Table 2.10:** Comparison the difference of eigenpairs and their derivatives obtained from the CIR process with those of a discretized generator.

Jacobi Process						
$K$	$\ \phi - \phi^A\ _2$	$\ \psi - \psi_x^A\ _2$	$\ \phi_x - \phi_x^A\ _2$	$\ \psi_x - \psi_x^A\ _2$	$\ \phi_{xx} - \phi_{xx}^A\ _2$	$\ \psi_{xx} - \psi_{xx}^A\ _2$
1	0.2687	0.2574	0.5295	1.3269	4.3172	30.7610
2	1.9747	1.9706	2.3178	3.9779	7.7879	50.9240
3	0.3088	0.3244	1.5420	1.3981	14.3080	33.9870

**Table 2.11:** Comparison the difference of eigenpairs and their derivatives obtained for the Jacobi process with those of a discretized generator.

### OU process

For the OU process, note that the spatial step size is  $dx = 0.11596$ . Estimation of the drift and diffusion are given in Figures 2.29 and 2.30, respectively. The linear fit and the 95% confidence bounds gives a drift with a slope of  $-1.4192 \pm 0.4945$  and a constant term of  $1.5167 \pm 0.8024$ . The true values are definitely in the interval. A quadratic fit  $p1 * x^2 + p2 * x + p3$  with 95% bounds is used for the squared of the diffusion. We obtained that only  $p3$  is significantly different than zero using a t-statistics test and is close to 1, its true value. More precisely,  $p1 = p2 = 0$  and  $p3 = 1.379 \pm 0.2877$ .

### CIR process

The spatial step used with 40 bins is  $dx = 0.092511$ . MATLAB financial toolbox provides a command *cir* to simulate the square root process. We use this command to generate paths and to estimate the drift and diffusion. We compare the latter reconstructed components with the ones obtained from generating paths using the transition probability density. The estimates are the same using both simulation techniques. We found no significant intercept for both the drift and the squared diffusion. The linear coefficient in the drift is  $p1 = -2.8739 \pm 1.5176$  which is a large confidence interval for a true value of -1. Part of the diffusion function could be recovered. The quadratic fit  $p1 * x^2 + p2 * x + p3$  for the squared diffusion gives significant coefficients at 95% confidence level:  $p1 = -0.6061 \pm 0.2562$ ,  $p2 = 2.1984 \pm 0.9809$  and  $p3 = 0$ . This implies that the drift has an extra linear term since  $p1$  is significantly different than zero. After taking the square root, the diffusion becomes a linear function of the state. The reconstructed drift and diffusion are

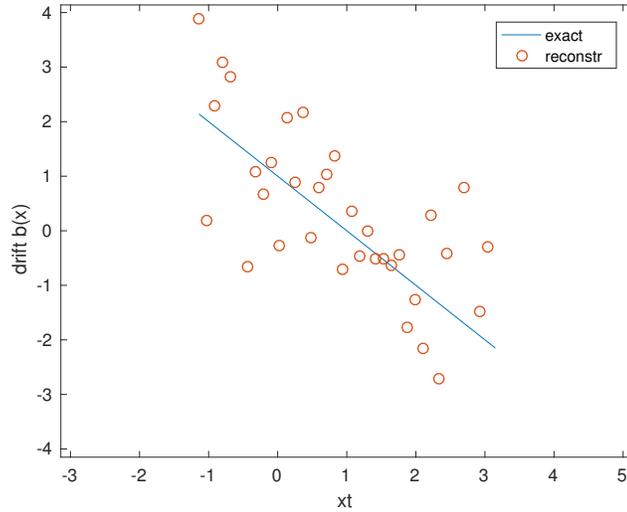
## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

shown in Figures 2.31 and 2.32.

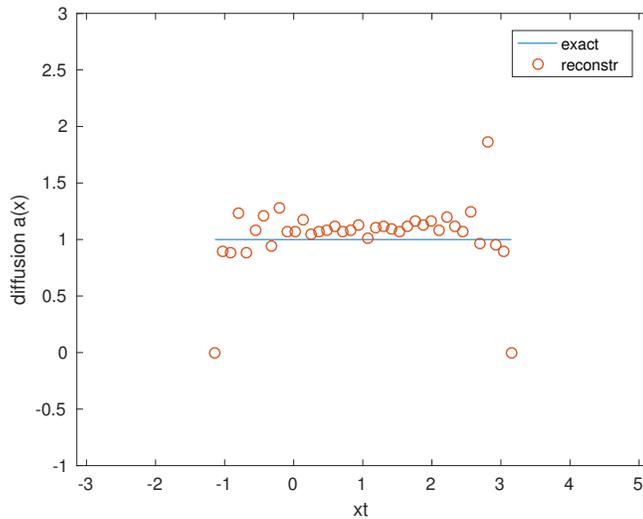
### **Jacobi process**

For the Jacobi process, note that the spatial step size is  $dx = 0.074826$  and 40 bins have been used. The drift and diffusion are given in Figures 2.33 and 2.34, respectively. The linear fit and the 95% confidence bounds can't be obtained for the drift. For the diffusion, a quadratic polynomial,  $p1 * x^2 + p2 * x + p3$ , was fitted and except the intercept,  $p3$ , the coefficients are significant with  $p1 = -1.4898 \pm 0.9608$ ,  $p2 = 4.9625 \pm 2.9611$ . Note that the true values of  $p1$  and  $p2$  are -1 and 3, respectively. Like the CIR process, only the general shape of the diffusion was recovered by the method.

## 2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

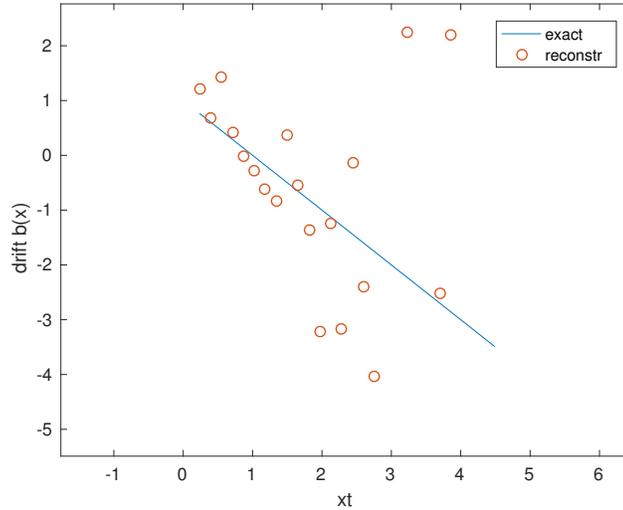


**Figure 2.29:** Reconstruction of the drift of an OU process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $b(x)$  are given. The true values are given by the line and estimates by the circles. Large approximation errors appear on the large positive values of the process

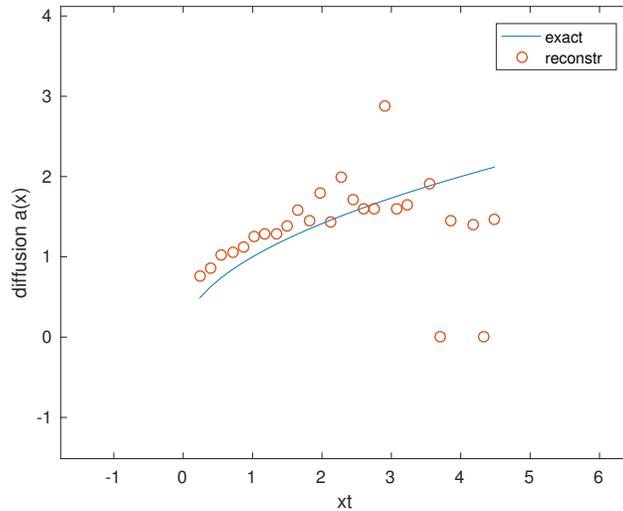


**Figure 2.30:** Reconstruction of the diffusion of an OU process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $a(x)$  are given. The true values are given by the line and estimates by the circles. Large approximation errors appear at the extreme values of the process

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS

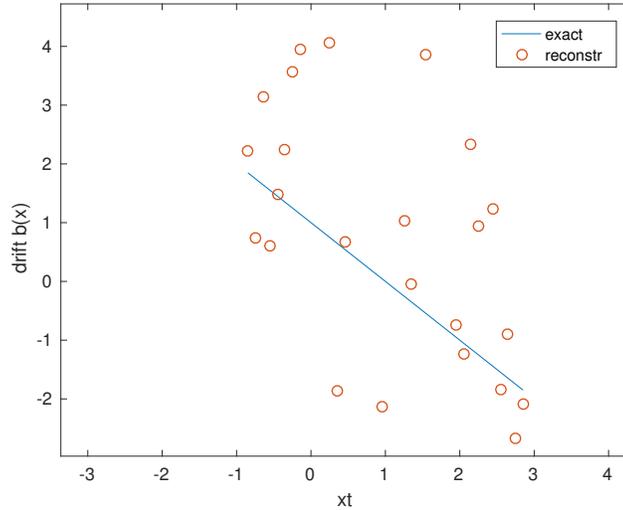


**Figure 2.31:** Reconstruction of the drift of a CIR process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $b(x)$  are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process.

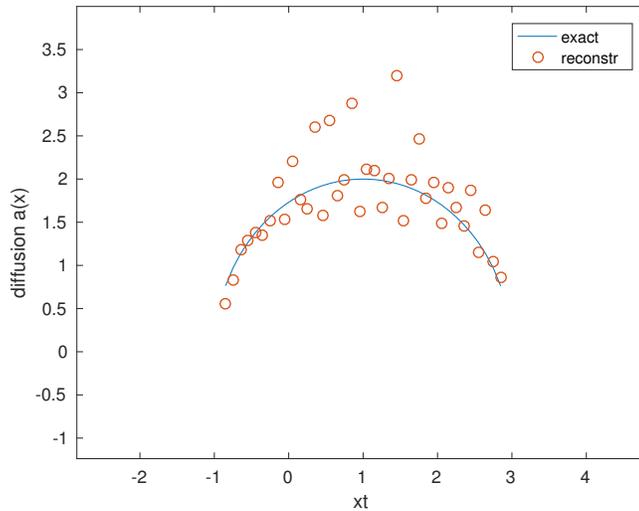


**Figure 2.32:** Reconstruction of the drift of a CIR process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $a(x)$  are given. The true values are given by the line and estimates by the circles. Diffusion for small values of the process are better approximated than for large values.

2.6. EXTENSION TO NON PERIODIC DRIFT AND DIFFUSION FUNCTIONS



**Figure 2.33:** Reconstruction of the drift of a Jacobi process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $b(x)$  are given. The true values are given by the line and estimates by the circles. These parameters fail to provide reconstruction.



**Figure 2.34:** Reconstruction of the diffusion of a Jacobi process using simulated data using  $T = 10^6$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas on the y-axis values of  $a(x)$  are given. The true values are given by the line and estimates by the circles. The shape of the diffusion is recovered but outliers appear when the process takes values around the long-term mean.

**Conclusion of this subsection:** Unlike the case of periodic functions, the reconstruction procedure using simulated data, and no filtering, fails to recover the true drift and diffusion. When a polynomial fit is possible, the 95% confidence intervals of the coefficients are relatively large. One reason could be the presence of outliers in the estimates of the drift and diffusion. As an example, approximation errors prevent a fit for the drift in the case of the JAC. The diffusion functional seen for all 3 processes is easier to fit than the drift.

## 2.7 Improved reconstruction

Without filtering, the spectral reconstruction fails to provide robust estimates of the non-periodic drift and diffusion coefficients. Numerically, there are many potential sources of errors, the transition probabilities, the eigenvalues and eigenfunctions, or the optimization. In this section, we study how the selection of parameters  $(T, h, M)$  could improve the reconstruction.

### 2.7.1 Large data sets

We check if an increase of the number of observations could significantly improve the reconstruction. We consider the parameters: subsampling time,  $h = 0.01$ , simulation time,  $dt = 10^{-4}$ , and the number of bins  $M$  is set to 40. This generates 10 times more data than before. The number of observations simulated was  $T = 10^7$ .

### **OU process**

For the OU process, note that the spatial step size is  $dx = 0.12854$ . The drift and diffusion are given in the Figures (2.35) and (2.36), respectively. The linear fit and the 95% confidence bounds give the drift with a slope of  $-2.6448 \pm 1.8666$  and a non significant constant i.e.  $p2 = 0$ . The slope is larger than its true value and the confidence interval is larger than with less data. A quadratic fit  $p1 * x^2 + p2 * x + p3$  for the diffusion gives  $p1 = p2 = 0$  and a constant term  $p3 = 1.242 \pm 0.0889$ . The constant diffusion is correctly recovered and close to the real value which is 1. When we increase further the number of sample points, 5 to 10 times the current value, the slope of the drift term is stiffer and for the diffusion estimate shows larger errors when the process takes large values.

### **CIR process**

The spatial step used with 40 bins used is  $dx = 0.14586$ . Even with the eigenvalues closed to the true one, eigenvalues are  $\{\lambda_1, \lambda_2, \lambda_3\} = \{1.1102e-13, -1.3421-2.7205\}$ , we found no significant fit for both the drift and diffusion. Again half of the diffusion is recovered but the presence of large errors when the process takes large values prevent a polynomial fit. We increase  $T$  further, multiplied it by 5, 10 ,20 times its value but found small improvements for the CIR reconstruction. When the process reaches large values, the errors continue to increase. The reconstructed drift and diffusion are shown in Figures (2.37) and (2.38).

### **Jacobi process**

For the Jacobi process, the spatial step size is  $dx = 0.074992$  and 40 bins have been used. The drift and diffusion are given in the Figures (2.39) and (2.40), respectively.

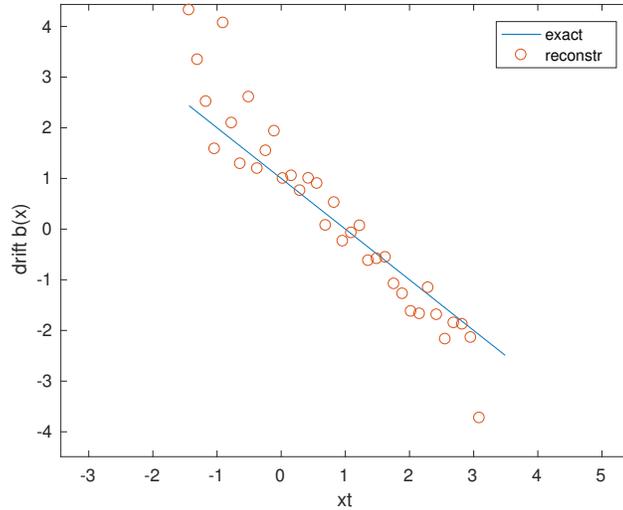
## 2.7. IMPROVED RECONSTRUCTION

---

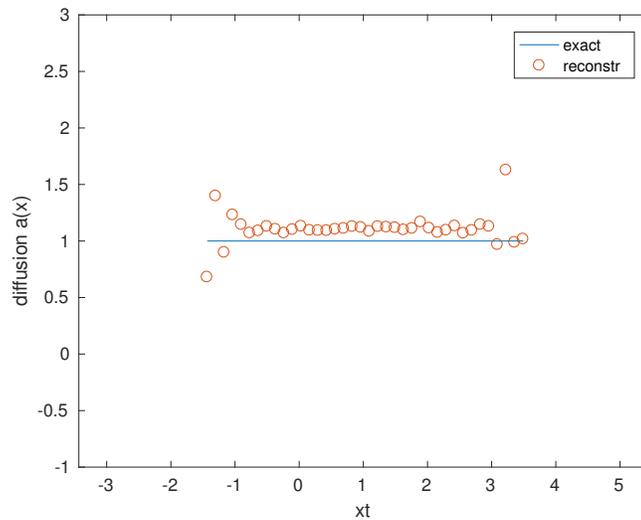
The linear fit and the 95% confidence bounds yields a drift with  $p1 = -0.9855 \pm 0.7325$  and a non significant intercept. Note that  $p1$  is closed to its true value of -1 and allows part of the drift to be recovered. Again a quadratic fit  $p1 * x^2 + p2 * x + p3$ , for the diffusion gives  $p1 = -1.0482 \pm 0.1828$ ,  $p2 = 3.1617 \pm 0.5644$  and no significant constant, i.e.  $p3 = 0$ . The significant coefficients are close to their true values -1 and 3, respectively. The Jacobi process drift and diffusion estimates improve when the number of observations 5, 10, or 20 times larger. We found a perfect match to the true value when the number of observations is greater than  $5 \times 10^7$ .

## 2.7. IMPROVED RECONSTRUCTION

---



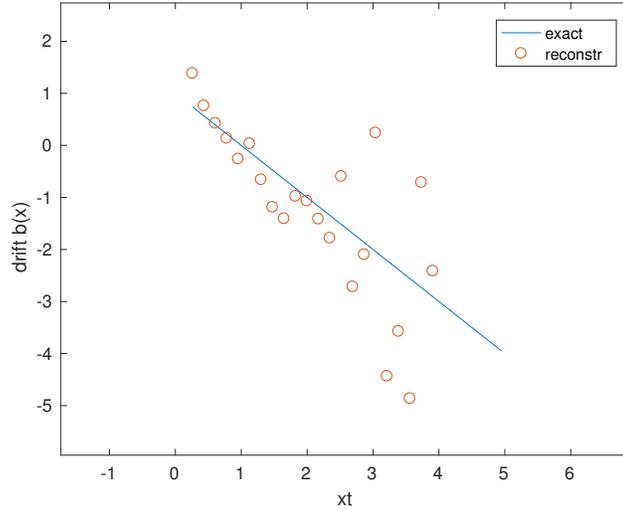
**Figure 2.35:** Reconstruction of the drift of an OU process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. The parameters provide an overestimated reconstruction. Large errors appear when the process takes extreme values.



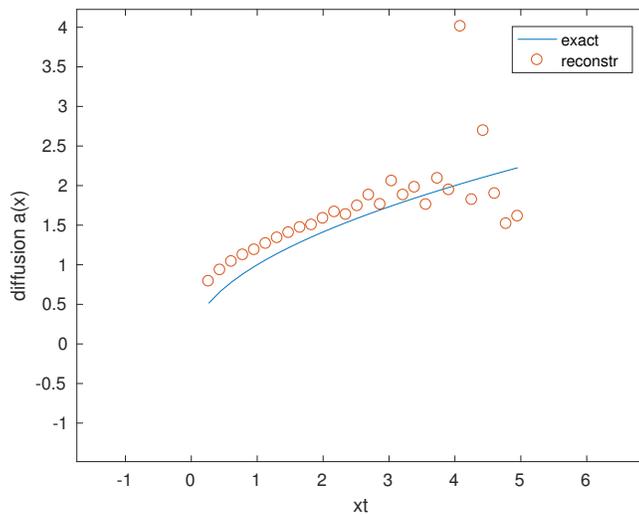
**Figure 2.36:** Reconstruction of the diffusion of an OU process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. The parameters provide an overestimated reconstruction of the diffusion. Large errors appear when the process takes extreme values.

## 2.7. IMPROVED RECONSTRUCTION

---



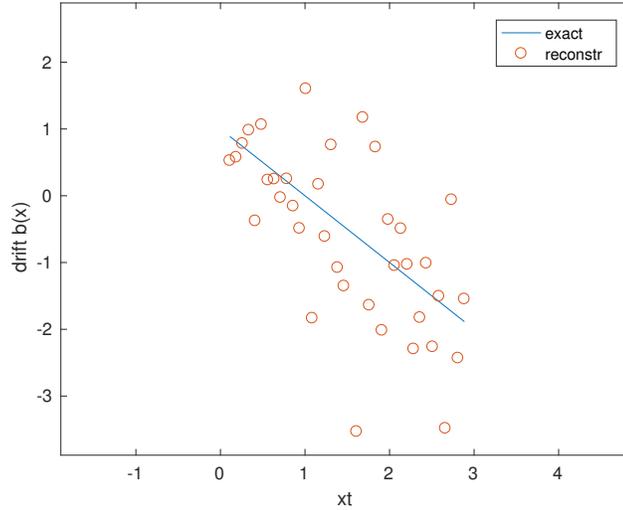
**Figure 2.37:** Reconstruction of the drift of a CIR process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 30$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process.



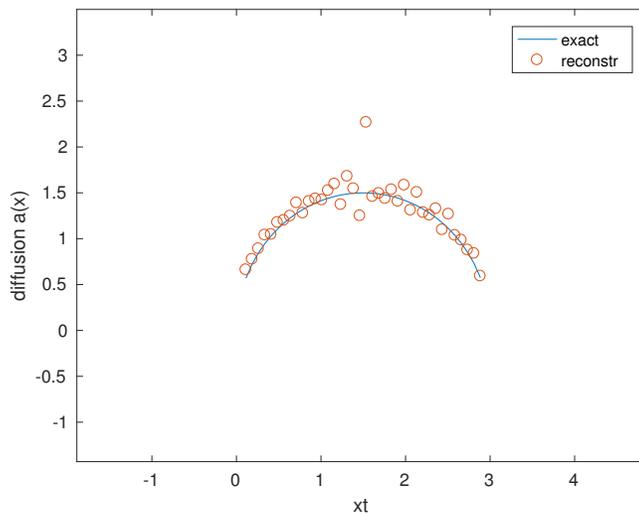
**Figure 2.38:** Reconstruction of the diffusion of a CIR process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 30$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. Large estimation errors could be found for large values taken by the process.

## 2.7. IMPROVED RECONSTRUCTION

---



**Figure 2.39:** Reconstruction of the drift of a Jacobi process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed drift are given. The true values are given by the line and estimates by the circles. Large errors could be found for any value of the process.



**Figure 2.40:** Reconstruction of the drift of a Jacobi process using simulated data using  $T = 10^7$ ,  $dt = 10^{-4}$ ,  $h = 0.01$ ,  $M = 40$ . The values taken by the process are given by  $xt$  on the x-axis. On the y-axis, the values of the reconstructed diffusion are given. The true values are given by the line and estimates by the circles. Large errors could be found for values close to the long-term mean.

**Conclusion of this subsection:** We tested if the increase in the number of sample points would reduce the approximation errors and improve the reconstruction of the drift and diffusion. The drift estimates for the OU process with an increase of data has a lot steeper slope while the errors in the diffusion estimates increase when the process takes extreme values. In summary, for the OU process, more data causes an overestimation of the drift and diffusion. The drift of the JAC process is partially recovered while the diffusion is fully reconstructed. The drift and diffusion of the CIR process could not be reconstructed even when the number sample was of the order  $10^8$ .

### 2.7.2 Errors with subsampling and spatial steps

We generated 100 paths using different values of  $h$  and numbers of bins  $M$ . For each process, we test how the choice of these parameters affects the estimation error that we define as the 2-norm of the difference between the true value and the estimated drift and diffusion at each discretization point. More formally we consider the following:  $\forall i = 1, \dots, M$ :

$$error_{1,drift} = \|b_i^{true} - \hat{b}_i\|_2, \quad error_{1,diffusion} = \|a_i^{true} - \hat{a}_i\|_2, \quad (2.44)$$

where  $b_i^{true}$  and  $a_i^{true}$  are the true values of the drift and diffusion at the point  $x_i$ , respectively. The corresponding estimated values are given by  $\hat{b}_i$  and  $\hat{a}_i$ . For each  $i$ , we can compute a vector of size  $100 \times 1$ . Large errors occur in the estimation especially at the extreme values taken by the process. We tried a second type of

## 2.7. IMPROVED RECONSTRUCTION

---

Nb. of bins(M)	Error Drift	Error Diff
20	7.3938	1.9578
40	4.1169	0.8256
60	2.292*10 <sup>9</sup>	8.3687
80	5.827*10 <sup>12</sup>	1.2620
100	4.9765*10 <sup>11</sup>	170.22

**Table 2.12:** Estimation errors obtained by varying the number of bins in the case of OU process

errors

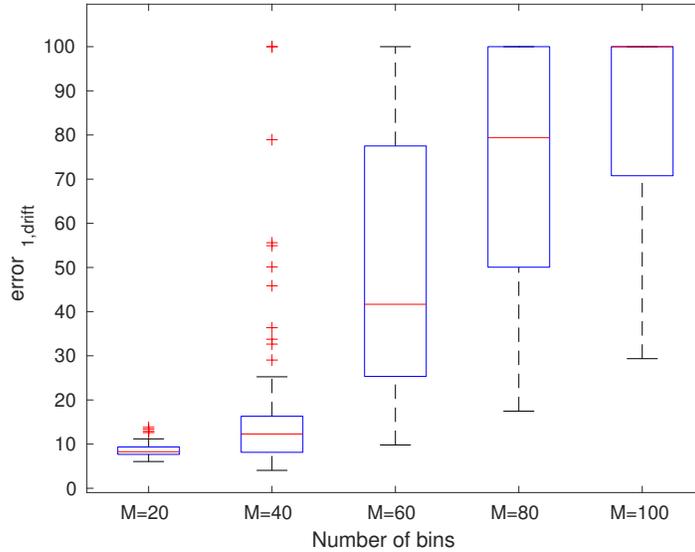
$$error_{2,drift} = \|b_i^{true} - \bar{b}_i\|_2, \quad error_{2,diffusion} = \|a_i^{true} - \bar{a}_i\|_2, \quad (2.45)$$

where  $\bar{b}_i = \frac{1}{100} \sum_{i=1}^{100} b_i$  and similarly  $\bar{a}_i = \frac{1}{100} \sum_{i=1}^{100} a_i$ . For each  $i$ , this error of type 2 gives one value. The reconstructed values at the extreme of the range of simulated data still affect this second type of errors. In addition to this difficulty, the calculation of errors is complicated by the fact that all bins need to have at least one data point. Not all randomly generated paths satisfy this condition. After multiple trials, we came to the conclusion that the values of  $h \leq 0.05$  produce fewer errors. Subsampling with this observational time step produced the best results. This gives us the following relationship between the simulation time and the subsampling time:  $h = \alpha dt$  with  $1 < \alpha \leq 10$ . The best values found were  $(dt, h) = (10^{-3}, 0.01)$ . Given these, we focused on the number of bins. We used the errors defined above in (2.45) and sum them for a given number of bins,  $M$ . This gave an overall value for the total error in the estimation of the drift and the diffusion that can be compared for different numbers of bins,  $M$ . The sum of the error made for different numbers of bins for the OU process is seen in Table 2.12. Estimation of the other processes show

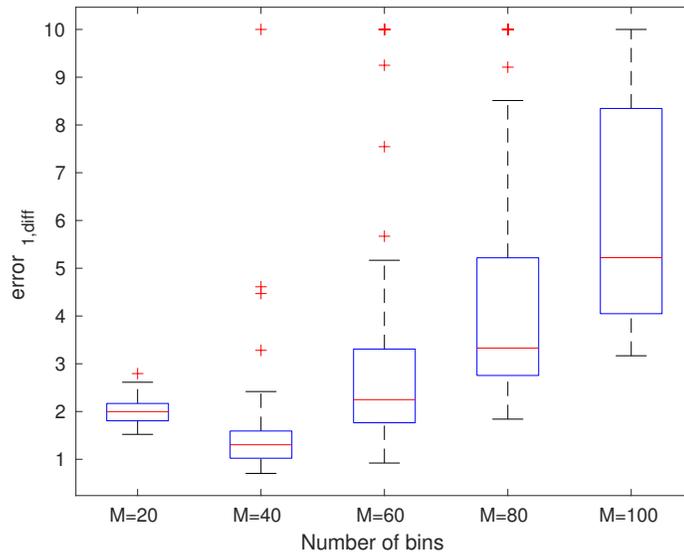
## 2.7. IMPROVED RECONSTRUCTION

---

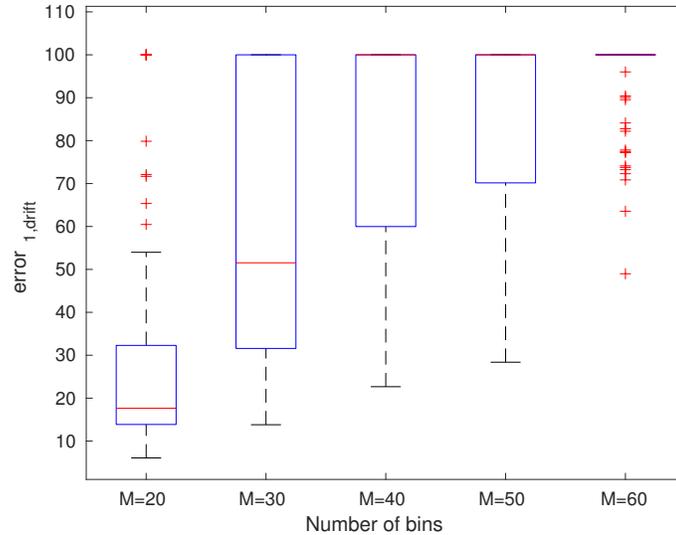
the same behavior seen in Table 2.12, the errors are larger when the number of bins increases. The best number of bins appear to be approximately 40. Sometimes a smaller number of bins is more appropriate for the CIR. For this process,  $M = 30$  also produces good results. To illustrate more visually how the estimation errors depend on the number of bins, we constructed box plots that show for a given number of bins the error defined by (2.44) for 100 paths. To reduce the effect of outliers on the plots, we just set for the drift that any error above 100 would be considered as an outlier and set to be equal to 100. For the diffusion, we use a threshold of 10. The following box plots, Figures 2.41 to 2.46, show that when the number of bins is greater than 40, the errors increased significantly for OU and JAC. This threshold is approximately equal to 30 for CIR. Note that the reconstruction of the components of JAC with 40 bins produces fewer errors and almost no outliers. The drift and diffusion reconstruction with more than 20 bins for the CIR process generated large errors for both the drift and the diffusion.



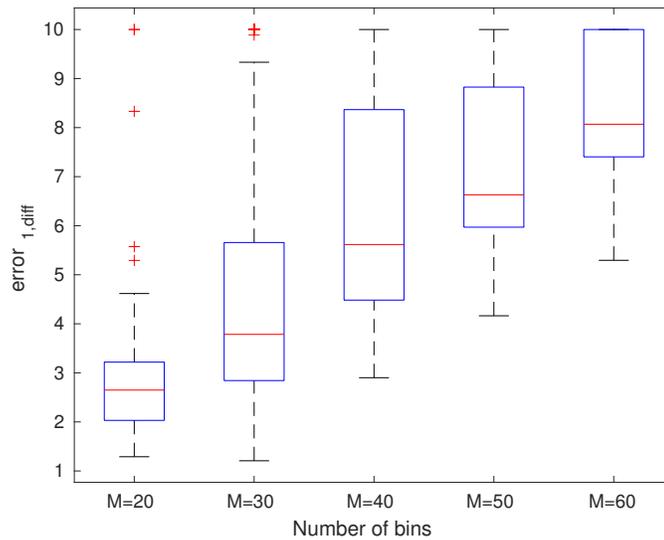
**Figure 2.41:** The number of bins for the OU process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. Number of bins between 20 and 40 provides less errors. When the number of bins is greater than 40, more outliers appear.



**Figure 2.42:** The number of bins for the OU process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set to 10. Observe that 40 bins give the smallest estimation errors for the diffusion.



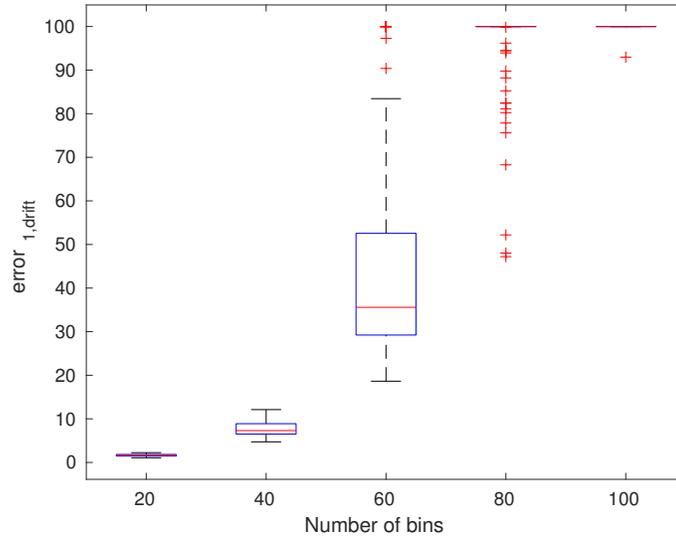
**Figure 2.43:** The number of bins for the CIR process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. Observe that compared to the OU process, the magnitude of errors made in the drift estimation is at least twice higher. The median is multiplied by at least 2 just by increasing the number of bins from 20 to 30.



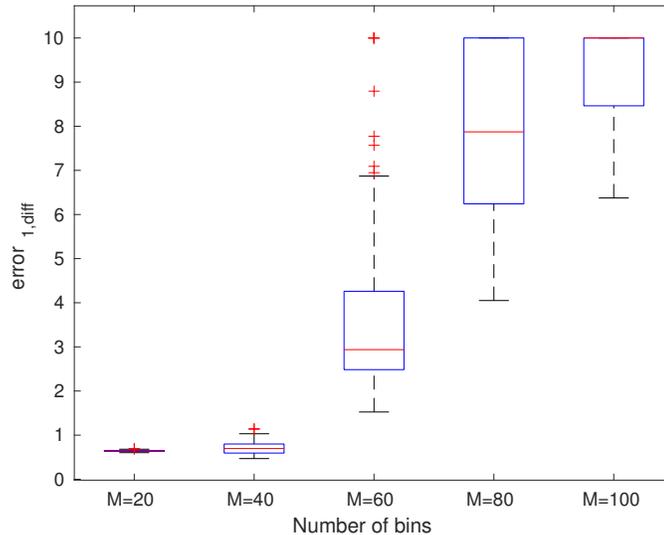
**Figure 2.44:** The number of bins for the CIR process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set to 10. The estimation errors are also larger than those of the OU. Small number of bins, between 20 and 40, gives less errors.

## 2.7. IMPROVED RECONSTRUCTION

---



**Figure 2.45:** The number of bins for the JAC process is given on the x-axis and the 2-norm of the difference between estimate and true drift on the y-axis. All values greater or equal to 100 are set to 100. The error value, when the number of bins is between 20 and 40, is smaller than 20. The error increases fourfold when the number of bins is greater than 40.



**Figure 2.46:** The number of bins for the JAC process is given on the x-axis and the 2-norm of the difference between estimate and true diffusion on the y-axis. All values greater or equal to 10 are set 10. Same pattern than the drift, a number of bins between 20 and 40 gives a better approximation.

**Conclusion of this subsection:** We found empirically that subsampling with  $h \leq 0.01$  improves the reconstruction while sampling less than 10 produces the best results. A large number of bins, and equivalently a small spatial step size, produces large errors. A number close to 40 bins appears to be optimal for the OU and JAC. Reconstruction of the drift and diffusion of the CIR generates the largest approximation errors. As suggested by the Figures 2.43 and 2.44, one way to reduce the estimation errors is to use a number of bins such that  $20 \leq M \leq 30$ .

### 2.7.2.1 Truncation of the outliers

The estimation seems to be worse at the extreme values of the process. This is expected since the process would visit these extreme states only a few number of times. Therefore, one way to improve the estimation is to only consider states that are most frequently visited by the process. For a mean reverting process as the one we have, this means that the estimation has to be restricted to the area close to the long term average. One simple way to restrict our estimation to the above-mentioned region is to truncate the transition probability matrix constructed from data. Hence, the approach that we chose is to look at the histogram of the data and to truncate bins with a low amount of points. The OU and JAC processes have a symmetric stationary distribution so both the leftmost and rightmost bins are truncated. The CIR process has a stationary density that follows a Chi-square distribution. For this process, more bins can be removed on the right tail than the left. Simulation parameters remain the same. The number of bins,  $M$ , is set to 40, the subsampling time,  $h = 0.01$ , and the number of observations  $T = 10^6$ .

## 2.7. IMPROVED RECONSTRUCTION

---

We include in the estimation observations in an interval sufficiently close to the mean. We introduce the following notations: the mean of the process is denoted by  $\mu_X$  and the variance  $\sigma_X$ . We found for each process cutoff points, denoted by  $c_L$  for the left tail cutoff value and  $c_R$  for the right one, and remove bins (same as states in the transition probability matrix) that are not in the interval  $[\mu_X - c_L * \sigma_X, \mu_X + c_R * \sigma_X]$ . The Jacobi process is simulated with a truncated scheme, so we didn't apply any additional truncation. The truncation significantly impacts the eigenvalues and eigenfunctions. The eigenfunctions are modified but remain close to their reference values before truncation. The first eigenvalues remained close to the true ones as shown in Table 2.13. The second and third eigenvalues after truncation are smaller than the theoretical ones. The eigenvalues are related to the weights. Smaller value in this case implies a small weight. To assess the reconstruction with

	True value	OU	CIR	JAC
$\lambda_1$	0	-1.7764e-13	-3.7007e-14	-1.1102e-13
$\lambda_2$	-1	-1.4100	-1.7317	-1.3010
$\lambda_3$	-2 (-3 for JAC)	-3.7508	-5.0499	-4.0999

**Table 2.13:** The eigenvalues after truncation of the extreme values the three affine processes are compared to their real value.

different subsampling times we computed error estimates using the definition (2.45). Subsampling at most every 10 observations gives the best reconstruction as shown in Table 2.14. We also plot the result of this table in Figures 2.47 and 2.48. One can clearly see that when  $h > 0.02$ , the errors increase for all processes. If  $h$  is small, i.e. ( $h < 0.003$ ), the errors, especially for the diffusion, also increases.

## 2.7. IMPROVED RECONSTRUCTION

$h$	OU		CIR		JAC	
	$error_{2,drift}$	$error_{2,diff}$	$error_{2,drift}$	$error_{2,diff}$	$error_{2,drift}$	$error_{2,diff}$
0.1	53.176	3.0154	977.24	2.5409	29.982	4.3950
0.05	62.299	1.7263	15.839	2.1124	22.874	1.9512
0.01	2.7779	0.2703	2.6650	0.2523	0.6659	0.3224
0.005	3.3931	0.4709	5.6534	0.4052	0.7565	0.5096
0.003	3.4695	0.7657	5.6758	0.7189	1.1344	0.8044
0.001	5.3535	2.1012	7.3795	1.9892	3.3086	2.0984

**Table 2.14:** This table gives the error corresponding to different subsampling time after applying a truncation of some values of the process histogram.

We use Table 2.14 to select the subsampling time for the reconstructions. We present the best cases for each process.

### OU process

The extreme values of the process are the ones that we want to eliminate. Empirically we found that  $c_L = c_R = 2$ . We focus the estimation on states in the interval  $[\mu_X - 2\sigma_X, \mu_X + 2\sigma_X]$ . This implies that around 5% of the data has been removed. The best results are obtained by subsampling every 10 observations. The Figures 2.49 and 2.50 show the results. The parameters are  $h = 0.01$ ,  $M = 40$ ,  $\Delta x = 0.074516$  and  $T = 10^6$ . The drift has a slope  $-1.17 \pm 0.312$  and a constant term  $1.149 \pm 0.397$ . For the diffusion,  $p1 = 1.076 \pm 0.065$  and  $p2 = p3 = 0$ . Apart of providing statistically significant coefficients close to their theoretical values one could notice that the confidence intervals are smaller than without truncation.

### CIR process

The best results are obtained when we sub-sample between every third and tenth data, i.e  $h \in [0.003, 0.01]$ . For illustration we present the case where  $h = 0.003$ ,

## 2.7. IMPROVED RECONSTRUCTION

---

$M = 30$ ,  $\Delta x = 0.0575$  and  $T = 10^6$ . The right tail should have a larger cut  $c_L = 1.5$  and  $c_R = 2$  so our interval is  $[\mu_X - 1.5\sigma_X, \mu_X + 2\sigma_X]$ . Using this interval removes 5% of the total number of observations, so we are left with most of the data. One can see in Figures 2.51 and 2.52 that the reconstruction has improved but still fails to recover the drift at the extreme values of the process. The drift has a slope  $-1.95 \pm 0.695$  and a constant term  $1.909 \pm 0.914$ . These values are almost twice the true values, both are 1, but have the correct sign. The quadratic fit  $p1 * x^2 + p2 * x + p3$  for the diffusion whose coefficients with 95% bounds for the diffusion is  $p1 = p3 = 0$ ; and  $p2 = 1.2795 \pm 0.3541$ . So the diffusion is a linear function of the process value as expected.

### Jacobi process

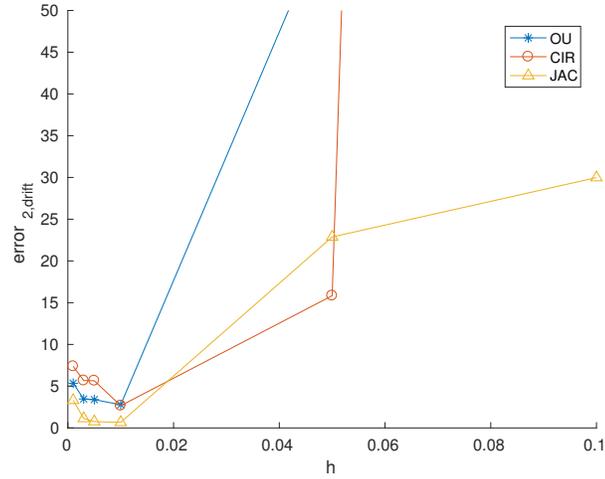
The Jacobi process data obtained from a truncated Euler scheme, requires no additional truncation since the extreme states have been removed during the data simulation. We therefore look at different subsampling times in the entire interval. Similar to the CIR process, good reconstructions are obtained when  $h \in [0.003, 0.01]$ . We present the results for  $h = 0.003$ ,  $M = 40$ ,  $\Delta x = 0.0074995$  and  $T = 10^6$  in Figures 2.53 and 2.54. We obtained: a drift with a slope  $-0.857 \pm 0.267$  and a constant term  $0.753 \pm 0.456$ . Both values are different than -1 and 1 respectively their true values. But the coefficients are statistically significant and relatively close compared to the case without truncation where no relationship could be found. The quadratic polynomial  $p1 * x^2 + p2 * x + p3$  whose coefficients with 95% bounds for the diffusion is  $p1 = -1.0145 \pm 0.0395$ ;  $p2 = 3.0367 \pm 0.1221$  and  $p3 = 0.2971 \pm 0.0796$ . the coefficients  $p1$  and  $p2$  match their true values -1 and 3 respectively. An extra constant

term is found.

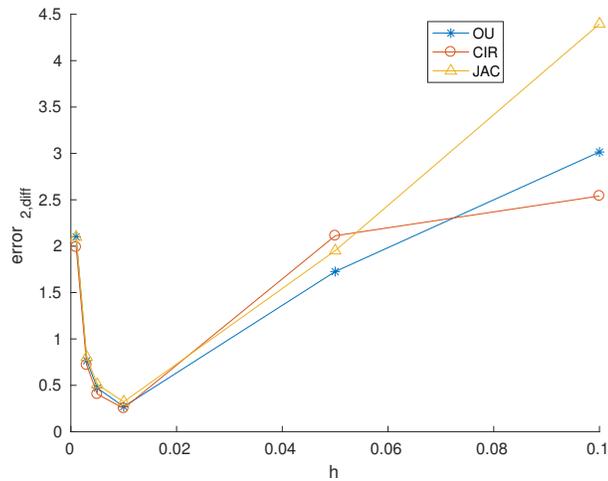
**Conclusion of this subsection:** By eliminating the rows and columns of the transition probability matrix which correspond to the tails of the stationary distribution we successfully remove the effect outliers have on the reconstruction and the polynomial fit on the estimates. We accomplished this truncation of the transition probability matrix by restricting our reconstruction on sub-interval close to the mean of the process,  $[\mu_X - c_L * \sigma_X, \mu_X + c_R * \sigma_X]$  where  $c_L$  and  $c_R$  are constants that depend on the process. The drift and diffusion of the OU and JAC processes are reconstructed much better than without truncation. The diffusion of the CIR process is fully reconstructed and its drift can only be partially reconstructed.

## 2.7. IMPROVED RECONSTRUCTION

---



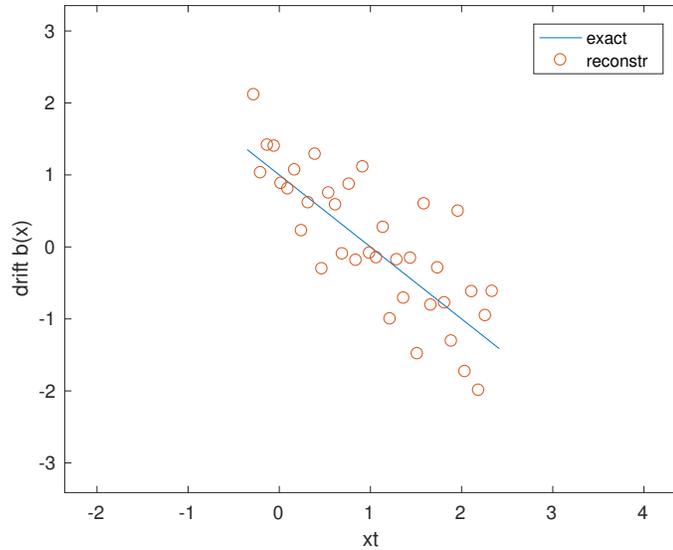
**Figure 2.47:** The  $error_{2,drift}$  for different subsampling times is given for the three affine processes after some values have been truncated. On the x-axis the time interval between consecutive values of the process, the subsampling time,  $h$ , is given. On the y-axis the values of the error  $error_{2,drift}$  are provided. We can see that the error on the drift of the Jacobi is smaller than the one of the other processes. Around  $h = 0.01$  the approximation error of the drift is minimal for the processes.



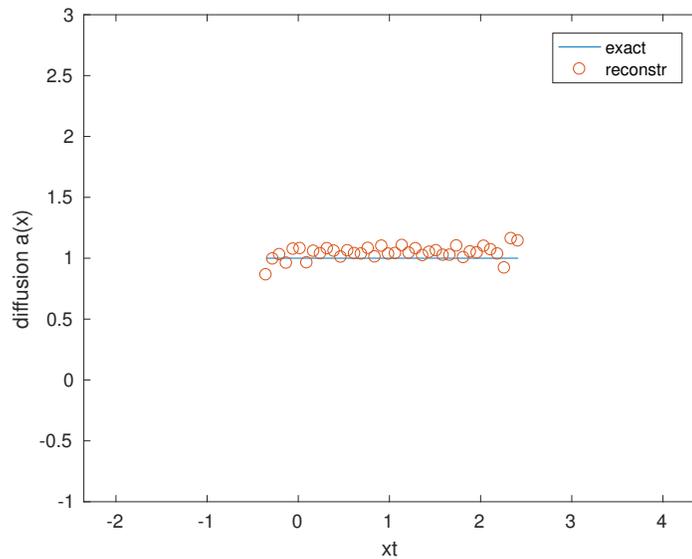
**Figure 2.48:** The  $error_{2,diff}$  for different subsampling times is given for the 3 affine processes after some values have been truncated. On the x-axis the time interval between consecutive values of the process, the subsampling time,  $h$ , is given. On the y-axis the values of the error  $error_{2,diff}$  are provided. We can see that the diffusion is better reconstructed for all 3 process when the subsampling time is close to 0.01. However when the subsampling time,  $h$ , is large, the Jacobi process generates the largest estimation errors for the diffusion.

## 2.7. IMPROVED RECONSTRUCTION

---



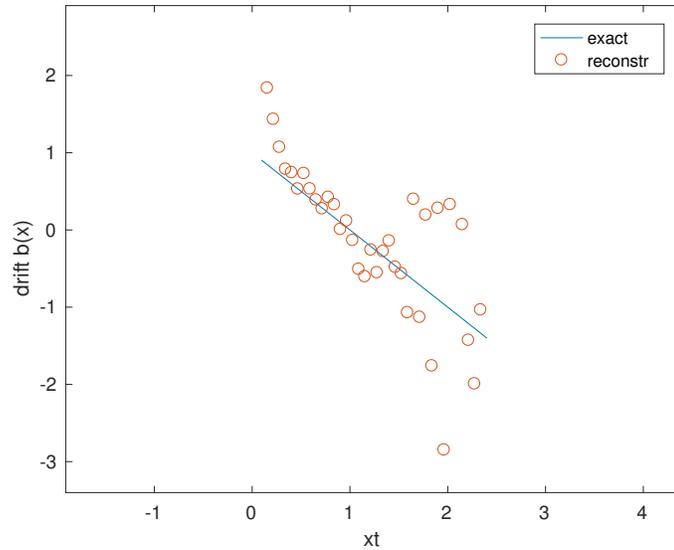
**Figure 2.49:** Reconstruction of the drift of an OU process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.



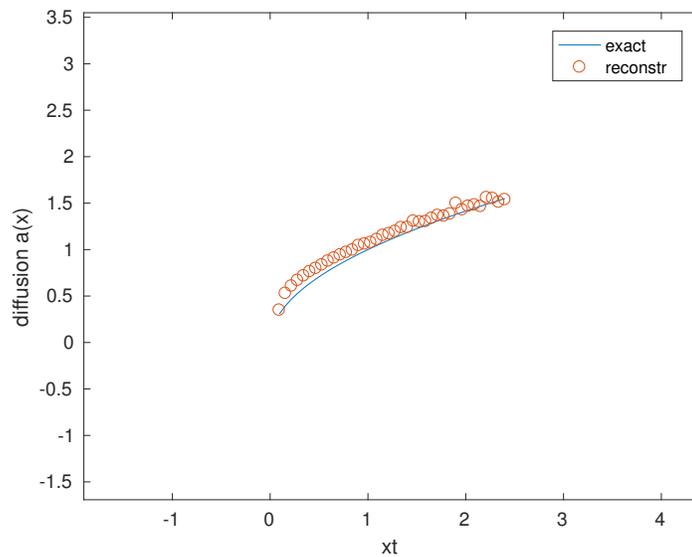
**Figure 2.50:** Reconstruction of the diffusion of an OU process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.01$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.

## 2.7. IMPROVED RECONSTRUCTION

---



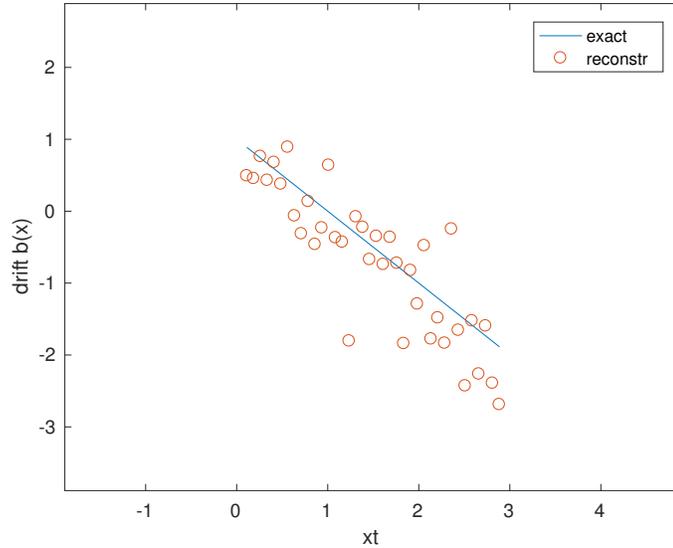
**Figure 2.51:** Reconstruction of the drift of a CIR process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.



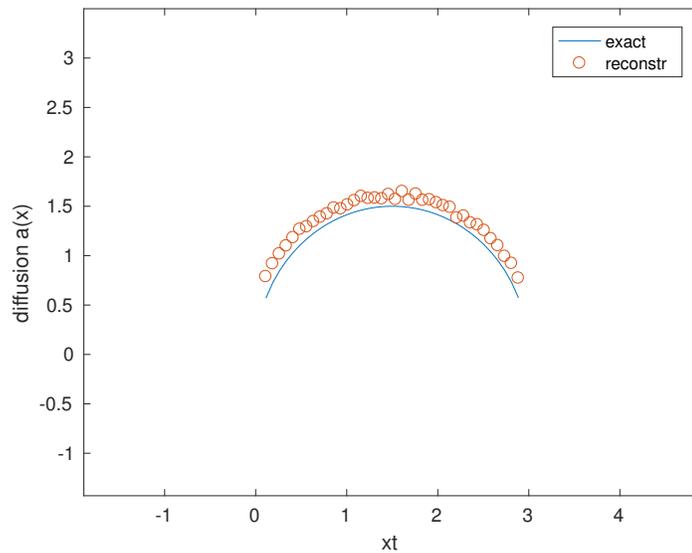
**Figure 2.52:** Reconstruction of the diffusion of a CIR process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.

## 2.7. IMPROVED RECONSTRUCTION

---



**Figure 2.53:** Reconstruction of the drift of a Jacobi process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $b(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.



**Figure 2.54:** Reconstruction of the diffusion of a Jacobi process using truncated histogram  $T = 10^6$ ,  $dt = 10^{-3}$ ,  $h = 0.003$ ,  $M = 40$ . The x-axis gives different values,  $xt$ , taken by the process whereas values of  $a(x)$  are given on the y-axis. The true values are given by the line and estimates by the circles.

## 2.8 Application: Modeling energy commodity prices dynamics

### 2.8.1 Motivation

Crude oil is the most commonly traded energy commodity in financial markets and it is considered a fundamental driver of most countries' macroeconomic performance. The recent fall of crude oil price has left the financial community stunned. From roughly \$ 100 per barrel prior to mid 2014, the price of oil in 2017 fluctuates around \$ 50, half of what it was a little more than three years ago. The important role in economic growth plays by this fossil fuel motivated various researchers to model its market price dynamics. Stochastic Differential Equations offer a simple and easy way to describe the behavior of the price of crude oil. Unlike the Geometric Brownian motion assumption, generally found in stock price models [8, 9], researchers commonly agree that commodities in general and crude oil in particular, should be modeled by a mean reverting process [28, 70, 77]. This mean-reversion feature of the oil price has been the object of different studies which tries to established its validity. The implications for forecasting are important: if mean reversion is observed, crude oil price could be considered predictable. On the other hand, if the price is assumed to follow a random walk type of dynamics, then in line with theories such as the famous efficient market hypothesis [25, 55] the price could be considered as unpredictable.

## 2.8.2 Literature review

The heart of the debate could be summarized by the following question *"Is Crude oil price mean reverting or a random walk?"*. This is not a new question. Geman [29] proposes two statistical tests of mean reversion in Oil and Natural gas prices. She used two well-known stationarity tests [40], Augmented Dickey Fuller (ADF) and Phillips Perron (PP) and reach the conclusion that the answer depends on the time period of interest. The time interval covered by her analysis went from January 1994 to October 2004. In this period, she found that the crude oil price mean-reverts from 1994 to 2000 and then the process became a random walk from 2000 to 2004. For natural gas, the same pattern could be observed, prices mean-reverts until 1999 and then behaves like a random walk. Skorodumov [71] also tested mean reversion in both oil and gas markets using historical data from 1990 to 2008. His test is based on the statistical significance of the estimated coefficients of an OU process. He concluded that for 10 years, 1991, 1994-1998, 2000, and 2003-2005, the crude oil price exhibits significant mean reversion coefficients. Chikobvu et al.(2013) [15] also investigated whether crude oil price followed a random walk or a mean reverting path. They used stationary tests like Geman [29,30] and a GARCH model with time varying coefficients. The data set comprised of monthly prices over the period of January 1980, to September, 2010. The ADF test and the GARCH model show significant mean reversion between 1980 and 1994 and random walk from 1994 to 2010. Therefore, all these studies showed that the price dynamics is constantly changing from mean reverting to random walk. The goal of this section is to use the spectral reconstruction to estimate the drift and diffusion of the crude oil price

process. A constant drift would suggest that the oil price behaves like a random walk when a linear one would be in favor of the mean reverting hypothesis. Apart of bringing more evidence on the shape of the drift, another contribution of this work is to identify the functional form of the diffusion process, what have not been considered in the previous studies.

### 2.8.3 Description of crude oil price data

There are different grades of crude oil based on the sulfur content and the density. A crude oil with low sulfur content would be qualified as sweet while a low density fuel would be referred as light. Different benchmarks to describe oil exist. The West Texas Intermediate (hereafter WTI) is a high grade crude oil, also known as the Texas light sweet, used primarily in the USA and mostly refined in the Midwest and Gulf Coast region of the country. The second popular benchmark is the Brent crude extracted in the northern sea. WTI and Brent are considered worldwide as benchmark for oil prices. In this work, we use WTI. The prices used in this study are part of the free data available in the website [www.histdata.com](http://www.histdata.com). It comprises 1 minute intervals of data of the main crude oil index, WTI, covering the time period from November 14<sup>th</sup> 2010 to December 31<sup>th</sup> 2016. The total number of data available from the website for this period is 1965766 observations. Note that within some trading days some quotations are missing, or values for entire days are just not reported. The presence of gaps in the data implies that the total number of observations available for a year varies. In the financial markets, Out of 365 days per year, almost a dozen days are labeled as market holidays, which leaves us with

## 2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---

approximately 350 full trading days. But in these data a regular year consists in 310 or 311 trading days at the most. Another point is that for each day not all data are reported. In the Table 2.15, the number of dates available per year are provided as well as for each year the total number of observations. For each 1-minute interval during the day trading session, the open bid quote, high bid quote, low bid quote, close bid quote are recorded. The crude oil price is then computed using an average of the open and close bid quotes.

The evolution of the WTI from 2010 to 2016 is presented in Figure 2.55. The oil extraction in the last two decades has been marked by a widespread use of Hydraulic fracturing also known as fracking, a stimulation technique that allowed to access the oil and gas in the shales reserves. Fracking created a boom in the oil supply. This increase in oil production in the USA, Canada but also China led to a global drop of the price of a barrel. We can easily distinguish in Figure 2.55, 3 potential state regimes in the crude oil markets for the last 7 years. Expressed in terms of years, we can see: a period 1 of high oil price from 2010 to 2013, a period 2 that covers the year 2014 that seems to be a year of transition from a high to low price regime. Finally the last period 3, 2015 to 2016, could be interpreted as the low oil price regime. The graphs for each regime are given in the Figures 2.56, 2.57 and 2.58. We also provide in Table 2.16 some summary statistics of the data. The average price for each period confirm the presence of high and low price regimes.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES  
DYNAMICS

---

	2010	2011	2012	2013	2014	2015	2016
Nb of dates	42	310	311	311	311	311	310
Nb of data	41143	318466	333544	320814	308208	326253	317338

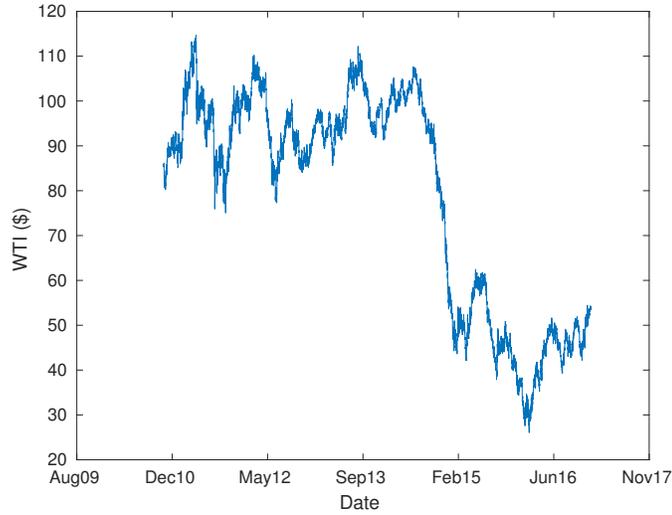
**Table 2.15:** Description of WTI 1-minute data set. Note that not everyday is reported and within a day the price for some minutes is missing. In this data set, a regular year appears to have between 310 and 311 days of quotations.

	Period 1	Period 2	Period 3	Period 1+2+3
Mean	95.4122	92.2897	46.087	78.773
Stand. dev.	7.3968	13.5881	7.2932	24.409
Skew	0.0637	-1.3931	-0.2293	-0.5748
Kurtosis	2.3871	4.0217	3.0853	1.7235
Minimum	75.040	52.470	26.040	26.040
Maximum	114.71	107.640	62.470	114.71

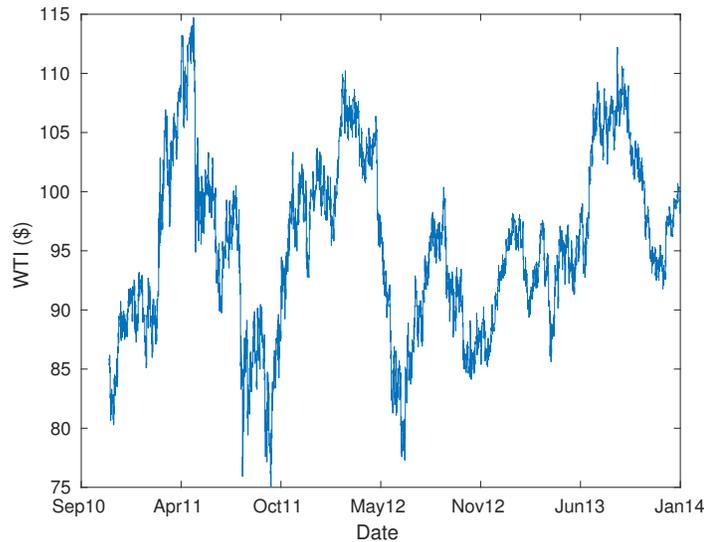
**Table 2.16:** Descriptive statistics of the WTI 1-minute dataset. The mean of each period shows three possible regimes in the last seven years. Period 1, 2010-2013, is a high oil price regime. Period 2, 2014, is a transition year during which the price of a barrel dropped. It was also a year of high uncertainty as shown by the standard deviation twice higher than during the two other periods. The Period 3, 2015-2016, is a low price regime with the average price per barrel less than half what it was in period 1.

## 2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---



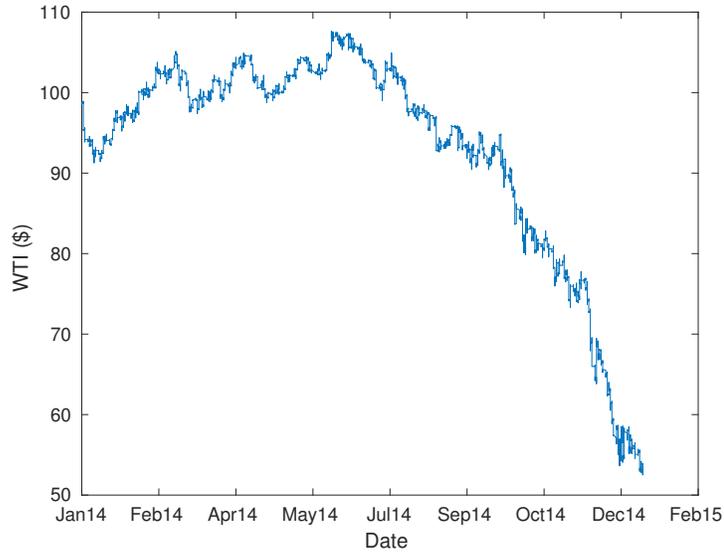
**Figure 2.55:** WTI 1-minute evolution from November 2010 to December 2016. The value of the benchmark is given by the average of the open and close bid at each minute. Graphically, one can distinguish 3 different regimes: the regime 1 with a value of WTI around \$ 90 from November 2010 to June 2014, regime 2 from June 2014 to January 2015 a fall of the price. And a regime 3 from January 2015 to December 2016 where the WTI oscillates around \$ 40 mark.



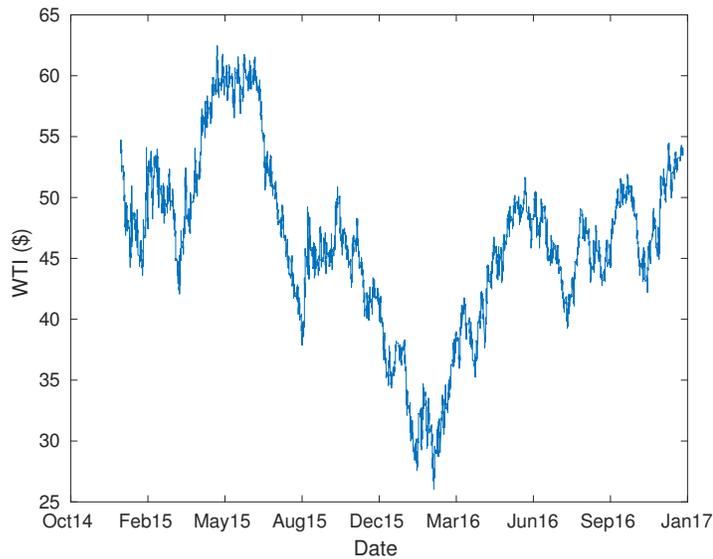
**Figure 2.56:** Period 1: High oil price regime, November 2010- to December 2013. This period is dominated by a WTI \$ 95 with a maximum value of \$114.83 on May 2<sup>nd</sup> 2011 after an increase of production by Saudi Arabia. The minimum price in this period was \$ 75.05 in October 04<sup>th</sup> 2011 due to recession fears.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES  
DYNAMICS

---



**Figure 2.57:** Period 2: 2014 the transition year. The price oscillates around \$100 from January 2014 until October 07<sup>th</sup> 2014 when it dropped below the \$90 mark. The maximum value of \$ 107.64 has been reached on June 12<sup>th</sup> 2014 while the lowest value \$ 52.51 was on December 31<sup>st</sup>2014



**Figure 2.58:** Period 3: Low oil price regime, from January 2015 to December 2016. The maximum value was \$ 62.45 occurred on May 6<sup>th</sup> 2015 and the minimum value was \$26.05 on February 11<sup>th</sup> 2016

### 2.8.4 Model of crude oil price dynamics

In order to model the oil price, we use the standard SDE representation introduced by [70, 77]. Let's  $S_t$  denote the crude oil spot price. A standard one factor model that describes the stochastic behavior of a commodity price has been introduced by [70]. The idea of this paper is to assume the dynamics of the commodity follows a diffusion process that can be constructed as follows:

$$\begin{aligned}\frac{dS_t}{S_t} &= b(S_t)dt + a(S_t)dW_t, \\ &= \kappa(\theta - \log(S_t))dt + \sigma dW_t,\end{aligned}\tag{2.46}$$

where  $\kappa$  is the speed of mean reversion,  $\theta$  is the long run mean and  $\sigma$  is the volatility. Defining the log price  $X_t = \log(S_t)$  and applying Ito's lemma to it the author in [70] obtained a SDE of an Ornstein-Uhlenbeck process

$$dX_t = \kappa(\theta^* - X_t)dt + \sigma dW_t,\tag{2.47}$$

with  $\theta^* = \theta - \frac{\sigma^2}{2\kappa}$ . This approach has been used as a reference to model the behavior of commodity prices. Note that parametric assumptions on the drift and diffusion have been formulated, the drift is linear and the volatility is constant. Unlike previous studies in the literature, such as [29, 70, 71] and others, except the log transformation of the price data, we are making no functional assumptions on the component of the SDE. In other words, using the spectral reconstruction methodology, we estimate the drift  $b(X_t)$  and diffusion  $a(X_t)$  of the following SDE:

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t,\tag{2.48}$$

where  $X_t = \log(S_t)$ .

### 2.8.5 Construction of the Markov chain and parameters selection

As in the previous section, we opt for a small number of eigenpairs and pick the first 3 only. As mentioned in the above description, gaps between observations prevent us from having data for every minute of each trading year. In order to avoid spurious transitions between states, we skip transition that didn't occur in a one minute time period. For example, suppose that we have data from 8:00 am today and missing data until tomorrow 8:00 am. We do not keep as valid the transition today to tomorrow but use 1 minute transitions from tomorrow 8:00 am onward. We also discard the transition from end of the year holidays (Christmas, New Year Eve, ...). This reduces the number of data, but guarantees that all the transitions recorded in the transition probability matrix occurred in a 1 minute time lapse. The time frequency of the data determined the scale of the parameters. In this application, we set the sampling frequency to the number of days divided by the total number of observations, T i.e.  $dt = \frac{\text{Number of days}}{\text{Number of observations}}$ . Another value for  $dt$  would just scale up or down the eigenvalues and the parameters but wouldn't change the reconstructed functional form of the drift and diffusion. The subsampling time,  $h$ , would be again a multiple of  $dt$  such as  $h = \alpha dt$  with  $\alpha \in \mathbb{N}$  and  $\alpha \geq 1$ . We would look at three cases  $\alpha = 2, 5, 10$  or in other words we sub-sample every second, fifth, and tenth observation. Increasing  $h$  implies discarding a certain amount of data. If for example, we sample every other data it means that we use only 50% of the data. If  $h = 10 * dt$  means that we use every other 10th value so 10% of the data. The process of subsampling reduces the sample size but, in the parametric case [4–7] and

non-parametric estimation of the previous section, has been shown to improve the quality of the estimators. The mean of the process would be denoted  $\mu_{X_t}$  and the volatility  $\sigma_{X_t}$ . A 95% confidence interval is given along with the parameters of a linear fit,  $p1 * x + p2$ , for the drift and quadratic fit,  $p1 * x^2 + p2 * x + p3$ , for the diffusion. An important feature of a mean-reverting process is the time that it takes to revert from its current level half way to its long term level. This half-life could be computed using the speed of reversion, which is the coefficient of the slope of the drift. We provide the half-life,  $t_{\frac{1}{2}}$ , for each subsampling time computed as follows

$$X_{t+1} = X_t e^{-\kappa t_{\frac{1}{2}}} \Rightarrow e^{-\kappa t_{\frac{1}{2}}} = \frac{X_{t+1}}{X_t} = 0.5 \Rightarrow t_{\frac{1}{2}} = \frac{\ln(2)}{\kappa}. \quad (2.49)$$

The value of  $t_{\frac{1}{2}}$  for each observational time step can be found by substituting the drift coefficient  $p1$  for  $\kappa$ .

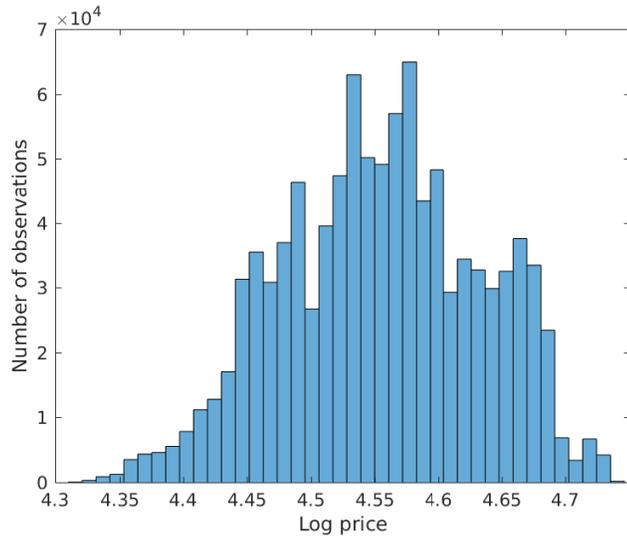
## 2.8.6 Spectral estimation applied to crude oil price data

### 2.8.6.1 Period 1: 2010-2013

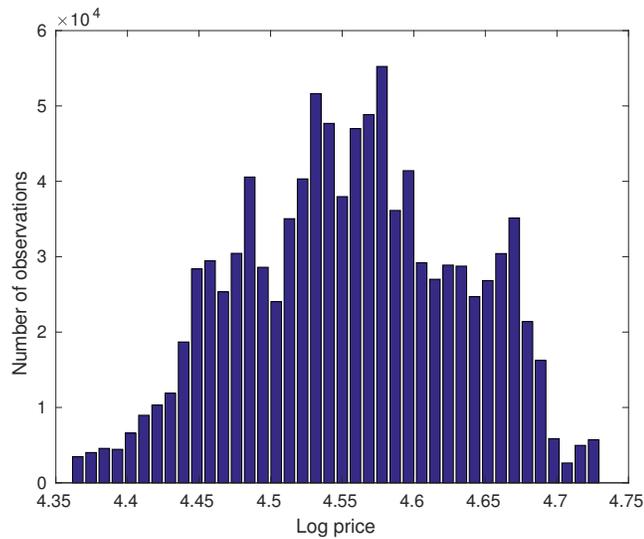
The total number of observations available for this period is equal 1013967. The histogram for the entire period is given in Figure 2.59. The sampling frequency is assumed to be  $dt = \frac{974}{1013967} = 0.00096$ . In order to remove outliers, we found that empirically better reconstruction is obtained with only log prices between  $[4.3607, 4.7302]$  that represents data in  $[\mu_{X_t} - 2.5\sigma_{X_t}, \mu_{X_t} + 2.25\sigma_{X_t}]$ . The histogram after truncation is given in Figure 2.60. We used  $M = 40$  bins. The spectral reconstruction still generates outliers in the estimated drift and diffusion.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---



**Figure 2.59:** Full data histogram of the period 2010-2013. The log price is on the x-axis and the number of observations on y-axis.



**Figure 2.60:** Truncated data histogram for the period 2010-2013. The log price is given on the x-axis and the number of observations on y-axis. We removed values on the both left and right tails such that each bin has at least 2000 observations.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---

$h$	$\Delta x$	Drift	Diff
2*dt	0.0092	p1= -0.14325 (-2.7024) p2= 0.65168 (2.7095)	p1=0.02606 (4.4935) p2= -0.23984 (-4.5599 ) p3=0.55367 (4.6427)
5*dt	0.0092	p1=-0.10082 (-2.9796) p2=0.45819 (2.9843)	p1= 0.01852 (4.5866) p2=-0.17029 (-4.6489 ) p3= 0.3927 (4.7283)
10*dt	0.0092	p1=-0.06467 (-2.3187) p2=0.29343 (2.3208)	p1=0.013639 (4.9519) p2=-0.12556 (-5.0296) p3=0.2898 (5.1254)

**Table 2.17:** Least squares fit for reconstruction during the period 2010-2013. The coefficients of a linear fit for the drift are provided for different subsampling times  $h$ . A quadratic fit was used for the diffusion and its coefficients are given for different values of  $h$ . The mean reverting assumption is validated for the drift whereas a non constant diffusion is identified

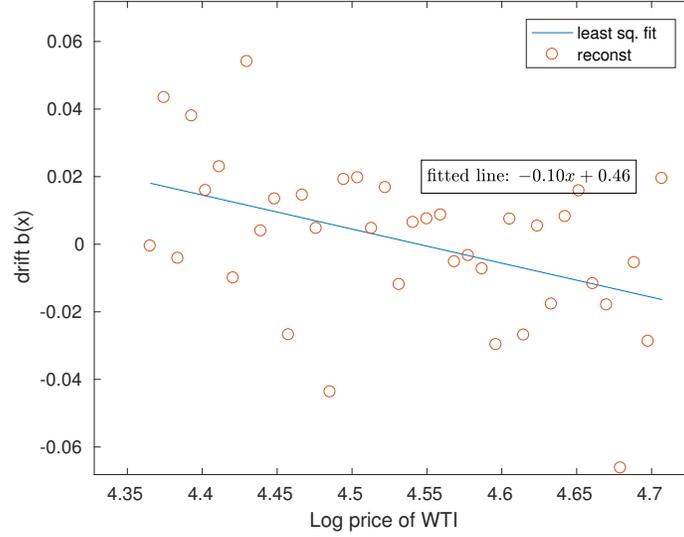
Model	$t_{\frac{1}{2}}$	RMSE Drift	RMSE Diff.
2*dt	4.84	0.033	0.000327
5*dt	6.87	0.021	0.000228
10*dt	10.72	0.017	0.000148

**Table 2.18:** Half-life and RMSE for period 1 2010-2013

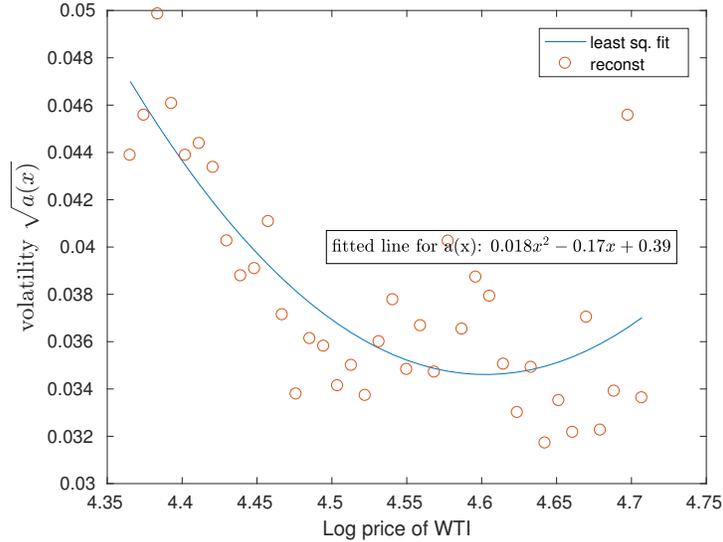
All the subsampling times confirmed the mean reversion hypothesis for the period from 2010 to 2013, see Table 2.17 and 2.18. The drift is linear with respect to the log price. Even if the mean reversion rate, the slope, changed with  $h$ , the long term mean which was equal to  $\frac{\rho^2}{\rho^1}$  was the same for all  $h$  and was equal to approximately 4.5. An unexpected finding is the rejection of the constant volatility assumption for this period. Indeed, the diffusion can be fitted with a quadratic function. So the estimated diffusion has a quadratic form. Figures 2.61 and 2.62 show the reconstructed drift and diffusion for the  $h = 5dt$  case. One limitation of these results is this estimation of the eigenvalues, fitted coefficients and, errors, measured by RMSE, decrease when  $h$  increases. The eigenfunctions were unaffected by a change in  $h$ . Future research will investigate this observation. Such behavior is an indication that more preliminary work on the data is required and suggest that some microstructures are currently unaccounted for. A second possibility might be that the drift and diffusion are time dependent and therefore might not be well captured by this method. Another possibility might be better described by a jump process or by fractional Brownian motion.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES  
DYNAMICS

---



**Figure 2.61:** Drift estimates with  $h = 5dt$ ,  $\Delta x = 0.0092$ , and  $M = 40$  for period from 2010 to 2013. The log price is given on the x axis and the value of the drift on the y axis. The estimates are in circles and the linear fit in line. The estimates are centered around a linear line with decreasing slope. This supports a mean reversion assumption for this period.



**Figure 2.62:** Diffusion estimates with  $h = 5dt$ ,  $\Delta x = 0.0092$ ,  $M = 40$  for period from 2010 to 2013. The log price is given on the x axis and the value of the diffusion on the y axis. The estimates are in circles and the linear fit in line. The estimates are distributed around a quadratic curve. This supports a non constant diffusion assumption for this period.

### 2.8.6.2 Period 3: 2015-2016

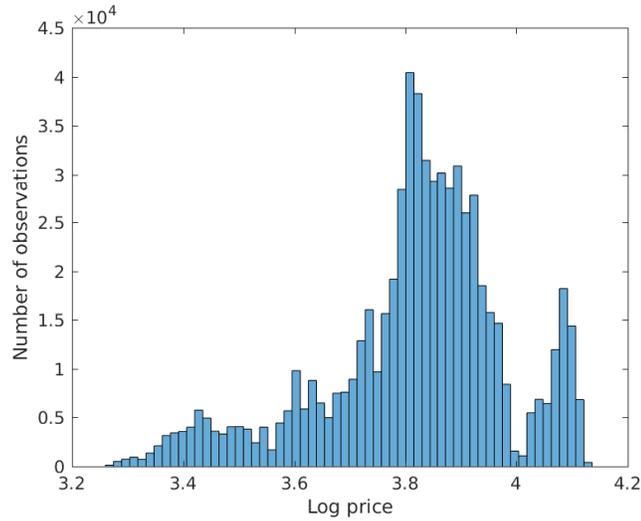
The total number of observations available for the period starting from January 2015 to December 2016 is 643591. We have less data for this time-frame. We obtain better estimations by increasing the number of bins by 20 to  $M = 60$ . The full histogram for the entire period between 2015-2016 is given in Figure 2.63. The stationary distribution is right skewed and different from the one of period 2010-2013. Only the log prices between  $[3.3581, 4.0008]$  that represents data in  $[\mu_{X_t} + 2.75\sigma_{X_t}, \mu_{X_t} + 1.1\sigma_{X_t}]$  are used in the reconstruction process. The histogram after truncation is given in Figure 2.64.

We present the estimated drift and diffusion in Figures 2.65 and 2.66 for the case  $h = 5dt$ . For this period, the random walk assumption seems to be validated, see Tables 2.19 and 2.20. No significant parameters can be found for the drift. The diffusion like in the previous case is quadratic. The half life coefficients are not significant in this case.

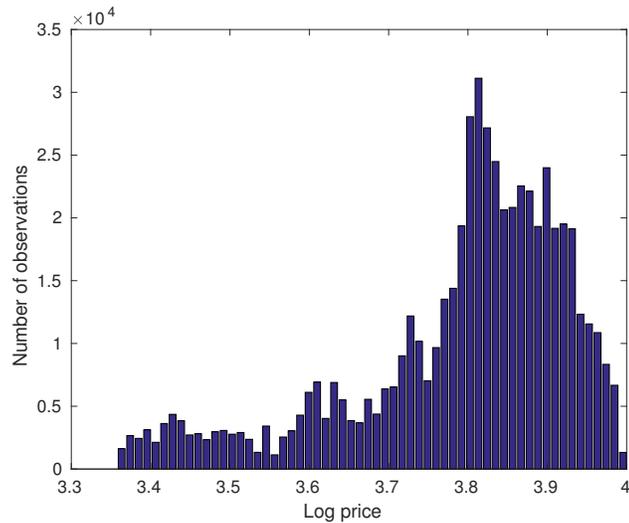
As before the eigenvalues and fitted coefficients in this estimation decrease when we increase  $h$ . The eigenfunctions are still unaffected by a change  $h$ . Further research on that matter is needed.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---



**Figure 2.63:** Full histogram for the period 2015-2016. The left tail contains bins with less than 2000 observations. Similarly on the right tail when the log price is close to 4, bins have a small number of observations.



**Figure 2.64:** Truncated histogram for the period 2015-2016. We removed the bins in both the left and right tails that contain less than 2000 observations.

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---

$h$	$\Delta x$	Drift	Diff
2*dt	0.0107	p1=0.042942 (0.68422 ) p2=-0.15922 (-0.6917 )	p1=0.01668 (4.3081) p2=-0.12738 (-4.4879) p3=0.24648 (4.7484)
5*dt	0.0107	p1= 0.07038 (1.1387) p2=-0.26605 (-1.1736)	p1= 0.01583 (4.5194) p2=-0.12042 (-4.6906) p3=0.23131 (4.9266)
10*dt	0.0107	p1=0.060668 (1.3582) p2=-0.22385 (-1.3664)	p1= 0.01391 (4.4745) p2=-0.10512 (-4.5914) p3=0.19951 (4.7649)

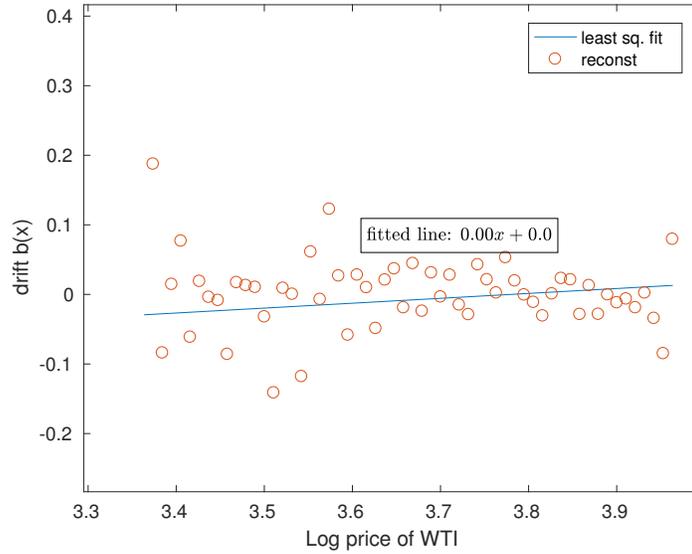
**Table 2.19:** Least Squares fit for reconstruction during the period 2015-2016. The coefficients of a linear fit for the drift are provided for different subsampling times  $h$ . A quadratic fit is used for the diffusion and its coefficients are given for different values of  $h$ . The random walk assumption for the drift failed to be rejected whereas a non-constant diffusion is identified.

Model	$t_{\frac{1}{2}}$	RMSE Drift	RMSE Diff.
2*dt	16.14	0.084	0.00082
5*dt	9.85	0.083	0.00074
10*dt	11.43	0.060	0.00066

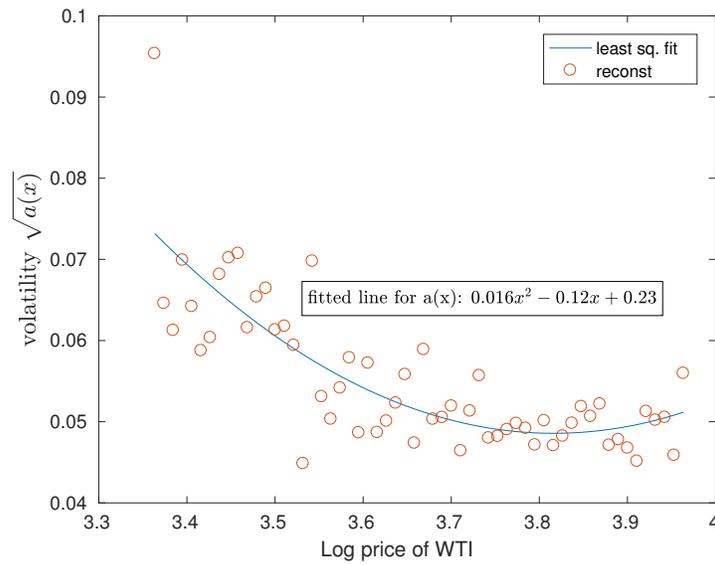
**Table 2.20:** Half-life and RMSE for period 3 from 2015 to 2016

2.8. APPLICATION: MODELING ENERGY COMMODITY PRICES DYNAMICS

---



**Figure 2.65:** Drift estimate with  $h = 5dt$ , and  $M = 60$ . The log price is given on the x axis and the value of the drift on the y axis. The estimates are in circles and the linear fit in line. The estimates are centered around the zero line what supports the random walk assumption.



**Figure 2.66:** Diff estimate with  $h = 5dt$  and  $M = 60$ . The log price is given on the x axis and the value of the diffusion on the y axis. The estimates are in circles and the linear fit in line. A quadratic fit well approximates the general distribution of the data, what supports the rejection of a constant drift assumption.

**Conclusion of this subsection:** The spectral estimation has been applied to crude oil prices. Two time periods have been selected for the estimation 2010-2013 and 2015-2016. The reconstruction of the SDE describing the crude oil price for each period has been performed. The first period validated the hypothesis of mean reversion in the oil price. The second period showed no significant drift coefficient and could be associated with a random walk process. These results are to be taken with caution since it appears that all the results scale with the observation time step. Further research on this observation will be done, however, this behavior might point to an omission in our model of some key data structures or the need for a methodology for SDE with jumps or fractional Brownian motion.

## 2.9 Conclusion Chapter 2

1. The reconstruction procedure has been tested to different affine processes relevant in many areas of science, engineering, and finance.
2. In most cases, diffusion can be reconstructed better than the drift.
3. Discarding a fixed high number of wave-number for Fourier filtering like in [18] appears too drastic. Instead, we tried instead to use a Parzen windows filtering scheme to smooth the entries of the Markov chain. The scheme has been done for a 1 and 2 dimensional filtering. A 1 dimensional filter is used where only the rows of the matrix are smoothed but in a 2 dimensional filter both rows and the columns are filtered. A triangular, or tent, and a Gaussian filter have been implemented. The results are close to the performance obtained by the

unfiltered matrix. The use of the Parzen windows produced good results for the periodic case.

4. For diffusion, of the OU, CIR, and JAC processes, both the shape and the value at different points were recovered while the shape of the drift was hardly recovered.
5. When the procedure used the discretized generator instead of the data-based transition probability matrix the drift and diffusion can be reconstructed regardless of the boundary conditions used.
6. Small differences between the eigenvalues and eigenfunctions of the data estimated transition probability matrix and these of the discretized generator seem to increase the errors in the reconstructed components.
7. Estimation procedures are sensitive to the choice of the subsampling time and the number of bins.

## CHAPTER 3

---

# CONDITIONAL EXPECTATION BASED RECONSTRUCTION OF DRIFT AND DIFFUSION

---

**The main goal of this chapter.** In this chapter, we analyze a second approach to the non-parametric estimation of the drift and diffusion of SDEs. This approach relies on conditional expectations formulas (2.2) introduced earlier in this thesis.

## 3.1 Introduction-Chapter 3

In the previous chapter, spectral reconstruction methods have been used to estimate the drift and diffusion functions of a SDE. One of the main conclusions was that the parameters used in the space and time discretization as well as the interval in which carries the estimation are crucial for an accurate estimation. An analysis of the sensitivity of the spectral reconstruction to these parameters have to take into account various type of errors. For example, we should consider errors from the construction of a random transition probability matrix, errors from the calculation of eigenvalues and eigenfunctions, errors from the generator's numerical derivatives, and also the least squares estimation errors in the optimization step. These considerations make the sensitivity analysis of the spectral approach complex and we refer to [18–20, 42] for more details. For simplicity and clarity reasons, we performed this sensitivity analysis in the context of conditional expectation based estimation, a second type of non-parametric technique. Consider a SDE:

$$dX_t = b(X_t)dt + a(X_t)dW_t. \quad (3.1)$$

with an initial condition,  $X(0) = x_0$ , and  $W_t$ , the standard Brownian Motion. The drift and diffusion have a statistical definition in terms of expectations, [24, 27, 64]:

$$\begin{aligned} b(x) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E(X_{t+\Delta t} - x | X_t = x), \\ a^2(x) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E[(X_{t+\Delta t} - x)^2 | X_t = x]. \end{aligned} \quad (3.2)$$

A discrete time estimator for the drift and for the diffusion can be constructed as follows. Consider  $N + 1$  observations  $\{X_i\}_{i=0\dots N}$  generated by a stochastic process

(3.1), a subsampling time step  $\Delta t$  with discrete estimators which are analogs of the conditional expectations in (3.2):

$$\begin{aligned}\tilde{b}(x) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\sum_{i=0}^N (X_{i+1} - x) \mathbb{1}_{\{X_i=x\}}}{\sum_{i=0}^N \mathbb{1}_{\{X_i=x\}}}, \\ \tilde{a}^2(x) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\sum_{i=0}^N (X_{i+1} - x)^2 \mathbb{1}_{\{X_i=x\}}}{\sum_{i=0}^N \mathbb{1}_{\{X_i=x\}}}.\end{aligned}\tag{3.3}$$

Note that the condition in the indicator function is the observed data point  $X_i$  is equal to one specific value. This event has a probability of zero but can be approximated by the event where  $X_i$  belongs to an interval centered at  $x$ . The state space of the process is discretized using a bin size of  $\Delta x$  and the event  $\{X_i = x\}$  is approximated by another which is  $X_i \in \left[ x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2} \right]$ . We assume in the following that  $x$  is the center of the given bin.

Our objective is to study for a particular interval, in the  $L^2$  sense, the behavior of the estimation error with respect to the spatial step  $\Delta x$ , the time step  $\Delta t$ , and the number of sample points involved in the estimation that we denote  $M$ . We consider a particular case of an OU process of the following form:

$$\begin{aligned}dX_t &= -\gamma X_t dt + \sigma dW_t, \\ X(0) &= X_0,\end{aligned}\tag{3.4}$$

where  $b(X_t) = -\gamma X_t$  and  $a(X_t) = \sigma$ , with  $\gamma$  and  $\sigma$  being two strictly positive constants. We set, without loss of generality,  $(\gamma, \sigma) = (1, 1)$ . Using standard Ito's calculus techniques, see [24], one can find that the solution of the SDE (3.4) is

$$X_t = X_0 e^{-\gamma t} + \sigma e^{-\gamma t} \int_0^t e^{\gamma s} dW_s.\tag{3.5}$$

The density of  $X_t$  is known to be a normal distribution

$$X_t \sim \mathcal{N} \left( X_0 e^{-\gamma t}, \frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma t}) \right).$$

Using the above solution, one can derive the following expression for the covariance of the process:

$$E(X_t X_s) = \frac{\sigma^2}{2\gamma} e^{-\gamma|t-s|} (1 - e^{-2\gamma(s \wedge t)}) + X_0^2 e^{-\gamma(t+s)}. \quad (3.6)$$

Discrete versions of these results for the OU process could be found in the literature, see [61]. Assume that these time series data,  $\{U_i\}_{i=0\dots N}$ , are collected. We have  $(N+1)$  direct observations,  $X_i = X_{t_i} = U_{i\Delta t}$  with  $i = 0, \dots, N$  extracted from the trajectory of an OU process during a time period  $[0, T]$  sub-sampled at discrete time steps,  $t_i = i\Delta t$ . The data points are observed at equal time intervals, i.e. they have a fixed subsampling such that  $0 = t_0 < \dots < t_i < \dots < t_N = T$  and  $t_{i+1} = t_i + \Delta t$ . Then  $X_{i+1}$  and  $X_i$  obey for any  $i$ :

$$X_{i+1} = X_i e^{-\gamma\Delta t} + \sigma e^{-\gamma t_{i+1}} \int_{t_i}^{t_{i+1}} e^{\gamma s} dW_s. \quad (3.7)$$

This expression by a change of variable becomes:

$$\begin{aligned} X_{i+1} &= X_i e^{-\gamma\Delta t} + \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s, \\ &= X_0 e^{-\gamma(i+1)\Delta t} + \sigma e^{-\gamma(i+1)\Delta t} \int_0^{(i+1)\Delta t} e^{\gamma s} dW_s. \end{aligned} \quad (3.8)$$

A finite difference version without stochastic integral, as in [4], is:

$$X_{i+1} = X_i e^{-\gamma\Delta t} + \sqrt{\frac{\sigma^2(1 - e^{-2\gamma\Delta t})}{2\gamma}} Z_i = X_0 e^{-\gamma(i+1)\Delta t} + \sqrt{\frac{\sigma^2(1 - e^{-2\gamma(i+1)\Delta t})}{2\gamma}} Z_i \quad (3.9)$$

where the  $Z_i$  are i.i.d standard Gaussian variables and are independent of  $X_0, \dots, X_N$  for each  $i = 1, \dots, N - 1$ . The data are not independent so a discretized version of the covariance (3.6) could also be derived.

## 3.2 Error Analysis

The first step in our analysis of the estimators (3.3) would be the study of the impact of the state-space discretization parameters,  $\Delta x$  and  $\Delta t$ , as well as the number of points in computing the conditional expectation. For this purpose, we will first look at the case where we know exactly the number of points used in each bin and then relax this assumption in the coming sections. We define two estimators,  $\hat{m}_1$  and  $\hat{m}_2$ , of the first and second moments respectively as:

$$\hat{m}_1 = \frac{\sum_{i=1}^N X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^N \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}, \quad \hat{m}_2 = \frac{\sum_{i=1}^N X_i^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^N \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}. \quad (3.10)$$

where  $B_{c_j}$  is an interval of size  $\Delta x$  centered at a point  $c_j$ . This interval is defined in more details in the following subsection.

### 3.2.1 Conditional expectation with fixed number of points

Our goal in this subsection is to have an estimate of the error made in each bin given that the number of points that is used in this bin is known. To compute this error, we restrict our attention to the interval  $[\mu_X - \sigma_X, \mu_X + \sigma_X]$ , where  $\mu_X$  and  $\sigma_X$  are respectively the mean and standard deviation of the process  $X_t$ . We refer to it as interval  $\mathcal{I}$ . We chose this interval to ensure that each partition would consist in

### 3.2. ERROR ANALYSIS

---

neighborhoods that are often visited by the trajectory. In this setting, we assume that we have exactly  $M$  points falling into a bin denoted  $B_{c_j}$  whose center is the point  $c_j$  and is of size  $\Delta x$ . More precisely, the bin is an interval obtained by partitioning  $\mathcal{I}$  into sub-intervals of size  $\Delta x$  centered at points  $c_j$  such that  $B_{c_j} = [c_j - \frac{\Delta x}{2}, c_j + \frac{\Delta x}{2}]$ . We use the notations  $c^\pm = c_j \pm \frac{\Delta x}{2}$  and also write  $B_{c_j} = [c^-, c^+]$ . Therefore, the denominator of the estimators (3.10) is just set to  $M$ .

$$\hat{m}_1 = \frac{1}{M} \sum_{i=1}^N X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}, \quad \hat{m}_2 = \frac{1}{M} \sum_{i=1}^N X_i^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}. \quad (3.11)$$

The above estimators (3.11) would estimate the first moment  $E[X_i | X_{i-1} = c_j]$  and the second one  $E[X_i^2 | X_{i-1} = c_j]$ , respectively.

#### Bias of the estimators

First, we would like to characterize the bias of the estimators. We start by computing the conditional expectation of the first estimator.

$$\begin{aligned} E[\hat{m}_1 | X_{i-1} = c_j] &= E\left[\frac{1}{M} \sum_{i=1}^N X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} \middle| X_{i-1} = c_j\right] \\ &= E\left[\frac{1}{M} \sum_{i=1}^M \left(X_{i-1} e^{-\gamma \Delta t} + \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s\right) \middle| X_{i-1} = c_j\right] \\ &= \frac{1}{M} \sum_{i=1}^M E\left[X_{i-1} e^{-\gamma \Delta t} \middle| X_{i-1} = c_j\right] = c_j e^{-\gamma \Delta t}. \end{aligned} \quad (3.12)$$

This conditional expectation of the first estimator gives the mean in terms of the center of the bin and is equal to the theoretical average of (3.7). Similarly, for the second moment, we obtained

$$\begin{aligned}
 E[\hat{m}_2 | X_{i-1} = c_j] &= E\left[\frac{1}{M} \sum_{i=1}^N \left(X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}\right)^2 \middle| X_{i-1} = c_j\right] \\
 &= E\left[\frac{1}{M} \sum_{i=1}^M \left(X_{i-1} e^{-\gamma \Delta t} + \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s\right)^2 \middle| X_{i-1} = c_j\right] \\
 &= \frac{1}{M} \sum_{i=1}^M E\left[X_{i-1}^2 e^{-2\gamma \Delta t} + \frac{\sigma^2(1 - e^{-2\gamma \Delta t})}{2\gamma} \middle| X_{i-1} = c_j\right] \\
 &= c_j^2 e^{-2\gamma \Delta t} + \frac{\sigma^2(1 - e^{-2\gamma \Delta t})}{2\gamma}
 \end{aligned} \tag{3.13}$$

The second moment also has no bias for this estimator. This expectation is a special case of a more general one involving  $X_{i-1} \in B_{c_j}$  where  $X_{i-1}$  doesn't have to be equal to the center, but can instead take any value in  $B_{c_j}$ . This more general condition would most likely lead to a bias of order  $\Delta x$  for the first estimator and  $(\Delta x)^2$  for the second one.

### 3.2.2 Crude error estimates

Recall that the conditional expectation of the OU process defined by equation (3.4) is  $E[X_{t+\Delta t} | X_t = c_j] = c_j e^{-\gamma \Delta t}$ .

### 3.2. ERROR ANALYSIS

---

The  $L_2$  error of  $\hat{m}_1$  is given by :

$$\begin{aligned}
& \left\| \hat{m}_1 - E[X_{t+\Delta t} | X_t = c_j] \right\|_2 = \left\| \hat{m}_1 - c_j e^{-\gamma \Delta t} \right\|_2 \\
& = \left\| \frac{1}{M} \sum_{i=1}^N X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} - c_j e^{-\gamma \Delta t} \right\|_2 \\
& = \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} e^{-\gamma \Delta t} + \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) - c_j e^{-\gamma \Delta t} \right\|_2 \\
& = \left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) e^{-\gamma \Delta t} + \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
& \text{(by triangular inequality)} \\
& \leq \left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) e^{-\gamma \Delta t} \right\|_2 + \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \quad (3.14) \\
& \text{(since } |X_{i-1} - c_j| \leq \frac{\Delta x}{2} \text{)} \\
& \leq \frac{\Delta x}{2} e^{-\gamma \Delta t} + \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma \Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
& \leq \frac{\Delta x}{2} e^{-\gamma \Delta t} + \frac{\sigma}{\sqrt{M}} \sqrt{\frac{1 - e^{-2\gamma \Delta t}}{2\gamma}} = \frac{\Delta x}{2} e^{-\gamma \Delta t} + \frac{1}{\sqrt{M}} \sqrt{\frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma \Delta t})} \\
& = \frac{\Delta x}{2} e^{-\gamma \Delta t} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \leq \frac{\sqrt{\Delta x}}{2} e^{-\gamma \Delta t} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \\
& \approx \frac{\sqrt{\Delta x}}{2} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}}, \quad \text{(for small } \Delta t \text{)}
\end{aligned}$$

Different cases appear with this bound:

1. When  $\Delta t$  is fixed, then the second term of the bound decays at a rate proportional to one over the square root of the number of points. An optimal scaling in this case would be  $\Delta x \propto \frac{1}{M}$ . This would allow the error to decay as  $\frac{1}{\sqrt{M}}$ . **So for a fixed  $\Delta t$ , having  $\Delta x \propto \frac{1}{M}$  would allow the error to decay at the rate  $\frac{1}{\sqrt{M}}$**

### 3.2. ERROR ANALYSIS

---

2. When  $\Delta t \rightarrow 0$  we need to balance the two terms in  $L^2$  error which leads to

$$\Delta x = \frac{\Delta t}{M}$$

Similar the  $L_2$  norm error of  $\hat{m}_2$  is given by

$$\begin{aligned}
& \left\| \hat{m}_2 - E[(X_{t+\Delta t})^2 | X_t = c_j] \right\|_2 \\
&= \left\| \hat{m}_2 - \left( \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma\Delta t}) + c_j^2 e^{-2\gamma\Delta t} \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} \right)^2 - \left( \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma\Delta t}) + c_j^2 e^{-2\gamma\Delta t} \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} e^{-\gamma\Delta t} + \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right)^2 - \left( \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma\Delta t}) + c_j^2 e^{-2\gamma\Delta t} \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1}^2 - c_j^2 \right) e^{-2\gamma\Delta t} + \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right)^2 \right. \\
&\quad \left. + \frac{1}{M} \sum_{i=1}^M 2X_{i-1} e^{-\gamma\Delta t} \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) - \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma\Delta t}) \right\|_2 \tag{3.15}
\end{aligned}$$

(using triangular inequality)

$$\begin{aligned}
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1}^2 - c_j^2 \right) e^{-2\gamma\Delta t} \right\|_2 + \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right)^2 \right\|_2 \\
&+ \left\| \frac{1}{M} \sum_{i=1}^M 2X_{i-1} e^{-\gamma\Delta t} \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 + \left\| \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma\Delta t}) \right\|_2 \\
&= \text{term ①} + \text{term ②} + \text{term ③} + \text{term ④}
\end{aligned}$$

### 3.2. ERROR ANALYSIS

---

Consider these four terms separately

$$\begin{aligned}
\text{term } \textcircled{1} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1}^2 - c_j^2 \right) e^{-2\gamma\Delta t} \right\|_2 \\
&= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( X_{i-1} + c_j \right) \right\|_2 \\
&= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( X_{i-1} - c_j + 2c_j \right) \right\|_2 \\
&= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right)^2 + \frac{1}{M} \sum_{i=1}^M 2c_j \left( X_{i-1} - c_j \right) \right\|_2 \\
&\quad \text{(using triangular inequality)} \\
&\leq e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right)^2 \right\|_2 + e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M 2c_j \left( X_{i-1} - c_j \right) \right\|_2 \\
&\leq e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 + 2c_j \left( \frac{\Delta x}{2} \right) \right] \quad , \text{ since } |X_{i-1} - c_j| \leq \frac{\Delta x}{2}
\end{aligned}$$

$$\begin{aligned}
\text{term } \textcircled{2} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right)^2 \right\|_2 \\
&= \sigma^2 e^{-2\gamma\Delta t} \frac{1}{M} \left\| \sum_{i=1}^M \left( \int_0^{\Delta t} e^{\gamma s} dW_s \right)^2 \right\|_2 \\
&\leq \frac{\sigma^2 e^{-2\gamma\Delta t}}{M} \left( \frac{e^{2\gamma\Delta t} - 1}{2\gamma} \right) \sqrt{2M + M^2} \\
&= \frac{\sigma^2}{2\gamma} \left( 1 - e^{-2\gamma\Delta t} \right) \sqrt{1 + \frac{2}{M}}
\end{aligned} \tag{3.16}$$

### 3.2. ERROR ANALYSIS

---

$$\begin{aligned}
\text{term } \textcircled{3} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma X_{i-1} e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) + \frac{2\sigma c_j}{M} \sum_{i=1}^M \left( e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&\text{(using triangular inequality)} \\
&\leq \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&\quad + \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma c_j e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&\text{(using the fact that } |X_{i-1} - c_j| \leq \frac{\Delta x}{2}\text{)} \\
&\leq \frac{\Delta x}{2} \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 + \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma c_j e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \left( \frac{\Delta x}{2} + c_j \right) \frac{2\sigma e^{-2\gamma\Delta t}}{\sqrt{M}} \sqrt{\left( \frac{e^{2\gamma\Delta t} - 1}{2\gamma} \right)}
\end{aligned} \tag{3.17}$$

The term  $\textcircled{4}$  remains unchanged. Combining all the terms gives:

$$\begin{aligned}
&\text{term } \textcircled{1} + \text{term } \textcircled{2} + \text{term } \textcircled{3} + \text{term } \textcircled{4} \\
&\leq e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 + 2c_j \left( \frac{\Delta x}{2} \right) \right] + \frac{\sigma^2}{2\gamma} \left( 1 - e^{-2\gamma\Delta t} \right) \sqrt{1 + \frac{2}{M}} \\
&\quad + \left( \frac{\Delta x}{2} + c_j \right) \frac{2\sigma e^{-2\gamma\Delta t}}{\sqrt{M}} \sqrt{\left( \frac{e^{2\gamma\Delta t} - 1}{2\gamma} \right)} + \frac{\sigma^2}{2\gamma} \left( 1 - e^{-2\gamma\Delta t} \right) \\
&= e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 + 2c_j \left( \frac{\Delta x}{2} \right) \right] + \frac{\sigma^2}{2\gamma} \left( 1 - e^{-2\gamma\Delta t} \right) \left( 1 + \sqrt{1 + \frac{2}{M}} \right) \\
&\quad + \frac{2\sigma e^{-2\gamma\Delta t}}{\sqrt{M}} \left( \frac{\Delta x}{2} + c_j \right) \sqrt{\left( \frac{e^{2\gamma\Delta t} - 1}{2\gamma} \right)} \\
&\approx c_j \Delta x + \sigma^2 \Delta t + 2c_j \sigma \left( \sqrt{\frac{\Delta t}{M}} \right), \quad \text{(for small } \Delta t\text{)}
\end{aligned} \tag{3.18}$$

Again the last line is obtained by taking a Taylor expansion of the exponential and

### 3.2. ERROR ANALYSIS

---

by discarding the second order terms. Unlike the first estimator, in addition to a term proportional to the spatial step size, we have a term linear in the subsampling time.

1. Given a number of points,  $M$ , when  $\Delta t$  is fixed if we choose  $\Delta x \propto \frac{1}{\sqrt{M}}$  the error decays as  $\Delta t + \frac{1}{\sqrt{M}}$ . Notice that the error doesn't decay to zero. **When  $\Delta t$  is fixed, an optimal spatial step size  $\Delta x \propto \frac{1}{\sqrt{M}}$  is picked to obtain the errors to decay at a rate of  $\frac{1}{\sqrt{M}}$  until they reach  $\Delta t$ .**
2. If  $M$  points are available in the bin, when  $\Delta t \rightarrow 0$  the spatial step size and the subsampling would have to approach zero at the same rate, e.g.  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{M}$ , to have a error to decay at the rate  $\frac{1}{M}$ . **Both the subsampling time and spatial step size should be inversely proportional to  $M$  to see the error decay at a rate of  $\frac{1}{M}$ .**

Combining error analysis for  $\hat{m}_1$  and  $\hat{m}_2$  we obtain the following scaling

1. for fixed  $\Delta t$ :  $\Delta x \propto \frac{1}{\sqrt{M}}$
2. for  $\Delta t \rightarrow 0$ ,  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{M^2}$ . In this case both  $L^2$  errors for  $\hat{m}_1$  and  $\hat{m}_2$  decay as  $\frac{1}{M}$ . However, scaling  $\Delta x \propto \frac{1}{M^2}$  produces extremely refined partitions for computing the conditional expectations. Recall that each bin required  $M$  data points. Thus, the total amount of data needed is of the order of  $M^3$  which can be prohibitively large. In this case, one may want to choose scaling

$$\Delta t \propto \frac{1}{M} \text{ and } \Delta x \propto \frac{1}{M}$$

### 3.2. ERROR ANALYSIS

---

Note that in the latter case the  $L^2$  error of the  $\hat{m}_1$  decays slower at  $\frac{1}{\sqrt{M}}$  due to the  $\sqrt{\Delta x}$  term. The  $L^2$  error for  $\hat{m}_2$  decays at rate  $\frac{1}{M}$ .

#### 3.2.3 Refined error estimates

##### Motivation

The  $L_2$  error estimates for the  $\hat{m}_1$  and  $\hat{m}_2$  provided in the previous section could be improved. The above derivation does not take into account the distribution of the different observations,  $X_{i-1}$  and rather used only the fact that the distance between each observation and the corresponding center in a given bin is less than half the spatial step size, i.e  $|X_{i-1} - c_j| \leq \frac{\Delta x}{2}$ . Another way to find an error bound is to consider the distribution of the observations falling inside a given bin. In other words, setting for a bin  $B_{c_j}$  the indicator function in (3.11) to 1 and estimate the value of the conditional expectation for this interval. Recall that each observation is extracted from an OU process such that for each  $i < j$ ,  $X_j$  conditional on  $X_i$  follows a normal distribution with mean  $X_i e^{-\gamma(j-i)\Delta t}$  and variance  $\frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma(j-i)\Delta t})$ . The distribution of all sample points over the real line is normal but the one conditional on being in one interval might not be normal. We have a distribution of points restricted to be in an certain interval of size  $\Delta x$  what gives us a normal random variable in a bounded interval. A normal random variable restricted to a certain interval is called a truncated normal distribution. Truncated distributions in general, and the truncated normal variable in particular, have been studied in various fields where they have have practical applications. For example hydrology (Pearson type III distribution )

or in econometrics (probit or tobit model) [13, 40, 49].

### Truncated normal distribution

Different types of truncation are possible, one sided left (right) truncation means that values of the random variable are greater (lower) than a certain threshold. There is also a two sided truncation, also known as doubly truncated distribution, which is a combination left and right truncations. In this section, we consider a doubly truncated normal distribution since the bins have a lower bound  $c^-$  and an upper one  $c^+$ . Next, we define the distribution. Suppose  $X \sim \mathcal{N}(\mu, \sigma^2)$  and takes value not on the entire real line  $\mathbb{R}$  but only in an interval  $[a, b]$ . Let's denote  $\phi(x)$  and  $\Phi(x)$  respectively the probability and cumulative density functions, henceforth, pdf and cdf, of a standard normal distribution.  $X$  conditional on  $a \leq X \leq b$  has a probability distribution  $f_X(x)$  on the support  $x \in [a, b]$  given by ( see [50])

$$f_X(x) = \begin{cases} \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})} & a \leq x \leq b \\ 0 & x \notin [a, b]. \end{cases} \quad (3.19)$$

In our work  $a = c_j - \frac{\Delta x}{2}$  and  $b = c_j + \frac{\Delta x}{2}$  and we define  $\alpha = \frac{a-\mu}{\sigma}$  and  $\beta = \frac{b-\mu}{\sigma}$ . The mean and the variance are given by ( see [12, 13, 50] )

$$\begin{aligned} E(X) &= \mu + \left( \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} \right) \sigma \\ Var(X) &= \sigma^2 \left[ 1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - \left( \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]. \end{aligned} \quad (3.20)$$

In our study for the data extracted from an OU process, the distribution of  $X_j$  conditional on  $X_i$  with for  $i < j$ , denoted hereafter  $X_j|X_i = c_j$ , has a mean  $\mu_{X_j} = X_i e^{-\gamma(j-i)\Delta t} = c_j e^{-\gamma(j-i)\Delta t}$  and a variance  $\sigma_{X_j}^2 = \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma(j-i)\Delta t})$ . Notice that we are interested in the case where  $j = i + 1$ , i.e.  $X_{i+1}|X_i$ . If the observation satisfies the

### 3.2. ERROR ANALYSIS

---

condition given by the indicator function, for a relatively small  $\Delta t$ , say  $\Delta t \leq 0.05$ , our bins are centered around the values  $c_j$ . If that is the case, each bin is centered around the mean of a Gaussian distribution and since the density is symmetric, we can simplify the expressions of the moments in (3.20). Using the symmetry of the standard normal distribution, we can deduce that for  $X_{i+1}|X_i$  we have  $\beta = \frac{\Delta x}{2\sigma_{X_{i+1}}}$ ,  $\alpha = \frac{c_j - \frac{\Delta x}{2} - c_j}{\sigma_{X_{i+1}}} = \frac{-\Delta x}{2\sigma_{X_{i+1}}} = -\beta$ . Also  $\phi(\alpha) = \phi(\beta)$  and  $\Phi(\beta) = 1 - \Phi(\alpha)$ . Now a small  $\Delta t$  loosens the dependence on the previous observation of the the mean and variance. We have for  $\Delta t < 0.05$  that  $\mu_{X_{i+1}} \approx X_i$  and by Taylor expanding the exponential in the variance,  $\sigma_{X_{i+1}}^2 \approx \sigma^2 \Delta t$ . We will then omit the subscript and denote the mean and variance of  $X_{i+1}|X_i$  by  $\mu_X$  and  $\sigma_X^2$ , respectively. Replacing these into the formulas (3.20) simplifies the above moments' definition as follows :

$$\begin{aligned} E(X) &= \mu_X, \\ \text{Var}(X) &= \sigma_X^2 \left[ 1 - \frac{\frac{\Delta x}{\sigma_X} \phi\left(\frac{\Delta x}{2\sigma_X}\right)}{2\Phi\left(\frac{\Delta x}{2\sigma_X}\right) - 1} \right]. \end{aligned} \tag{3.21}$$

#### **Autocorrelation of the points inside the bin**

As mentioned before, the values of the OU process are correlated but one would like to know if the observations that fall within a particular bin are themselves correlated. We first provide a numerical argument. We simulated 250000 sample points of an OU process, kept the values in one standard deviation of the mean, i.e values in  $[\mu_X - \sigma_X, \mu_X + \sigma_X]$ . Then we partition this interval into 1, 2, 3... 7 bins. We pick a bin, the first one for example, and record the value of the autocorrelation of order 1 to 5 of the data within the bin. For a stochastic process,  $X_t$ , the autocorrelation of order  $k$ , denoted  $r_k$ , measures the correlation between  $X_t$  and  $X_{t+k}$ , where  $k = 0 \dots K$ ,

### 3.2. ERROR ANALYSIS

---

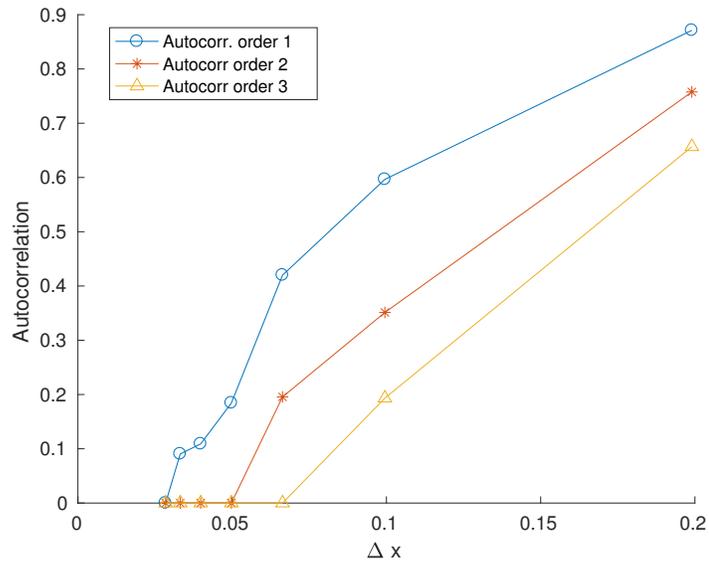
$K < T$ . The formula is defined by

$$r_k = \frac{C_k}{C_0} \quad \text{with} \quad C_k = \frac{1}{T-1} \sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (3.22)$$

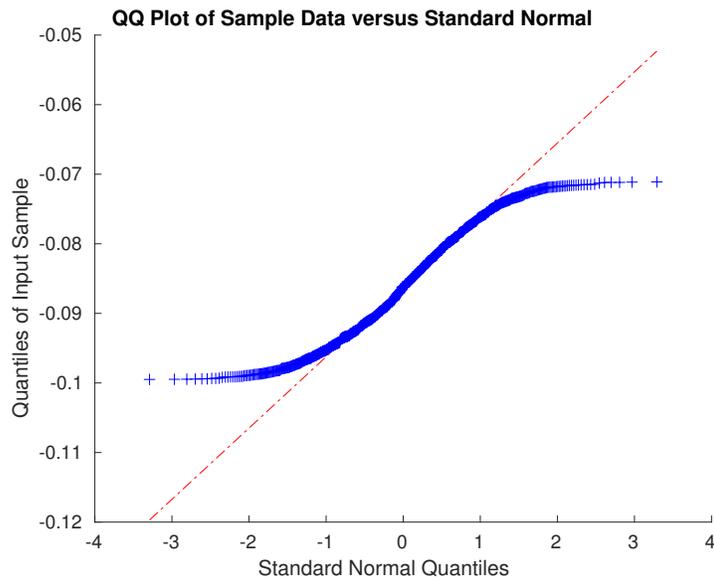
where  $\bar{X}$  and  $C_0$  are the sample mean and variance of the time series  $X_t$ , respectively. We empirically observed that autocorrelation is a function of the spatial step  $\Delta x$ . Small spatial step, i.e. large number of bins, implies little or no autocorrelation. Figure 3.1 below shows how fast the autocorrelation inside one bin decreases. We start with no partition, so 1 bin,  $\Delta x = 0.2$  and stop after dividing the interval into 7 bins, where  $\Delta x = 0.028$ . When we have no partition,  $\Delta x = 0.2$ , autocorrelation of order 1 is 0.8912, of order 2 is 0.7577, and of order 3 is 0.6567. After creating 3 intervals,  $\Delta x = 0.066$  the autocorrelation of order 1 and 2 are 0.42 and 0.1956 respectively while the autocorrelation of order 3 vanished. With 4 bins,  $\Delta x = 0.049$ , we have only an autocorrelation of order 1 whose value was 0.1844. When the interval was divided into 7 bins, and  $\Delta x = 0.028$ , no autocorrelation was found in any bin. This is an argument in favor of using the independence assumption for the points inside the bins. A QQplot was done to see how close the data inside the bin are from a normal distribution. We can see from Figure 3.2 that the extreme values of the sample data are far from being from a normal. This is due to the truncation of the tails of a normal probability density. On the other hand, the data between the first standard normal quantiles show a perfect match with the sample data quantiles. This shows that these data are close to normally distributed points.

### 3.2. ERROR ANALYSIS

---



**Figure 3.1:** This graph shows the autocorrelation of order 1,2, and 3 of the data recorded in the first bin formed by the different partition of the interval  $[\mu_X - \sigma_X, \mu_X + \sigma_X]$  into bins. After dividing the interval by 4 we have no autocorrelation of order 2 and 3. When the interval is divided by 7 the data does not show autocorrelation.



**Figure 3.2:** The QQplot shows that the data inside the first bin are normal around the mean but differ greatly in the tails.

**Autocorrelation of the points inside the bin, a probabilistic argument**

To explain the decay of the autocorrelation when spatial step size decreases, we consider the question of how to pick  $\Delta x$ . One drawback of an arbitrary discretization is a wrong choice of parameters,  $\Delta t$  and  $\Delta x$ , could create dependence between observations in each bin. This should not be surprising since given a fixed subsampling frequency, a large bin has a higher probability to contain consecutive observations. If the bin width is too small the probability of it not being visited increases, which creates a need for a large number of points. To address this issue, we consider the probability that consecutive observations are in the same bin. First, we assume that our process lives in a measure space  $(\Omega, \mathcal{F}, P)$  where  $\Omega$  is the state space,  $\mathcal{F}$  is the sigma algebra of subsets of  $\Omega$ , and a probability measure,  $P$ . In this setting, we want to understand how space-time discretization impacts  $P(|X_{t+\Delta t} - X_t| > \Delta x)$ . The SDE gives us the distribution of the increments,  $\Delta X_t = X_{t+\Delta t} - X_t$ , conditional on  $X_t$  which is Gaussian with a mean of,  $-X_t \Delta t$ , and unit variance. Although the distribution of each  $\Delta X_t$  is Gaussian, the direct computation of the probability would involve a state-dependent expectation, which generates an extremely complicated expression. The value of the probability,  $P$ , itself is not important only its connection with  $\Delta t$  and  $\Delta x$  is of interest. We used an equivalent probability measure to study the connection. The OU process under a different probability measure is just a Brownian motion. The change of the probability measure can be obtained using the Girsanov theorem, [8, 64, 65]. Recall the OU SDE:

$$dX_t = -X_t dt + dW_t, \tag{3.23}$$

Using the Girsanov theorem, a new probability,  $Q(\cdot)$  on  $(\Omega, \mathcal{F})$ , equivalent to  $P$ ,

### 3.2. ERROR ANALYSIS

---

could be defined for any subset  $A \subset \Omega$ . For such  $A$ , we would have:

$$Q(A) = \int_A Z(w) dP(w), \quad (3.24)$$

where  $Z(w) = \exp\{\int_0^t -X_s dW_s - \frac{1}{2} \int_0^t X_s^2 ds\}$  is the Radon Nikodym derivative. Under the probability measure,  $Q$ , a Brownian motion,  $B_t$ , could be defined on the filtration,  $\mathcal{F}$ , such that the increment,  $\Delta X_t$ , is equal to this process, i.e.  $\Delta X_t = \Delta B_t$ . The increments under  $Q$  are then normally distributed with a mean of zero and variance  $\Delta t$ . Instead of analyzing  $P(|X_{t+\Delta t} - X_t| > \Delta x)$ , we can compute  $Q(|X_{t+\Delta t} - X_t| > \Delta x) = Q(|\Delta X_t| > \Delta x)$ . The absolute value of a normal density with a mean of zero is called a half-normal distribution and the moments are defined. Since our random variable is non-negative and should be greater than  $\Delta x > 0$ , we can apply Markov's inequality and obtain

$$\begin{aligned} Q(|\Delta X_t| > \Delta x) &\leq \frac{E^Q|\Delta X_t|}{\Delta x}, \\ &= \frac{\Delta t}{\Delta x} \sqrt{\frac{2}{\pi}}. \end{aligned} \quad (3.25)$$

where  $E^Q$  is the expectation under the probability measure  $Q$ . For  $\Delta x$  too small this bound is not informative. However, we can see that if  $\Delta t$  goes to zero faster than  $\Delta x$ , observations would tend to be in the same bin. The ratio of the time and spatial steps affect the correlations inside the bin, which could impact the estimation. Since  $P$  and  $Q$  share the same null sets, this conclusion would be true for both probability measures.

#### **$L^2$ error analysis with independence assumption**

The same behavior can be observed inside any bin. This allows us to assume that the observations inside the bin are independent for some values of  $\Delta x$ , e.g., when

### 3.2. ERROR ANALYSIS

---

the number of bins is greater than 7. In addition to this observation, we can assume that the sample points identically follow a truncated normal distribution. In order to simplify the notation, we omit the subscript  $X_i$  and define the truncated variance as

$$\eta^2 = \sigma_X^2 \left[ 1 - \frac{\frac{\Delta x}{\sigma_X} \phi\left(\frac{\Delta x}{2\sigma_X}\right)}{2\Phi\left(\frac{\Delta x}{2\sigma_X}\right) - 1} \right].$$

The variance doesn't depend on the centers of the bin. Each event  $\{X_{i+1}|X_i = c_j\}$  has a truncated Gaussian distribution,  $\mathcal{TN}(c_j, \eta^2)$ . Using this set up, we can find alternative bounds for the estimators  $\hat{m}_1$  and  $\hat{m}_2$  by computing

$$\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) \right\|_2 \text{ in (3.14) and } \left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j)^2 \right\|_2 \text{ in (3.16)}$$

using their corresponding probability distribution. We start with  $\hat{m}_1$  and the 2-norm term in (3.14). Given that  $X_{i-2} = c_j$ , we have  $X_{i-1} \in B_{c_j}$  and  $\{X_{i-1}|X_{i-2} = c_j\} \sim \mathcal{TN}(c_j, \eta^2)$ . Moments of a standardized doubly-truncated random variable works like the non-truncated case, see [12, 13]. So we can standardize like a normal distribution. Therefore, we have  $\{X_{i-1} - c_j|X_{i-2} = c_j\} \sim \mathcal{TN}(0, \eta^2)$ . Consider the 2-norm term. By the law of iterated expectation [24, 64], one can see that the argument is the mean of the  $M$ -independent truncated normal distribution centered at zero. When  $M$  is large, by the central limit theorem for a truncated normal distribution [13] this mean follows a normal  $\mathcal{N}(0, \frac{\eta^2}{M})$ .

### 3.2. ERROR ANALYSIS

---

The  $L^2$  term squared can be computed as:

$$\begin{aligned}
\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) \right\|_2^2 &= E \left[ \left( \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) \right)^2 \right] = \frac{1}{M^2} E \left[ \left( \sum_{i=1}^M (X_{i-1} - c_j) \right)^2 \right] \\
&= \frac{1}{M^2} \sum_{i=1}^M E \left( X_{i-1} - c_j \right)^2 = \frac{\eta^2}{M} \\
&= \frac{1}{M} \sigma_X^2 \left[ 1 - \frac{\frac{\Delta x}{\sigma_X} \phi\left(\frac{\Delta x}{2\sigma_X}\right)}{2\Phi\left(\frac{\Delta x}{2\sigma_X}\right) - 1} \right]
\end{aligned} \tag{3.26}$$

From line 2 to 3, we used the independence assumption of the points inside the bin.

**Lemma 3.2.1.** For any  $y > 0$ ,  $\left( 1 - \frac{2y\phi(y)}{2\Phi(y)-1} \right) < y^2$ .

*Proof.* Using a Taylor approximation of the cumulative of a normal distribution suggested in [56] we can write:

$$\begin{aligned}
\Phi(y) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \left[ y + \frac{y^3}{3} + O(y^5) \right] = \frac{1}{2} + \phi(y) \left[ y + \frac{y^3}{3} + O(y^5) \right] \\
\Rightarrow 2\Phi(y) - 1 &= 2\phi(y) \left[ y + \frac{y^3}{3} + O(y^5) \right] \\
\Rightarrow 1 - \frac{2y\phi(y)}{2\Phi(y) - 1} &= 1 - \frac{2y\phi(y)}{2y\phi(y) \left[ 1 + \frac{y^2}{3} \right]} = 1 - \frac{1}{1 + \frac{y^2}{3}} \\
&= 1 - \frac{3}{3 + y^2} = \frac{y^2}{3 + y^2} < y^2
\end{aligned}$$

□

Using the Lemma 3.2.1:

$$\begin{aligned}
\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) \right\|_2 &= \sqrt{\frac{\eta^2}{M}} \\
&= \sqrt{\frac{1}{M} \sigma_X^2 \left[ 1 - \frac{\frac{\Delta x}{\sigma_X} \phi\left(\frac{\Delta x}{2\sigma_X}\right)}{2\Phi\left(\frac{\Delta x}{2\sigma_X}\right) - 1} \right]} \\
&\leq \frac{\sigma_X}{\sqrt{M}} \sqrt{\left( \frac{\Delta x}{2\sigma_X} \right)^2} = \frac{1}{\sqrt{M}} \frac{\Delta x}{2}
\end{aligned} \tag{3.27}$$

### 3.2. ERROR ANALYSIS

---

One can see that the 2 norm term is dependent on the number of points in the bin  $M$  unlike the first approach. From this, we can find a different bound for the first moment estimator  $\hat{m}_1$ :

$$\begin{aligned}
\|\hat{m}_1 - E[X_{t+\Delta t}|X_t = c_j]\|_2 &= \|\hat{m}_1 - c_j e^{-\gamma\Delta t}\|_2 \\
&\leq \left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j) e^{-\gamma\Delta t} \right\|_2 \\
&\quad + \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&< \frac{1}{\sqrt{M}} \frac{\Delta x}{2} e^{-\gamma\Delta t} + \frac{1}{\sqrt{M}} \sqrt{\frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma\Delta t})} \quad (3.28) \\
&= \frac{1}{\sqrt{M}} \frac{\Delta x}{2} e^{-\gamma\Delta t} + \frac{1}{\sqrt{M}} \sigma_X \\
&= \frac{1}{\sqrt{M}} \frac{\Delta x}{2} e^{-\gamma\Delta t} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \\
&\leq \frac{\sqrt{\Delta x}}{\sqrt{M}} e^{-\gamma\Delta t} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \approx \frac{\sqrt{\Delta x}}{\sqrt{M}} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}}
\end{aligned}$$

1. Given a number of points  $M$ , when  $\Delta t$  is fixed the error decays as  $\frac{1}{\sqrt{M}}$ . So the choice of the spatial step should not influence the decay of the errors. **No optimal spatial step seems necessary for the errors to decay at rate  $\frac{1}{\sqrt{M}}$ .**
2. If  $M$  points are available in the bin, when  $\Delta t \rightarrow 0$  choosing a spatial step size to approach zero at the same rate, e.g.  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{M}$ , would make the error to decay at the rate  $\frac{1}{M}$ . **In this case, both the subsampling time and spatial step size are inversely proportional to  $M$  as well as the decay rate of the estimation errors.**

### 3.2. ERROR ANALYSIS

---

Alternatively, the second moment estimator bound could also be derived by using the probability distribution of  $X_{i+1}|X_i$  to compute  $\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j)^2 \right\|_2$ . The distribution of this summation can be found using the Gaussian distribution. In fact, recall that  $X_{i-1} - c_j$  follows a  $\mathcal{TN}(0, \eta^2)$ . Using moment generating functions, the square of a truncated normal could be shown to follow a truncated non-central Chi square i.e.  $(X_{i-1} - c_j)^2 \sim \eta^2 \chi_1$ . We omit this proof. The summation of Chi square with one degree of freedom is a Chi square whose degree of freedom is the number of variables in the sum. We can write  $\sum_{i=1}^M (X_{i-1} - c_j)^2 \sim \eta^2 \chi_M$ . We can finally divided both side by M and obtain the distribution  $\frac{\eta^2}{M} \chi_M$ . The computation of the mean is straightforward:

$$E \left[ \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j)^2 \right] \leq \frac{1}{M} \eta^2 M = \eta^2 \quad (3.29)$$

The inequality is obtained assuming that the mean of a truncated Chi square is smaller than the mean of an un-truncated Chi square distribution. We obtain the second moment using the central Chi square moments.

$$\begin{aligned} E \left[ \left( \frac{1}{M} \sum_{i=1}^M \left( \frac{X_{i-1} - c_j}{\eta} \right)^2 \eta^2 \right)^2 \right] &= \frac{\eta^4}{M^2} E \left[ \left( \sum_{i=1}^M \left( \frac{X_{i-1} - c_j}{\eta} \right)^2 \right)^2 \right] \\ &= \frac{\eta^4}{M^2} E \left[ \sum_{i=1}^M \left( \frac{X_{i-1} - c_j}{\eta} \right)^4 \right. \\ &\quad \left. + 2 \sum_{k=1}^M \sum_{l>k}^M \left( \frac{X_{k-1} - c_j}{\eta} \right)^2 \left( \frac{X_{l-1} - c_j}{\eta} \right)^2 \right] \\ &\leq \frac{\eta^4}{M^2} \left[ 2M + M(M-1) \right] = \frac{\eta^4}{M} + \eta^4 \end{aligned} \quad (3.30)$$

We assumed again that the variance of a truncated distribution is smaller than the un-truncated one. We also used the independence of increments,  $(X_{k-1} - c_j)^2$  and

### 3.2. ERROR ANALYSIS

---

$(X_{l-1} - c_j)^2$ , to find the above expectation. We can then write:

$$\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j)^2 \right\|_2 \leq \sqrt{\frac{\eta^4}{M} + \eta^4} = \eta^2 \sqrt{\frac{1}{M} + 1}.$$

From Lemma 3.2.1, we have that  $\eta^2 < \left(\frac{\Delta x}{2}\right)^2$ . Then

$$\left\| \frac{1}{M} \sum_{i=1}^M (X_{i-1} - c_j)^2 \right\|_2 \leq \eta^2 \sqrt{\frac{1}{M} + 1} < \left(\frac{\Delta x}{2}\right)^2 \sqrt{\frac{1}{M} + 1}.$$

We can again look at the four different terms in (3.15) of the  $L_2$  error computed for the second moment estimator  $\hat{m}_2$ . The second and last terms remain the same while the first and third ones are modified.

$$\begin{aligned} \text{term } \textcircled{1} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1}^2 - c_j^2 \right) e^{-2\gamma\Delta t} \right\|_2 \\ &= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( X_{i-1} + c_j \right) \right\|_2 \\ &= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( X_{i-1} - c_j + 2c_j \right) \right\|_2 \\ &= e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right)^2 + \frac{1}{M} \sum_{i=1}^M 2c_j \left( X_{i-1} - c_j \right) \right\|_2 \\ &\quad \text{(by the triangular inequality)} \\ &\leq e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right)^2 \right\|_2 + e^{-2\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M 2c_j \left( X_{i-1} - c_j \right) \right\|_2 \\ &\quad \text{(using the distribution of } \{X_{i-1} - c_j\}) \\ &< e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 \sqrt{1 + \frac{1}{M}} + 2c_j \left( \frac{\Delta x}{2} \right) \left( \frac{1}{\sqrt{M}} \right) \right]. \end{aligned} \tag{3.31}$$

### 3.2. ERROR ANALYSIS

---

We can similarly find an upper bound for the third term as follows:

$$\begin{aligned}
\text{term } \textcircled{3} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma X_{i-1} e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) + \frac{2\sigma c_j}{M} \sum_{i=1}^M \left( e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&\quad + \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma c_j e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&\quad + \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma c_j e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \text{term (3a)} + \text{term (3b)}.
\end{aligned}$$

$$\begin{aligned}
\text{term (3a)} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma(X_{i-1} - c_j) e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= 2e^{-\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \frac{2}{M} e^{-\gamma\Delta t} \left\| \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
&= \frac{2}{M} e^{-\gamma\Delta t} \sqrt{E \left( \sum_{i=1}^M \left( X_{i-1} - c_j \right) \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right)^2} \\
&= \frac{2}{M} e^{-\gamma\Delta t} \sqrt{M\eta^2\sigma_X^2} = \frac{2}{\sqrt{M}} e^{-\gamma\Delta t} \eta\sigma_X \\
&< \frac{2\sigma_X}{\sqrt{M}} e^{-\gamma\Delta t} \left( \frac{\Delta x}{2} \right) = \sigma_X e^{-\gamma\Delta t} \left( \frac{\Delta x}{\sqrt{M}} \right).
\end{aligned}$$

(3.32)

$$\begin{aligned}
 \text{term (3b)} &= \left\| \frac{1}{M} \sum_{i=1}^M \left( 2\sigma c_j e^{-2\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 \\
 &= 2c_j e^{-\gamma\Delta t} \left\| \frac{1}{M} \sum_{i=1}^M \left( \sigma e^{-\gamma\Delta t} \int_0^{\Delta t} e^{\gamma s} dW_s \right) \right\|_2 = 2c_j e^{-\gamma\Delta t} \frac{\sigma_X}{\sqrt{M}} \quad (3.33)
 \end{aligned}$$

$$\begin{aligned}
 \text{term ③} &= \text{term (3a)} + \text{term (3b)} \\
 &< \sigma_X e^{-\gamma\Delta t} \left( \frac{\Delta x}{\sqrt{M}} \right) + 2c_j e^{-\gamma\Delta t} \frac{\sigma_X}{\sqrt{M}}.
 \end{aligned}$$

$$\begin{aligned}
 \|\hat{m}_2 - E[(X_{t+\Delta t})^2 | X_t = c_j]\|_2 &= \text{term ①} + \text{term ②} + \text{term ③} + \text{term ④} \\
 &< e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 \sqrt{1 + \frac{1}{M}} + 2c_j \left( \frac{\Delta x}{2} \right) \left( \frac{1}{\sqrt{M}} \right) \right] + \sigma_X^2 \sqrt{1 + \frac{2}{M}} + \sigma_X e^{-\gamma\Delta t} \left( \frac{\Delta x}{\sqrt{M}} \right) \\
 &\quad + 2c_j e^{-\gamma\Delta t} \frac{\sigma_X}{\sqrt{M}} + \sigma_X^2 \\
 &= e^{-2\gamma\Delta t} \left[ \left( \frac{\Delta x}{2} \right)^2 \sqrt{1 + \frac{1}{M}} + 2c_j \left( \frac{\Delta x}{2} \right) \left( \frac{1}{\sqrt{M}} \right) \right] + \sigma_X^2 \left( 1 + \sqrt{1 + \frac{2}{M}} \right) \\
 &\quad + \sigma_X e^{-\gamma\Delta t} \left( \frac{\Delta x}{\sqrt{M}} + \frac{2c_j}{\sqrt{M}} \right) \\
 &\approx \left( \frac{c_j \Delta x}{\sqrt{M}} \right) + \sigma^2 \Delta t + \sigma \sqrt{\Delta t} \left( \frac{\Delta x}{\sqrt{M}} + \frac{2c_j}{\sqrt{M}} \right) \\
 &= \left( \frac{c_j \Delta x}{\sqrt{M}} \right) + \sigma^2 \Delta t + 2c_j \sigma \left( \sqrt{\frac{\Delta t}{M}} \right) + \sigma \sqrt{\Delta t} \left( \frac{\Delta x}{\sqrt{M}} \right) \\
 &= \left( \frac{c_j + \sigma \sqrt{\Delta t}}{\sqrt{M}} \right) \Delta x + \sigma^2 \Delta t + 2c_j \sigma \left( \sqrt{\frac{\Delta t}{M}} \right)
 \end{aligned}$$

The bound for the second moment has the same behavior than the crude bound for  $\hat{m}_2$  derived without the independence assumption.

1. Given a number of points,  $M$ , when  $\Delta t$  is fixed, the error doesn't go to zero but should converge to  $\Delta t$  when  $M$  is large and we chose  $\Delta x \propto \frac{1}{\sqrt{M}}$ .

### 3.2. ERROR ANALYSIS

---

When  $\Delta t$  is fixed and we choose a spatial step size  $\Delta x \propto \frac{1}{\sqrt{M}}$ , the errors decay at a rate of  $\frac{1}{\sqrt{M}}$  to  $\Delta t$ .

2. With  $M$  points in the bin, when  $\Delta t \rightarrow 0$ , we have the same conclusion than in the case of the crude bound of  $\hat{m}_2$ . Choosing a spatial step size that approaches zero at the same rate than the subsampling time, e.g.  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{M}$ , would make errors decay at a rate of  $\frac{1}{M}$ . **In this case, the subsampling time, spatial step size, and the decay rate of the estimation errors are inversely proportional to the number of points in the bin,  $M$ .**

In summary the different upper bounds obtained are:

#### Crude error bounds

$$\begin{aligned} \|\hat{m}_1 - E[X_{t+\Delta t}|X_t = c_j]\|_2 &\leq \frac{\sqrt{\Delta x}}{2} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \\ \|\hat{m}_2 - E[(X_{t+\Delta t})^2|X_t = c_j]\|_2 &\leq c_j \Delta x + \sigma^2 \Delta t + 2c_j \sigma \left( \sqrt{\frac{\Delta t}{M}} \right) \end{aligned} \tag{3.34}$$

#### Refined error bounds

$$\begin{aligned} \|\hat{m}_1 - E[X_{t+\Delta t}|X_t = c_j]\|_2 &\leq \frac{\sqrt{\Delta x}}{\sqrt{M}} + \sigma \frac{\sqrt{\Delta t}}{\sqrt{M}} \\ \|\hat{m}_2 - E[(X_{t+\Delta t})^2|X_t = c_j]\|_2 &\leq \left( \frac{c_j + \sigma \sqrt{\Delta t}}{\sqrt{M}} \right) \Delta x + \sigma^2 \Delta t + 2c_j \sigma \left( \sqrt{\frac{\Delta t}{M}} \right) \end{aligned}$$

### 3.2.4 Numerical investigation

The goal of this subsection is to numerically justified the existence of the bounds (3.34). We generate some paths and compute the errors and compare them to the values of the upper bounds (3.34) at each discretization point. Parameters used were

### 3.2. ERROR ANALYSIS

---

$T = 250000$ ,  $\Delta t = 0.01$ , the number of points in the bin  $B_{c_j}$  centered at a point  $c_j$  is denoted by  $M_{c_j} = \{50, 100, 200, 300, 500, 600, 700, 800, 900, 1000\}$ . The number of simulated paths is denoted by  $N_{sim}$ , which is set at  $N_{sim} = 2000$ . The bin chosen is the one whose center is equal to the long term mean plus  $\Delta x$ . Using a log log scale, we compare a line of slope  $-0.5$  with the graph of the relationship between estimation errors and the number of points inside the bin. If slopes are the same it would confirm that the errors have a decay rate of  $\frac{1}{M}$ . To compute the error at each point, we chose the Root Mean Square Error (RMSE) defined as:

$$error_{b_j} = \sqrt{\frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} |b_j - b_j^{true}|^2},$$

$$error_{a_j} = \sqrt{\frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} |a_j - a_j^{true}|^2}.$$

where  $b_j, a_j, b_j^{true}, a_j^{true}$  are the computed and true values of  $\hat{m}_1$  and  $\hat{m}_2$  respectively. The estimation errors are computed and compared to the crude and refined bounds.

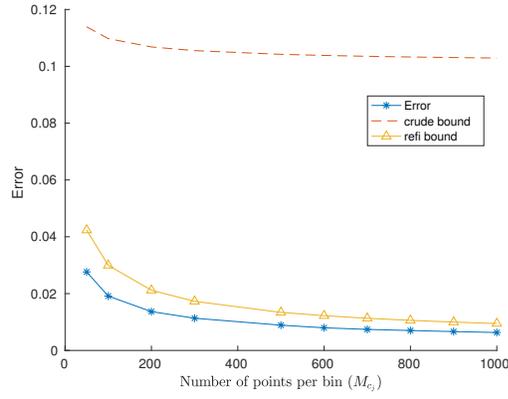
**Test 1:**  $\Delta x = 0.039801$ ,  $c_j = 0.039801$ , **subsampling time step**  $\Delta t = 0.01$ , **simulation time step**,  $dt = 0.001$  **while the number of bins is 5**

Figures 3.3 and 3.4 show that for a fixed sub-sampling time the errors decay at a rate inversely proportional to  $M_{c_j}$ . At  $M_{c_j} = 1000$  the error of the  $\hat{m}_1$ ,  $error_{b_j}$ , is 0.006006. The error is below the refined bound at 0.0094711. The crude estimator is much larger at 0.10291. Similarly for the  $\hat{m}_2$  error, when  $M_{c_j} = 1000$ , the error  $error_{a_j}$  is 0.0008938. The crude bound is 0.012 while the refined is 0.011. The refined bound for the  $\hat{m}_1$  has the same slope as the error estimate whereas for the  $\hat{m}_2$  the crude and refined bounds are equal. In the log log scale plots, Figures 3.5

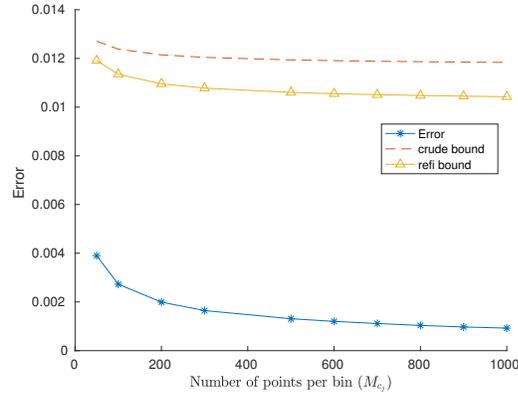
### 3.2. ERROR ANALYSIS

---

and 3.6, we can fit a linear regression for the  $\hat{m}_1$  and  $\hat{m}_2$ . The slope of the  $\hat{m}_1$  error is  $-0.497 \pm 0.013$  and constant is  $-1.665 \pm 0.079$ . The slope of the  $\hat{m}_2$  error is  $-0.492 \pm 0.008$  and constant is  $-3.603 \pm 0.051$ . The slopes of the errors of  $\hat{m}_1$  and  $\hat{m}_2$  are close to  $-0.5$ , what confirms the decay rate of  $\frac{1}{\sqrt{M_{c_j}}}$ .



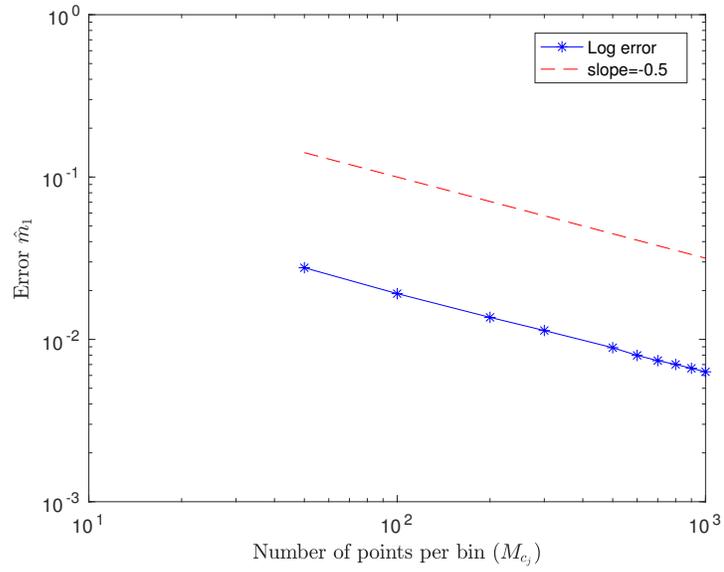
**Figure 3.3:** Error and bounds for  $\hat{m}_1$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . The number of points per bin,  $M_{c_j}$ , is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The errors are closely bounded by the refined bound whereas the crude one is about five times higher than the error and refined bound.



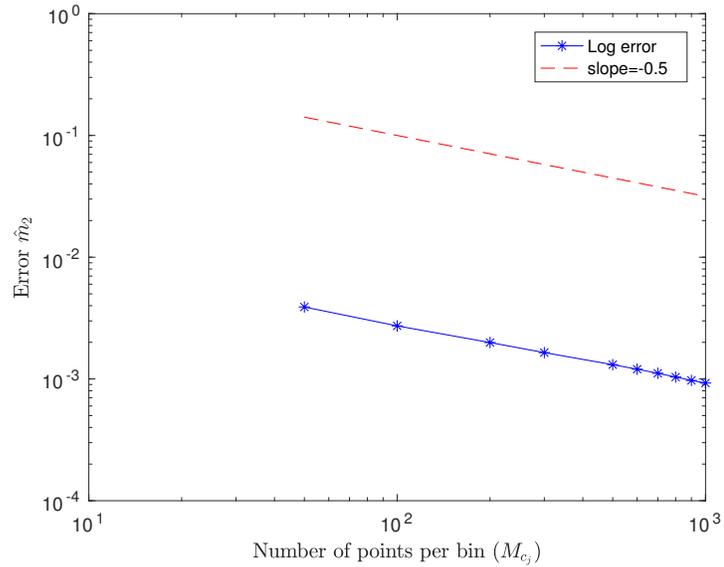
**Figure 3.4:** Error and bounds for  $\hat{m}_2$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . The number of points per bin,  $M_{c_j}$ , is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The errors are relatively smaller than both crude and refined bounds.

### 3.2. ERROR ANALYSIS

---



**Figure 3.5:** Log error vs  $M_{c_j}$ ,  $\hat{m}_1$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . In the log log scale, the number of bins is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The  $\frac{1}{\sqrt{M}}$  decay rate is confirmed for  $\hat{m}_1$ .



**Figure 3.6:** Log error vs  $M_{c_j}$ ,  $\hat{m}_2$ ,  $\Delta x = 0.039801$ ,  $\Delta t = 0.01$ . In the log log scale, the number of bins is given in the x axis and the corresponding errors in the y axis. The estimation errors are in star line. The  $\frac{1}{\sqrt{M}}$  decay rate is confirmed for  $\hat{m}_2$ .

### 3.2. ERROR ANALYSIS

---

**Test 2: Fixed subsampling time step,  $\Delta t = 0.01$ , and  $\Delta x = \frac{1}{M_{c_j}}$**

We would like to test the optimal scaling when the subsampling time is fixed and the spatial time step is the inverse of the number of points in the bin. The number of bins is not fixed but changes with  $\Delta x$ . The number of points in a bin centered at  $c_j$  is denoted by  $M_{c_j}$ . In this test, we look at the errors for  $M_{c_j} = \{50, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ . The number of simulated paths was set at  $N_{sim} = 2000$ . Table 3.1 gives the parameters used to compute the errors of the different paths.

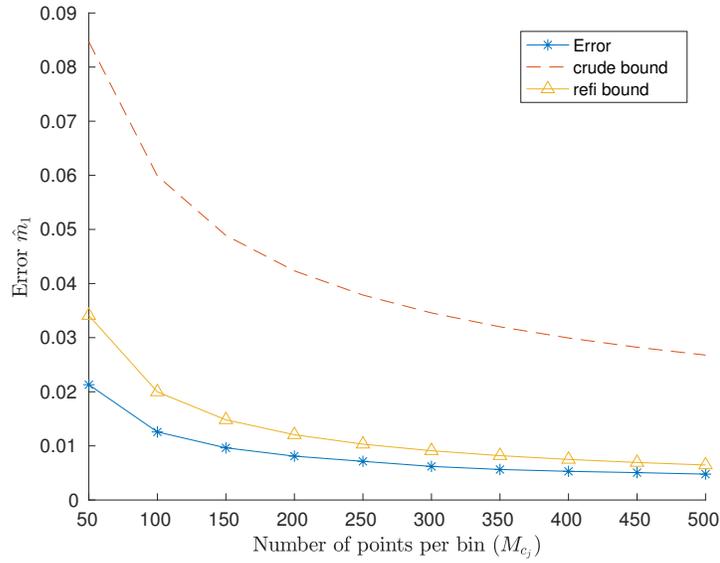
$M_{c_j}$	$\Delta t$	$\Delta x$
50	0.01	0.022112
100	0.01	0.010474
150	0.01	0.006862
200	0.01	0.005103
250	0.01	0.004061
300	0.01	0.003373
350	0.01	0.002884
400	0.01	0.002519
450	0.01	0.002236
500	0.01	0.002010

**Table 3.1:** This table gives us the parameters used to test optimal scaling when  $\Delta t$  is fixed and  $\Delta x = \frac{1}{M_{c_j}}$ .

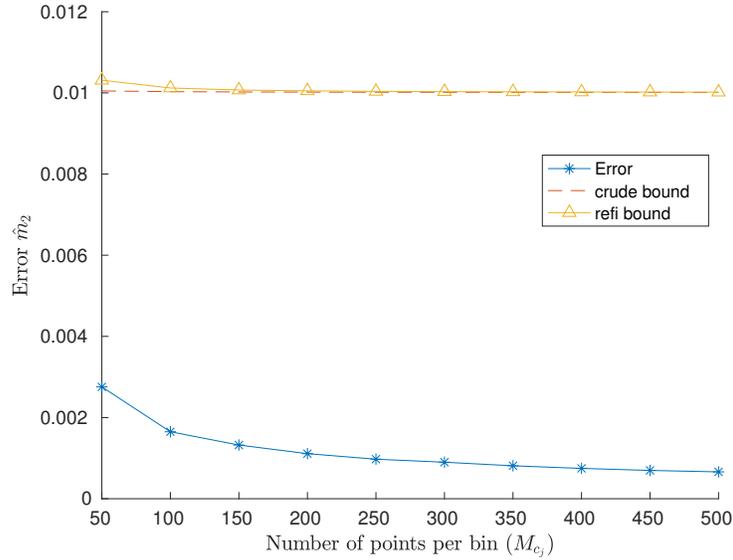
At  $M_{c_j} = 500$ , the error of the  $\hat{m}_1$ ,  $error_{b_j}$ , is 0.00479. The refined bound is higher at 0.006477 while the crude estimator is larger at 0.02689. Similarly for the  $\hat{m}_2$ , when  $M_{c_j} = 500$ , the error  $error_{a_j}$  is 0.000672 when both bounds are 0.01005. In the log scale, the slope of the errors of  $\hat{m}_1$  is  $-0.649 \pm 0.043$  and the constant is  $-1.329 \pm 0.236$ . The slope of the errors of  $\hat{m}_2$  is  $-0.692 \pm 0.068$  and the constant is  $-3.075 \pm 0.371$ . Choosing  $\Delta x \propto \frac{1}{M_{c_j}}$  makes the  $L^2$  errors decay faster.

### 3.2. ERROR ANALYSIS

---



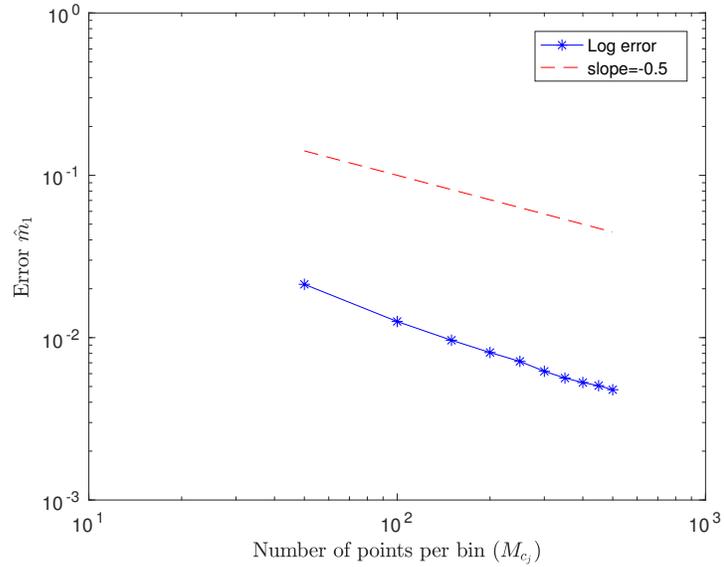
**Figure 3.7:** The errors for  $\hat{m}_1$  is plotted with the two bounds. The subsampling time  $\Delta t$  is fixed at 0.01 and the spatial step size is  $\Delta x = \frac{1}{M_{c_j}}$ . Both bounds are above the error values.



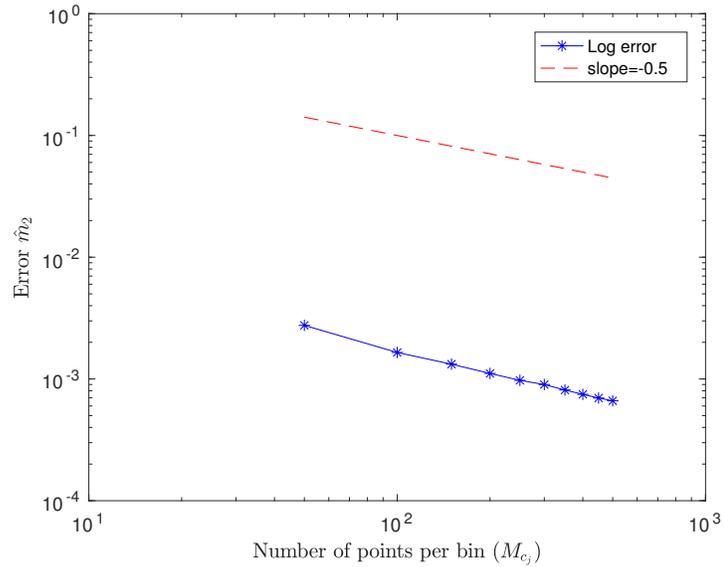
**Figure 3.8:** The errors for  $\hat{m}_2$  is plotted with the two bounds. The subsampling time  $\Delta t$  is fixed at 0.01 and the spatial step size is  $\Delta x = \frac{1}{M_{c_j}}$ . Again both bounds are close to the value of the time step  $\Delta t$ . The errors are smaller than the bounds.

### 3.2. ERROR ANALYSIS

---



**Figure 3.9:** The error of  $\hat{m}_1$  in the log log scale when  $\Delta t = 0.01$  and  $\Delta x = \frac{1}{M_{c_j}}$ . The slope of the error is  $-0.649 \pm 0.043$ . The slope of the log errors of  $\hat{m}_1$  being smaller than  $-0.5$  implies that the estimation errors decay a little bit slower than expected.



**Figure 3.10:** The error of  $\hat{m}_2$  in the log log scale when  $\Delta t = 0.01$  and  $\Delta x = \frac{1}{M_{c_j}}$ . The slope of the error is equal to  $-0.692 \pm 0.068$ . The slope of the log errors of  $\hat{m}_2$  being smaller than  $-0.5$  implies that the estimation errors decay a little bit slower than expected.

### 3.2.5 Relaxing assumption of the fixed number of points

In order to analyze the estimation errors, of  $\hat{m}_1$  and  $\hat{m}_2$ , when the assumption of a fixed number of points has been relaxed, we change the definition of the variable  $M$ . The total number of data points generated from the process (3.7) is now denoted by  $M$ , and the observations are given by  $X_i$  for  $i = 1 \dots M$ . The connection with the previous notation is that we now use  $N = M$ . For each sample point, we observe that  $X_i \sim N(0, \sigma_{X_i}^2)$  with  $\sigma_{X_i}^2 = \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma i \Delta t})$ . The spatial discretization is done using bins defined as  $B_{c_j} = [c^-, c^+] = [c_j - \frac{\Delta x}{2}, c_j + \frac{\Delta x}{2}]$ . In this set up, the first moment estimator is given by

$$\hat{m}_1 = \frac{\sum_{i=1}^M X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}. \quad (3.35)$$

When we relax the assumption of a fixed number of points in each bin, the denominator of  $\hat{m}_1$  can be written as a summation of indicator functions like in (3.35). Since the non stationary case is more complicated, we restricted our analysis of this estimator for a stationary OU process, where each observation follows a Gaussian  $N(0, \sigma_X^2)$ . Before studying the distribution of  $\hat{m}_1$ , we discuss separately about the distribution of the denominator and numerator of  $\hat{m}_1$

#### 3.2.5.1 Denominator

The denominator in (3.35) represents the total number of times the process visits the bin,  $B_{c_j}$ . Notice that the value of each indicator function,  $\mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$ , of this denominator is either zero or one. It can be represented by a random variable that follows a Bernoulli distribution whose parameter is denoted by  $p_{B_{c_j}}$ , the probability

### 3.2. ERROR ANALYSIS

---

of the observation  $X_{i-1}$  to be in the bin  $B_{c_j}$ . In other terms,  $p_{B_{c_j}}$  is the probability of the event  $\{X_{i-1} \in B_{c_j}\}$ . The OU process follows a normal distribution so this probability can be obtained by computing the difference  $\Phi(c^+) - \Phi(c^-)$  where  $\Phi(\cdot)$  is the cumulative normal distribution. Using the independence of observations, we can write the sum of Bernoulli as a binomial distribution with the parameters,  $M$  and  $p_{B_{c_j}}$ . In other words, we can write  $\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} \sim Bin(M, p_{B_{c_j}})$ . We omit the subscript to simplify the notation of the probability, i.e. we use  $p$  instead of  $p_{B_{c_j}}$ . When the number of observations,  $M$ , is very large, we can use the Central Limit Theorem (CLT), a normal approximation of the binomial distribution, [67, 68]. When  $M$  is sufficiently large, by CLT we can use the convergence in distribution:

$$Bin(M, p) \longrightarrow N(Mp, Mp(1 - p)). \quad (3.36)$$

#### 3.2.5.2 Numerator

Unlike the case where the number of points is fixed and a sum of truncated normal distributions could be used, the exact number of points in the bin is not known. As argued for the denominator of the estimator (3.35), the indicator function is a random variable that could be represented by a Bernoulli distribution with parameter,  $p$ . We examine the sum of products of Bernoulli and Normal distributions in the numerator. The product of a discrete and continuous density is called a mixed distribution [53]. Similar work has been done by [22] for the sum of independent product of Bernoulli and exponential distribution. Following [22], we derive the density of the numerator of (3.35).

### Product of Bernoulli and Gaussian distributions

We consider the data generated by an OU process,  $X_i$ ,  $i = 1 \dots M$ , to define two new random variables,  $Y_i = \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$  and  $Z_i = X_i Y_i$ . We know that  $P(Y_i = 1) = p$  and  $P(Y_i = 0) = 1 - p$ . By Bayes theorem, [53, 68], we have:

$$\begin{aligned} P(Z_i \leq z) &= P(Y_i = 0)P(Z_i \leq z|Y_i = 0) + P(Y_i = 1)P(Z_i \leq z|Y_i = 1), \\ &= (1 - p) P(Z_i \leq z|Y_i = 0) + p P(Z_i \leq z|Y_i = 1). \end{aligned}$$

The first probability,  $P(Z_i \leq z|Y_i = 0)$ , could be understood as follows. When  $Y_i = 0$  the product  $Z_i$  is equal to zero and if  $z < 0$  then  $P(Z_i \leq z|Y_i = 0) = 0$ . However, if  $z \geq 0$ , the probability is one. We then have

$$P(Z_i \leq z|Y_i = 0) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0. \end{cases} \quad (3.37)$$

This function is called the step or Heaviside function and it is denoted here by  $H(z)$ . The same argument can be used for  $Y_i = 1$ , but in both cases the solution is equal to the cumulative of the variable  $X_i$  that we denoted  $\Phi(X)$ . We therefore have  $P(Z_i \leq z|Y_i = 1) = \Phi(X)$ . The cumulative distribution  $F_{Z_i}$  is

$$\begin{aligned} F_{Z_i}(z) = P(Z_i \leq z) &= (1 - p) P(Z_i \leq z|Y_i = 0) + p P(Z_i \leq z|Y_i = 1), \\ &= (1 - p) H(z) + p F_{X_i}(z). \end{aligned}$$

where  $F_{X_i}(z)$  is the cumulative of the normally distributed process,  $X_i$ . By differentiating the cumulative with respect to  $z$ , assuming that the derivative of the Heaviside function is the Dirac delta distribution  $\frac{\partial H(z)}{\partial z} = \delta(z)$ , we obtain the following probability density  $f_{Z_i}$ :

$$f_{Z_i}(z) = P(Z_i \leq z) = (1 - p) \delta(z) + p f_{X_i}(z),$$

where  $f_{X_i}(z)$  is the normal density function with a mean of zero and a variance of  $\sigma_X^2$ .

### Sum of independent products of Bernoulli and Gaussian distributions

We consider the distribution of the sum of M products of Zs i.e.

$$S_M = \sum_{k=1}^M Z_k. \quad (3.38)$$

Note that the data are not independent. Indeed the covariance of any  $Z_k$  and  $Z_l$ , for  $k$  and  $l$  two integers less than M, would be equal to either zero or  $E[X_k X_l]$ . The latter is given in equation (3.6) and is not generally zero. But we argued in the previous section that the points in one bin could be assumed independent if one uses a sufficiently small bin size. Using this fact, we assume independence of the nonzero points  $Z_k$  in the bin,  $k = 1 \dots M$ . Similarly to [22], we can prove that the density of our mixed distribution has the following form:

$$f_{S_M}(x) = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} f_{X^{*k}}(x)$$

where  $f_{X^{*0}}(x) = \delta(x)$  and for  $k \geq 1$  we have that  $f_{X^{*k}}(x)$  the k-fold convolution of normal distribution. Recall that  $X^{*k}$  means  $X_1 * X_2 * \dots * X_k$  and for independent normal distribution  $X^{*k} \sim N(0, k\sigma_X^2)$ . For details of the derivation of this density, its distribution and moments, we refer to [22] since our results mirror theirs. The  $n^{\text{th}}$  moment of the sum (3.38) would be:

$$E[S_M^n] = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} E[(X^{*k})^n]$$

### 3.2. ERROR ANALYSIS

---

Since the  $X^{*k} \sim N(0, k\sigma_X^2)$  the moments of  $k$  fold convolutions could be written as :

$$E[(X^{*k})^n] = \begin{cases} 0 & n \text{ is odd} \\ (\sigma_X \sqrt{k})^n (n-1)!! & n \text{ is even} \end{cases} \quad (3.39)$$

The double factorial,  $!!$ , in this expression is defined as

$$n!! = \begin{cases} n(n-2) \dots 5.3.1 & n > 0 \text{ and } n \text{ is odd,} \\ n(n-2) \dots 6.4.2 & n > 0 \text{ and } n \text{ is even,} \\ 1 & n = 0, -1. \end{cases} \quad (3.40)$$

Using this definition, the  $n^{\text{th}}$  moments are then defined as

$$E[S_M^n] = \begin{cases} 0 & n \text{ is odd} \\ \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} (\sigma_X \sqrt{k})^n (n-1)!! & n \text{ is even} \end{cases} \quad (3.41)$$

When  $n = 1$  we have  $E[S_M^n] = 0$ . The variance, or second moment in this case, is obtained by plugging  $n = 2$  in the above expectation and using a change of variable what gives:

$$E[S_M^2] = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} k\sigma_X^2 = Mp\sigma_X^2. \quad (3.42)$$

Following the same derivation as in [22], we found that the moment generating function  $M_{SM}(t)$  has the form

$$M_{SM}(t) = \left[ (1-p) + p e^{\frac{\sigma_X^2 t^2}{2}} \right]^M. \quad (3.43)$$

Using the CLT, [67, 68], this density is approximately Gaussian when  $M$  is large enough. In other terms, we can write that

$$f_{S_M}(x) \longrightarrow N(0, pM \sigma_X^2). \quad (3.44)$$

### 3.2.5.3 Distribution of the estimator $\hat{m}_1$

The estimator  $\hat{m}_1$  is a ratio of two random variables that can both be separately approximated by a normal distribution. We can write schematically:

$$\hat{m}_1 = \frac{\sum_{i=1}^M X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \sim \frac{N(0, pM \sigma_X^2)}{N(Mp, Mp(1-p))} \quad (3.45)$$

In this particular case, the denominator cannot be zero, since the number of elements in the bin is an integer number between 1 and  $M$ , to produce a strictly positive normally distributed variable in the denominator. The exclusion of non-negative values for the denominator would result in a truncated Cauchy distribution which unlike the classical one has all its moments [62, 63]. The problem with this approach is that we do not have a ratio of standard normal variables. Instead, we used another approach based on the normal approximation of the ratio of Gaussian variables, [21, 31, 43, 47, 57]. The idea is as follows. Take two random variables,  $X$  and  $Y$ , with respective distributions,  $N(\mu_x, \sigma_x^2)$  and  $N(\mu_y, \sigma_y^2)$ . Suppose that you are interested in the distribution of  $Z = \frac{X}{Y}$ . Key parameters according to [21] are the variables,  $(\beta, \sigma_x, \rho, \delta_y)$ , where,  $\beta = \frac{\mu_x}{\mu_y}$ ,  $\rho = \frac{\sigma_y}{\sigma_x}$ , and  $\delta_y = \frac{\sigma_y}{\mu_y}$ . The joint density of  $Y$  and  $Z$  is  $f_{Y,Z}$  and the marginal density of  $Z$  is,  $f_Z$ . The joint density can be expressed as :

$$f_{Y,Z}(y, z; \beta, \rho, \delta_y, \sigma_x) = f_Z(z; \beta, \rho, \delta_y) f_{Y|Z}(y|z; \beta, \sigma_x, \rho, \delta_y). \quad (3.46)$$

The conditional density is Gaussian but what we are looking for is the first term in the right hand-side. Its expression is given by (see [21, 57]):

$$f_Z(z; \beta, \rho, \delta_y) = \frac{\rho}{\pi(1 + \rho^2 z^2)} \left\{ \exp\left[-\frac{\rho^2 \beta^2 + 1}{2\delta_y^2}\right] + \sqrt{\frac{\pi}{2}} \rho \operatorname{erf}\left(\frac{\rho}{\sqrt{2}}\right) \exp\left[-\frac{\rho^2(z - \beta)^2}{2\delta_y^2(1 + \rho^2 z^2)}\right] \right\} \quad (3.47)$$

### 3.2. ERROR ANALYSIS

---

with  $q = \frac{(1+\beta\rho^2z)}{\delta_y\sqrt{1+\rho^2z^2}}$ . The cumulative distribution of  $Z$ , is denoted by  $F_Z$  and can be approximated by a simpler distribution than  $F_Z$ . It's shown in [21,47] that when  $\delta_y \rightarrow 0$ , we can find a function  $F^*(z)$  that is converging uniformly to  $F_Z(z)$ . This function introduced by [47] is defined as

$$F^*(z) = \Phi\left(\frac{z\mu_y - \mu_x}{\sqrt{\sigma_x^2 + z^2\sigma_y^2}}\right) = \Phi\left(\frac{z - \beta}{\delta_y\sqrt{\rho^{-2} + z^2}}\right). \quad (3.48)$$

Another important result in [47] is the upper bound of the absolute difference of  $F_Z$  and  $F^*$  depends on  $\delta_y$ , i.e.

$$|F_Z(z; \beta, \rho, \delta_y) - F^*(z)| \leq \Phi\left(-\delta_y^{-1}\right) \quad (3.49)$$

An important limitation of  $F^*$  is it is not a probability distribution since

$$F^*(-\infty) = \Phi(-\delta_y^{-1}); \quad F^*(\infty) = 1 - \Phi(-\delta_y^{-1}). \quad (3.50)$$

A normal approximation to  $F^*$  and  $F_Z$  has been presented in [21] and is summarized by the following theorem:

**Theorem 3.2.2.** *Theorem 1 in Dias et Frances et al. (2013) [21] Let  $X$  be a normal random variable with positive mean  $\mu_x$ , variance  $\sigma_x^2$  and coefficient of variation  $\delta_x$  such that  $0 < \delta_x < \lambda \leq 1$ ,  $\lambda$  is a known constant . For every  $\epsilon > 0$ , there exists  $\gamma(\epsilon) \in (0, \sqrt{\lambda^2 - \delta_x^2})$  and also a normal random variable  $Y$  independent of  $X$ , with positive mean  $\mu_y$  variance  $\sigma_y^2$  and coefficient of variation  $\delta_y$  that satisfies the conditions,  $0 < \delta_y \leq \gamma(\epsilon) \leq \sqrt{\lambda^2 - \delta_x^2} < \lambda$  for which the following result holds: for any  $z$  in  $I = [\beta - \frac{\sigma_z}{\lambda}, \beta + \frac{\sigma_z}{\lambda}]$  we have that  $|G(z) - F_Z(z)| < \epsilon$ .*

This theorem is fundamental for the study of our estimators. Different remarks can be made. First notice that this result is only valid in a specific interval  $I$ . The

### 3.2. ERROR ANALYSIS

---

second remark is that the theorem cannot be used for the bin centered at zero. This leads to a third remark where the bins under consideration should only be centered at positive values. When the center of a bin  $c_j > 0$ , the mean of the numerator is negative and the theorem cannot be used. In that case, we can restrict our analysis to bins centered at positive values. The distribution of an OU process with a zero long-term mean is symmetric around zero so bins with negative centers should be filled in the same way as positive bins. Therefore, we can use the theorem assuming  $c_j < 0$  in the following lines.

$$\begin{aligned}
\|\hat{m}_1 - E[X_{t+\Delta t}|X_t = c_j]\|_2 &= \|\hat{m}_1 - c_j e^{-\gamma\Delta t}\|_2 \\
&= \left\| \frac{\sum_{i=1}^M X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} - c_j e^{-\gamma\Delta t} \right\|_2 \\
&= \left\| \frac{\sum_{i=1}^M (X_i - c_j e^{-\gamma\Delta t}) \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \right\|_2
\end{aligned} \tag{3.51}$$

The distribution of this ratio has parameters  $\mu_x = -c_j e^{-\gamma\Delta t}$ ,  $\mu_y = Mp$ ,  $\sigma_x^2 = Mp\sigma_X^2$  and  $\sigma_y = Mp(1-p)$ . According to Theorem 3.2.2, the distribution of the ratio in the interval  $I = [\beta - \frac{\sigma_z}{\lambda}, \beta + \frac{\sigma_z}{\lambda}]$  is  $N(\beta, \sigma_z^2)$  where

$$\begin{aligned}
\beta &= \frac{\mu_x}{\mu_y} = \frac{-c_j e^{-\gamma\Delta t}}{Mp} \\
\sigma_z^2 &= \beta^2(\delta_x^2 + \delta_y^2) = \left( \frac{-c_j e^{-\gamma\Delta t}}{Mp} \right)^2 \left( \frac{\sigma_x^2}{(-c_j e^{-\gamma\Delta t})^2} + \frac{\sigma_y^2}{Mp(1-p)} \right) \\
&= \left( \frac{-c_j e^{-\gamma\Delta t}}{Mp} \right)^2 \left( \frac{Mp\sigma_X^2}{(-c_j e^{-\gamma\Delta t})^2} + \frac{Mp(1-p)}{(Mp)^2} \right)
\end{aligned} \tag{3.52}$$

The ratio in (3.51) has a mean equal to  $\frac{-c_j e^{-\gamma\Delta t}}{Mp}$  and a standard deviation is given

### 3.2. ERROR ANALYSIS

---

using

$$\begin{aligned}
\sigma_z &= \sqrt{\left(\frac{-c_j e^{-\gamma \Delta t}}{Mp}\right)^2 \left(\frac{Mp\sigma_X^2}{(-c_j e^{-\gamma \Delta t})^2} + \frac{Mp(1-p)}{(Mp)^2}\right)} \\
&= \left(\frac{-c_j e^{-\gamma \Delta t}}{Mp}\right) \sqrt{\frac{Mp\sigma_X^2}{(-c_j e^{-\gamma \Delta t})^2} + \frac{1-p}{Mp}} \\
&= \left(\frac{1}{Mp}\right) \sqrt{Mp\sigma_X^2 + \frac{-c_j e^{-\gamma \Delta t}(1-p)}{Mp}} \\
&= \sqrt{\frac{\sigma_X^2}{Mp} + \frac{-c_j e^{-\gamma \Delta t}(1-p)}{(Mp)^3}}
\end{aligned} \tag{3.53}$$

Taylor expansion to the first order of the process variance yields  $\sigma_X^2 = \frac{\sigma^2(1-e^{2\gamma\Delta t})}{2\gamma} \approx \sigma^2\Delta t$ . The value,  $Mp$ , approximates the number of points in a bin that we assumed is sufficiently large when we consider points of the OU that are close to the long-run mean. When  $M$  is large, we neglected the last term in the standard deviation and obtain

$$\sigma_z = \sqrt{\frac{\sigma_X^2}{Mp} + \frac{-c_j e^{-\gamma \Delta t}(1-p)}{(Mp)^3}} \approx \sqrt{\frac{\sigma_X^2}{Mp}} = \frac{\sigma_X}{\sqrt{Mp}} \tag{3.54}$$

The probability that a sample point is in a bin centered at  $c_j$ ,  $p = P_{c_j} = \Phi(c_j + \frac{\Delta x}{2}) - \Phi(c_j - \frac{\Delta x}{2})$ , can be approximated using the mean value theorem as

$$\begin{aligned}
p &= \Phi\left(c_j + \frac{\Delta x}{2}\right) - \Phi\left(c_j - \frac{\Delta x}{2}\right) \\
&= \int_{c_j - \frac{\Delta x}{2}}^{c_j + \frac{\Delta x}{2}} \phi(t) dt \\
&\stackrel{*}{=} \phi(t^*) \Delta x \quad \text{using integral mean value theorem}
\end{aligned}$$

where  $t^* \in [c_j - \frac{\Delta x}{2}, c_j + \frac{\Delta x}{2}]$ . Hence, the variance of the ratio is given by

$$\sigma_z = \frac{\sigma_X}{\sqrt{Mp}} = \frac{\sigma\sqrt{\Delta t}}{\sqrt{M\phi(t^*)\Delta x}} = \frac{\sigma}{\sqrt{M\phi(t^*)}} \sqrt{\frac{\Delta t}{\Delta x}}. \tag{3.55}$$

$$\begin{aligned}
 \|\hat{m}_1 - c_j e^{-\gamma \Delta t}\|_2 &= \left\| \frac{\sum_{i=1}^M (X_i - c_j e^{-\gamma \Delta t}) \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \right\|_2 \\
 &= \sqrt{\left( \frac{\sigma}{\sqrt{M\phi(t^*)}} \sqrt{\frac{\Delta t}{\Delta x}} \right)^2 + \left( \frac{-c_j e^{-\gamma \Delta t}}{Mp} \right)^2} \quad (3.56) \\
 &\approx \frac{\sigma}{\sqrt{M\phi(t^*)}} \sqrt{\frac{\Delta t}{\Delta x}} \quad (\text{for } Mp \text{ large})
 \end{aligned}$$

1. For  $\hat{m}_1$ , when  $\Delta t$  is fixed, the error would decay at a rate of  $\frac{1}{\sqrt{M}}$  if the spatial step size is chosen to be equal to the subsampling time step. So  $\Delta x \propto \Delta t$
2. For  $\hat{m}_1$ , when  $\Delta t \rightarrow 0$ , we would need the spatial step  $\Delta x$  to go to zero slower than  $\Delta t$ . So if  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{\sqrt{M}}$ , the error would scale at the rate of  $\frac{1}{M}$ .

**We can see that for  $\hat{m}_1$ , for fixed  $\Delta t$ , a good choice of  $\Delta x$  would be  $\Delta x = \Delta t$ . In this case, the error would decay at a rate of  $\frac{1}{\sqrt{M}}$ . When  $\Delta t \rightarrow 0$ , we would need the spatial step  $\Delta x$  to reach zero slower than  $\Delta t$ . So if  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{\sqrt{M}}$ , the error would scale at a rate of  $\frac{1}{M}$ .**

### Sum of independent products of Bernoulli and Chi square distributions

Similar analysis can be done for the second estimator defined as

$$\hat{m}_2 = \frac{\sum_{i=1}^M X_i^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \quad (3.57)$$

The only difference is that now the square of the observations is used in the numerator. Since each observation follows a Gaussian distribution, the square of a

### 3.2. ERROR ANALYSIS

---

standard normal distribution is a Chi square. Here, we have a non-central Chi square where  $\forall i, \sigma_X^2 X_i^2 \sim \chi_1$ . We rewrote the estimator

$$\begin{aligned} \hat{m}_2 &= \frac{\sum_{i=1}^M \left( \frac{\sigma_X X_i}{\sigma_X} \right)^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} = \frac{\sum_{i=1}^M \sigma_X^2 \left( \frac{X_i}{\sigma_X} \right)^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \\ &= \frac{\sigma_X^2 \sum_{i=1}^M U_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \end{aligned} \quad (3.58)$$

where  $U_i = \frac{X_i^2}{\sigma_X^2}$  and  $U_i \sim \chi_1$ . We then have the sum of the product of Chi square and Bernoulli distributions. We just mirrored the approach for the sum of independent product of Normal-Bernoulli and derive the density, distribution, and moments. We have a k-fold convolution of Chi square of degree 1 that gives us a Chi square with degree equal to k. We define the sum as  $S_M = \sum_{i=1}^M U_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$  and following the approach in [22], we can show that the density of our mixed distribution has the following form:

$$f_{S_M}(x) = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} f_{U^{*k}}(x)$$

where  $f_{U^{*0}}(x) = \delta(x)$  and  $\forall k \geq 1$ . We have that  $f_{U^{*k}}(x)$ , the k-fold convolution of the Chi square distributions, such that  $U^{*k} \sim \chi_k$ . As shown previously, the  $n^{\text{th}}$  moment could be computed as

$$E[S_M^n] = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} E[(U^{*k})^n]$$

The expectation on the right-hand side is easy to handle, since the moments of a Chi square with k degrees of freedom, denoted by  $\chi_k$ , are well-known:

$$E[(U^{*k})^n] = k(k+2)(k+4) \dots (k+2n-2) = 2^n \frac{\Gamma(n + \frac{k}{2})}{\Gamma(\frac{k}{2})} \quad (3.59)$$

### 3.2. ERROR ANALYSIS

---

By the binomial theorem and using the moments of  $\chi_k$ , when  $n = 1$ ,  $E[S_M] = Mp$ ; while when  $n = 2$ ,  $E[S_M^2] = Mp(1-p) + (Mp)^2 + 2Mp$  and the variance,  $Var[S_M] = Mp(3-p)$ . The  $n^{th}$  moments formula is given by

$$E[S_M^n] = \sum_{k=0}^M \binom{M}{M-k} p^k (1-p)^{M-k} 2^n \frac{\Gamma(n + \frac{k}{2})}{\Gamma(\frac{k}{2})} \quad (3.60)$$

Following the approach in [22], the moment generating function  $M_{S_M}(t)$  has the form

$$M_{S_M}(t) = \left[ (1-p) + p \frac{1}{(1-2t)^{\frac{k}{2}}} \right]^M, \quad t < \frac{1}{2} \quad (3.61)$$

As the Normal-Bernoulli case, when  $M \rightarrow \infty$  or is large, we use the CLT [67, 68] to give an approximation of the distribution of  $S_M$ :

$$f_{S_M}(x) \longrightarrow N(Mp, Mp(3-p)) \quad (3.62)$$

#### 3.2.5.4 Distribution of the estimator $\hat{m}_2$

The distribution of the ratio of the estimator,  $\hat{m}_2$ , after approximating the numerator and denominator by a normal is

$$\begin{aligned} \hat{m}_2 &= \frac{\sigma_X^2 \sum_{i=1}^M U_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} = (\sigma_X^2) \frac{\sum_{i=1}^M U_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \\ &\sim \sigma_X^2 \frac{N(Mp, Mp(3-p))}{N(Mp, Mp(1-p))} = \frac{N(Mp, Mp(3-p)\sigma_X^4)}{N(Mp, Mp(1-p))} \end{aligned} \quad (3.63)$$

$$\Rightarrow \hat{m}_2 \sim \frac{N(Mp, Mp(3-p)\sigma_X^4)}{N(Mp, Mp(1-p))} \quad (3.64)$$

Exactly like  $\hat{m}_1$ , let X and Y be the numerator and denominator of the distribution of  $\hat{m}_2$  with  $\mu_x = Mp$ ,  $\mu_y = Mp$ ,  $\sigma_x^2 = Mp(3-p)\sigma_X^4$ , and  $\sigma_y^2 = Mp(1-p)$ . The ratio

### 3.2. ERROR ANALYSIS

---

is again denoted by  $Z = \frac{X}{Y}$  and the parametrization:

$$\begin{aligned}
\beta &= \frac{\mu_x}{\mu_y} = 1 \\
\rho &= \frac{\sigma_y}{\sigma_x} = \frac{1}{\sigma_X^2} \sqrt{\frac{Mp(1-p)}{Mp(3-p)}} = \frac{1}{\sigma_X^2} \sqrt{\frac{1-p}{3-p}} = \frac{1}{\sigma_X^2} \sqrt{1 - \frac{2}{3-p}} \\
\delta_x &= \sigma_X^2 \sqrt{\frac{3-p}{Mp}} \\
\delta_y &= \sqrt{\frac{1-p}{Mp}}
\end{aligned} \tag{3.65}$$

The 2-norm that we would like to study is:

$$\begin{aligned}
\|\hat{m}_2 - E[(X_{t+\Delta t})^2 | X_t = c_j]\|_2 &= \left\| \hat{m}_2 - \left( \sigma_X^2 + c_j^2 e^{-2\gamma\Delta t} \right) \right\|_2 \\
&= \left\| \frac{\sigma_X^2 \sum_{i=1}^M U_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} - \left( \sigma_X^2 + c_j^2 e^{-2\gamma\Delta t} \right) \right\|_2 \\
&= \left\| \frac{\sum_{i=1}^M \left( \sigma_X^2 U_i - (\sigma_X^2 + c_j^2 e^{-2\gamma\Delta t}) \right) \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \right\|_2
\end{aligned} \tag{3.66}$$

The above term is the ratio of normal random variables

$$\frac{N(Mp - (\sigma_X^2 + c_j^2 e^{-2\gamma\Delta t}), Mp(3-p)\sigma_X^4)}{N(Mp, Mp(1-p))}.$$

Also, we assume that  $p > \frac{(\sigma_X^2 + c_j^2 e^{-2\gamma\Delta t})}{M}$  in order for the mean of the numerator to be positive. Using the Theorem 3.2.2, the distribution of the ratio is  $N(\beta, \sigma_z^2)$  where

$$\begin{aligned}
\beta &= \frac{\mu_x}{\mu_y} = 1 \\
\sigma_z^2 &= \beta^2(\delta_x^2 + \delta_y^2) = \sigma_X^4 \left( \frac{3-p}{Mp} \right) + \frac{1-p}{Mp}
\end{aligned} \tag{3.67}$$

### 3.2. ERROR ANALYSIS

---

The 2 norm is then

$$\begin{aligned}
& \|\hat{m}_2 - E[(X_{t+\Delta t})^2 | X_t = c_j]\|_2 \\
&= \left\| \frac{\sum_{i=1}^M \left( \sigma_X^2 U_i - (\sigma_X^2 + c_j^2 e^{-2\gamma\Delta t}) \right) \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}} \right\|_2 \\
&= \sqrt{\sigma_X^4 \left( \frac{3-p}{Mp} \right) + \frac{1-p}{Mp} + 1} \\
&\leq \sqrt{1 + \sigma_X^4 \left( \frac{3}{Mp} \right) + \frac{1}{Mp}} \quad , \text{ since } 0 < p < 1 \\
&\leq \sqrt{1 + (1 + \sigma_X^4) \left( \frac{3}{Mp} \right)} \approx \sqrt{1 + \frac{3(1 + (\sigma^2\Delta t)^2)}{Mp}} \\
&\approx \sqrt{1 + \frac{3(1 + (\sigma^2\Delta t)^2)}{M\phi(t^*)\Delta x}}
\end{aligned} \tag{3.68}$$

The last two lines were derived using a Taylor expansion of the variance around  $\Delta t$  and an approximation of the probability distribution using the mean value theorem.

1. For  $\hat{m}_2$ , when  $\Delta t$  is fixed, the error won't reach zero. If  $\Delta x \propto \frac{1}{M}$ , the bound depends on  $\Delta t$ .
2. For  $\hat{m}_2$ , When  $\Delta t \rightarrow 0$ , we would need the spatial step,  $\Delta x$ , to approach zero slower than  $\Delta t$  for the error to be bounded. For  $\Delta t \propto \frac{1}{M}$  and  $\Delta x \propto \frac{1}{M}$ , the error would be bounded by  $\sqrt{1 + \frac{1}{M^2}}$ .

In summary, our upper bounds for the errors when the number of points in each bin is not fixed are:

$$\begin{aligned} \|\hat{m}_1 - E[X_{t+\Delta t}|X_t = c_j]\|_2 &\leq \frac{\sigma}{\sqrt{M\phi(t^*)}} \sqrt{\frac{\Delta t}{\Delta x}} \\ \|\hat{m}_2 - E[(X_{t+\Delta t})^2|X_t = c_j]\|_2 &\leq \sqrt{1 + \frac{3(1 + (\sigma^2\Delta t)^2)}{M\phi(t^*)\Delta x}} \end{aligned}$$

For  $\hat{m}_1$ , for a fixed  $\Delta t$ , choosing a  $\Delta x$  equal to  $\Delta t$  makes the error decay at a rate of  $\frac{1}{\sqrt{M}}$ . For  $\hat{m}_2$ , the error will not decay to zero regardless of the choice of parameters  $\Delta t$  and  $\Delta x$ . We conclude for fixed  $\Delta t$  or  $\Delta t \rightarrow 0$ , an optimal choice of  $\Delta x$  that would help to have a fast decay rate of the estimation errors is  $\Delta x \propto \frac{1}{\sqrt{M}}$ .

### 3.3 Conditional expectation based estimators of the drift and diffusion

The two estimators  $\hat{m}_1$  and  $\hat{m}_2$  provide insights on how the space and time discretization steps can affect the first two moments. Our primary objective was to use the information about the conditional expectation to produce the same type of results for the drift and diffusion estimators (3.3). However, applying the same type of derivations were tedious. We chose to circumvent this issue by using a different approach based on the relationship between our estimators,  $\hat{m}_1$  and  $\hat{m}_2$ , and the least square estimates for an OU process. The least squares method has been introduced in the beginning of the nineteenth century with the work of mathematicians such as

Legendre and Gauss. The formulas of least squares coefficients have been studied for a long time and are today part of the basic toolbox of any mathematician [40]. Indeed, the ratio estimators we are studying could be seen as a special case of a generalized linear squares.

### Weighted Least Squares connection

Consider the following simple regression:  $y_i = \beta x_i + u_i \quad \forall i = 1 \dots M$ , where  $M$  is the total number of observations,  $y_i$  and  $x_i$  are the independent and dependent variables respectively,  $\beta$  is a constant coefficient, and  $u_i$  is an error term. The Ordinary Least Squares (OLS) method assumes that all the data are of equal importance in the estimation whereas Weighted Least Squares (WLS) allows us to vary the weight of each observation. Assume for example that we have a known matrix of weights,  $W = \{w_i\}, \forall i = 1, \dots, M$ . Each observation can receive a weight in an estimation whose goal would be:

$$\text{Minimize}_{\beta} \sum_{i=1}^M w_i (y_i - \beta x_i)^2. \quad (3.69)$$

By taking the derivative and setting it equal to zero, we obtain the WLS coefficient:

$$\hat{\beta} = \frac{\sum_{i=1}^M w_i x_i y_i}{\sum_{i=1}^M w_i x_i^2} \quad (3.70)$$

Now, if we choose the weight to be equal to the inverse of the value of each observation, e.g.,  $w_i = \frac{1}{x_i} = x_i^{-1}$  we can see that that our estimate is a ratio of sums:

$$\hat{\beta} = \frac{\sum_{i=1}^M y_i}{\sum_{i=1}^M x_i} \quad (3.71)$$

This WLS estimate of  $\beta$  can be related to our estimators (3.10) by replacing first  $x_i$  by  $\mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$  and then  $y_i$  by either  $X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$  or  $X_i^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$  to obtain either  $\hat{m}_1$  or  $\hat{m}_2$ . The inconvenience with WLS is that values of the process close

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

to zero will generate extremely large weights. The second, and crucial, drawback is in econometrics, the weights are positive values whereas in the current setting using  $w_i = x_i^{-1}$  does not guarantee that  $w_i > 0$ . In fact, WLS is a special case of the Generalized Least Squares approach that has been used to estimate models in the presence of heteroskedasticity [40]. In these problems, the weights are strictly positive since they are proportional to the noise variance of each observation. So the use of WLS in our modeling of the estimators (3.10) is a little bit odd since the weights can't be negative.

#### Ordinary Least Squares connection

To avoid this issue, we use the coefficients  $\beta$  from a simple regression done by OLS. The estimator would be the average value, or square value, for each bin and this can be obtained by running a simple linear regression without an intercept:

$$y_i = \beta x_i + u_i \quad \forall i = 1 \dots M. \quad (3.72)$$

Since OLS has been used for over a century, the statistics of the estimator are well known, [40], and could be used to understand our ratio estimators (3.10). In order to estimate the value of our estimators at the point  $x_j$ , we have to look at all the points that fall inside one bin,  $B_{c_j}$ . A linear regression that captures this evaluation involves setting the independent variable  $x_i$  of (3.72) to the dummy variable  $\mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}$ . The dependent variables are given by the product of the dummy by the observation, or their squares. Our two estimators,  $\hat{m}_1$  and  $\hat{m}_2$ , are calculated via the following regressions:

$$\begin{aligned} X_i \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} &= \beta_1 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} + u_{1,i}, \\ X_i^2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} &= \beta_2 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} + u_{2,i}. \end{aligned} \quad (3.73)$$

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

where  $u_{k,i} \sim (0, \sigma_{u_k}^2)$  for  $k = 1, 2$ . Minimizing the squared errors of the regressions (3.73) yields  $\hat{\beta}_1 = \hat{m}_1$  and  $\hat{\beta}_2 = \hat{m}_2$ . The variance of these coefficients do not generally depend on the independent variable [40]:

$$V(\hat{\beta}_k) = \frac{\sigma_{u_k}^2}{\sum_{i=1}^M \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}}}, \quad k = 1, 2 \quad (3.74)$$

#### Scaling of drift and diffusion estimators

Recall, the general form of the SDE:

$$dX_t = b(X_t)dt + a(X_t)dW_t$$

The discretized version of this

$$X_{i+1} - X_i = b(X_i)\Delta t + a(X_i)\sqrt{\Delta t}z_i$$

where  $z_i \sim \mathbb{N}(0, 1)$  and  $i = 1 \dots M$ . A special example of such process is the OU process where  $b(X_i) = b_i = \gamma X_i$  with  $\gamma < 0$  and  $a(X_i) = a_i = \sigma$ .

The impact of the discretization on the two following estimators of the drift and diffusion,  $\hat{b}_j$  and  $\hat{a}_j$ , is investigated:

$$\hat{b}_j = \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}; \quad \hat{a}_j = \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i)^2 \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}. \quad (3.75)$$

#### Upper bound for the drift

Again, we write the drift estimator

$$\hat{b}_j = \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}. \quad (3.76)$$

As argued before, the conditional expectation could be associated to a simple linear regression of the observation on a constant. A bound of the standard deviation

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

of the difference between the estimator and the true value of drift,  $\|\hat{b}_j - b_j\|_2$ , could be found. This bound is explicitly:

$$\begin{aligned}
\|\hat{b}_j - b_j\|_2 &= \left\| \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} - b_j \right\|_2 \\
&= \left\| \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} - b_j \frac{\Delta t \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\Delta t \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2 \\
&= \frac{1}{\Delta t} \left\| \frac{\sum_{i=0}^M (X_{i+1} - X_i - b_j \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2.
\end{aligned} \tag{3.77}$$

Again, note that the term inside the norm is a OLS coefficients that can be obtained by running the following linear regression

$$(X_{i+1} - X_i - b_j \Delta t) \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} = \beta_1 \mathbb{1}_{\{X_{i-1} \in B_{c_j}\}} + u_{1,i}, \tag{3.78}$$

where  $u_{1,i} \sim (0, \sigma_{u_1}^2)$ . The unknown coefficients of this regression are  $\beta_1$  and  $\sigma_{u_1}^2$ . We can calibrate these parameters using information on the discretized SDE. This equation could be written as:

$$X_{i+1} - X_i = b(X_i) \Delta t + a(X_i) \sqrt{\Delta t} z_i = b_j \Delta t + a_j \sqrt{\Delta t} z_i. \tag{3.79}$$

This latter equality mixed continuous and discrete time notation. It takes into account the drift and diffusion estimates obtained from points  $X_i$  over a bin center at  $c_j$ , with a subscript index  $j$  not  $i$ . The expectation is conditional on the observation being in the bin. A subscript, like  $E_j$ , is used to emphasize that the expectation is conditional on being in the bin  $j$ . We abuse notation and omit the subscript. The expectation of (3.79)

$$E((X_{i+1} - X_i - b_j \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}) = \beta_1 \mathbb{1}_{\{X_i \in B_{c_j}\}} + E(u_{1,i}). \tag{3.80}$$

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

One basic assumption of the OLS is that  $E(u_{1,i}) = 0$ . This implies that the expectation of the deviation is  $\beta_1$ . Using this implication and the discretized SDE we can write that the unconditional expectation of (3.78):

$$E((X_{i+1} - X_i - b_j \Delta t)) = \beta_1 = 0. \quad (3.81)$$

Similarly, for the variance of (3.79)

$$\begin{aligned} \text{Var}((X_{i+1} - X_i - b_j \Delta t)) &= E((X_{i+1} - X_i - b_j \Delta t)^2) \\ &= \text{Var}(u_{1,i}) = \sigma_{u_1}^2 \\ &= a_j^2 \Delta t \quad , \text{ which is the variance of an OU process.} \end{aligned} \quad (3.82)$$

The asymptotic variance of the OLS coefficient in (3.78) is as defined in (3.74)

$$V(\hat{\beta}_1) = \frac{\sigma_{u_1}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} = \frac{a_j^2 \Delta t}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}. \quad (3.83)$$

The 2 norm of the drift deviation is

$$\begin{aligned} \|\hat{b}_j - b_j\|_2 &= \frac{1}{\Delta t} \left\| \frac{\sum_{i=0}^M (X_{i+1} - X_i - b_j \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2 = \frac{1}{\Delta t} \sqrt{V(\hat{\beta}_1)}, \\ &= \frac{1}{\Delta t} \sqrt{\frac{\sigma_{u_1}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}} = \frac{1}{\Delta t} \sqrt{\frac{a_j^2 \Delta t}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}, \\ &= \frac{1}{\sqrt{\Delta t}} \sqrt{\frac{a_j^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}} = \frac{\sqrt{\Delta x}}{\sqrt{\Delta t}} \sqrt{\frac{a_j^2}{\Delta x \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}. \end{aligned} \quad (3.84)$$

For an OU process  $a_j$  is a constant so (3.84) is well posed. This expression has two main drivers the ratio  $\sqrt{\frac{\Delta x}{\Delta t}}$  and the term  $\frac{a_j}{\sqrt{\Delta x \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}$ . The denominator of the latter term is a portion of the process histogram. It means that the error is inversely proportional to the number of points in the bin of size  $\Delta x$ . Infrequently visited bins

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

have a large third term and error, while bins regularly visited have a bounded square root term. Another important point is by reducing  $\Delta x$  alone will not reduce the error since a smaller  $\Delta x$  would imply a smaller bin size and a greater possibility of no points in the bin which leads to a higher squared root term.

#### Upper bound for the diffusion

Consider the diffusion estimator:

$$\hat{a}_j^2 = \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i)^2 \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}. \quad (3.85)$$

The 2 norm of the deviation from the true diffusion is expressed as

$$\begin{aligned} \|\hat{a}_j^2 - a_j^2\|_2 &= \left\| \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i)^2 \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} - a_j^2 \right\|_2 \\ &= \left\| \frac{1}{\Delta t} \frac{\sum_{i=0}^M (X_{i+1} - X_i)^2 \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} - a_j^2 \frac{\Delta t \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\Delta t \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2 \\ &= \frac{1}{\Delta t} \left\| \frac{\sum_{i=0}^M ((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2. \end{aligned} \quad (3.86)$$

The corresponding regression for this deviation is

$$((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}} = \beta_2 \mathbb{1}_{\{X_i \in B_{c_j}\}} + u_{2,i}, \quad (3.87)$$

where  $u_{2,i} \sim (0, \sigma_{u_2}^2)$ . Here the unknown coefficients are  $\beta_2$  and  $\sigma_{u_2}^2$ . Again by setting  $E(u_{2,i}) = 0$  would imply that the expectation of the deviation is equal to  $\beta_2$  as shown:

$$E(((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}) = \beta_2 = 0 \quad (3.88)$$

The variance of the diffusion deviation is

$$\begin{aligned} \text{Var}(((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}) &= E((((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}})^2), \\ &= \text{Var}(u_{2,i}) = \sigma_{u_2}^2. \end{aligned} \quad (3.89)$$

### 3.3. CONDITIONAL EXPECTATION BASED ESTIMATORS OF THE DRIFT AND DIFFUSION

---

The asymptotic variance of the OLS coefficient in (3.87) would be:

$$V(\hat{\beta}_2) = \frac{\sigma_{u_2}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}. \quad (3.90)$$

Here, notice that the variance of the coefficient  $\hat{\beta}_2$ , depends on the variance of the error term  $\sigma_{u_2}^2$  that we assumed constant when we chose OLS as an inference method. In econometrics [40], a estimator,  $s^2$ , of this variance exists. It follows a non-central Chi squared distribution with  $M - 1$  degrees of freedom, i.e.  $s^2 \sim \frac{\sigma_u^2}{M-1} \chi_{M-1}$ . When the number of points,  $M$ , is large we can see that the variance decreases. We do not use this argument but just the OLS assumption of a constant variance. This allowed us to write the bound for the diffusion deviation as:

$$\begin{aligned} \|\hat{a}_j^2 - a_j^2\|_2 &= \frac{1}{\Delta t} \left\| \frac{\sum_{i=0}^M ((X_{i+1} - X_i)^2 - a_j^2 \Delta t) \mathbb{1}_{\{X_i \in B_{c_j}\}}}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}} \right\|_2 = \frac{1}{\Delta t} \sqrt{V(\hat{\beta}_2)}, \\ &= \frac{1}{\Delta t} \sqrt{\frac{\sigma_{u_2}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}} = \frac{1}{\Delta t} \sqrt{\frac{\sigma_{u_2}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}, \\ &= \frac{1}{\Delta t} \sqrt{\frac{\sigma_{u_2}^2}{\sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}, \\ &= \frac{\sqrt{\Delta x}}{\Delta t} \sqrt{\frac{\sigma_{u_2}^2}{\Delta x \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}}}. \end{aligned} \quad (3.91)$$

This expression, like for the drift, depends on the ratio of  $\Delta x$  and  $\Delta t$  as well as the histogram  $\Delta x \sum_{i=0}^M \mathbb{1}_{\{X_i \in B_{c_j}\}}$ . A small denominator in the third term would increase the error, which implies the error is inversely proportional to the number of points in the bin of size  $\Delta x$ . Compared the drift estimator, there is a square root only on the spatial step but not the time step. This means that for a small  $\Delta x$ , the drift error bound would have a smaller error than the diffusion error.

## 3.4 Conclusion Chapter 3

1. An error analysis of the non-parametric estimation of the drift and diffusion of a SDE has been done. The goal was to study the impact of the sample size, time, and space discretization on the estimation errors.
2. In practical applications, a large amount of data on a given stochastic process are sampled at a fixed time interval,  $\Delta t$ . On the other hand, a researcher has control of any state space discretization. Let  $\Delta x$  be the spatial step size of a given state-space partition. Here spatial discretization implies a binning procedure i.e. partitioning the state space into bins of equal size  $\Delta x$ . In this context, understanding how  $\Delta t$  and  $\Delta x$  affects the estimation errors on the drift and diffusion is crucial to obtain reliable reconstruction of the components of a SDE.
3. We focus on a non-parametric estimation based on conditional expectations. We also limited the scope of our analysis to an Ornstein Uhlenbeck (OU) process.
4. We define, using time and space discretization, two conditional expectation estimators,  $\hat{m}_1$  and  $\hat{m}_2$ , and conduct an error analysis on them.
5. Assuming that the number of points in each bin is fixed, and equal to  $M$ , when the observational time step,  $\Delta t$ , is fixed we found that the  $L^2$  errors made when estimating the drift decays at a rate of  $\frac{1}{\sqrt{M}}$ .
6. When we relaxed the assumption that the number of points in each bin is fixed,

we found for a fixed  $\Delta t$  that the  $L^2$  errors when estimating the drift decays at a rate of  $\frac{1}{\sqrt{M}}$  while the diffusion error did not decay to zero but was bounded.

7. We look at conditional expectation estimator of the drift and diffusion over a certain bin,  $j$ , denoted  $\hat{b}_j$  and  $\hat{a}_j$ , respectively. We found that for, a fixed  $\Delta t$ , the  $L^2$  errors depend on the ratio  $\frac{\Delta x}{\Delta t}$  and the number of points inside the bin. Bins that are frequently visited have a smaller estimation errors because the square root term in the upper bound of the error is finite. Also, we found that choosing a nonzero spatial step size smaller or equal to the subsampling time produces small values of the ratio  $\frac{\Delta x}{\Delta t}$  and by consequence implies smaller estimation errors for both the drift and diffusion.
8. More research needs to be done for the case where the number of points is not fixed.

# APPENDIX A

---

## APPENDIX-CHAPTER 2

---

### **Brief overview of the spectral reconstruction method using Finite Element (FE)**

The finite difference approach exploits a close relationship between the spectrum of the infinitesimal generator and that of the data estimated transition probabilities matrix to estimate the best possible drift and diffusion terms. Issues such as model misspecification can be avoided using this approach. Unfortunately, one major limitation is the magnitude of the approximation errors that could be generated by the method itself. Indeed the method requires computing approximation for the derivatives of the eigenfunctions to perform the numerical optimization. Since no

conditions are imposed on the spectrum except of being discrete, finite differences of these functions appear as a major source of errors.

Crommelin and Vanden-Eijnden introduced in [20] a modification of the optimization error in order to avoid differentiating. The key idea is to use a Galerkin approximation of the eigentriplets that allow to compute the infinitesimal generator without using derivatives of the eigenfunctions, but instead differentiating test functions. This approach is based on the same key relationship between the infinitesimal generator and the Markov chain generated by the data and the finite difference could be seen as a special case of this more general approach. We summarize the key idea of this approach using the same notation as in [20].

Consider a diffusion process  $X_t$  on the space  $\Omega$  with invariant measure  $\mu$  that admits an invariant density  $\rho$ . Consider that the infinitesimal generator  $\mathcal{L}$  lies on a the Sobolev space  $H^2(\Omega, \mu)$  and the operator  $P_t$  on a larger space  $\mathcal{F} = L^2(\Omega, \mu)$ . The eigentriplets are denoted as  $\{\psi_k(x), \phi_k(x), \lambda_k(x)\}$ .

Using the Galerkin method implies that the domain of the conditional expectation operator  $P_t$  is approximated by its projection into a finite dimensional subspace denoted  $\mathcal{F}_{\mathcal{M}} \subset \mathcal{F}$ . The set of independent functions  $f_i : \Omega \rightarrow \mathbb{R}$ , for  $i = 1..M$  is a basis of  $\mathcal{F}_{\mathcal{M}}$ . A weak formulation of the eigenvalue problem (2.24) is expressed as:

$$\langle P_t \phi_k, f_i \rangle_\rho = \langle \Omega_k \phi_k, f_i \rangle_\rho \tag{1}$$

for all  $i = 1..M$ ,  $k \in \mathbb{N}$  and  $\rho$  (the invariant density of the process) is used as the weight function in the above inner product.

The Galerkin approximation of the right eigenfunctions, denoted  $\phi_k^g$ , on the basis

of function of  $\mathcal{F}_M$  is:

$$\phi_k^g = \sum_{i=1}^M v_{ki} f_i(x) \quad (2)$$

where the  $v_{ki} \in \mathbb{C}$  are expansion coefficients.

A matrix formulation of this problem can be obtained by first representing all the expansion coefficient by a matrix  $V$ , the basis function and the mapping operator by respectively a symmetric matrix  $R$  and another matrix  $T$  whose entries are as follows:

$$V_{ij} = \{v_{ki}\}_{i=1..M, k \in \mathbb{N}}, \quad (3)$$

$$R_{ij} = \langle f_i, f_j \rangle_\rho, \quad (4)$$

$$T_{ij} = \langle P_t f_i, f_j \rangle_\rho. \quad (5)$$

The following generalized eigenvalue problem for  $P_t$  is expressed as

$$VT = D_\Lambda VR, \quad (6)$$

where  $D_\Lambda$  is a diagonal matrix containing the eigenvalues of  $P_t$ . Similar problem than in [18] could be formulated for the adjoint of the infinitesimal generator with the same eigenvalues but with left eigenfunctions  $\psi_k$ . In this approach, instead of using the left eigenfunctions, functions  $\xi_k$  are derived from the generalized eigenvalue problem could be defined using the invariant density as

$$\psi_k = \rho \xi_k. \quad (7)$$

The eigenvalue problem gives Galerkin approximation of these functions:

$$TW^* = RW^* D_\Lambda, \quad (8)$$

where  $W$  is the matrix of the expansion coefficients  $w_{ki}$  of the functions  $\xi_k^g$  such that

$$\xi_k^g = \sum_{i=1}^M w_{ki} f_i(x), \quad (9)$$

$$VTW^* = D_\Lambda. \quad (10)$$

In the first method, [18, 19] the residuals of the discretized eigenfunctions of both the infinitesimal generator and its adjoint were minimized. In the latest approach [20] only the infinitesimal generator is integrated against a set of test functions to construct a weak formulation of the eigenvalue problem and the residuals generated are minimized using different weights to estimate the drift and diffusion. The general objective function appears as follows:

$$\tilde{E}(b, a) = \sum_{k=1}^K \sum_{i=1}^N \alpha_{ki} \left| \langle \xi_k, \mathcal{L}(b, a)\sigma_i \rangle_\rho - \lambda_k \langle \xi_k, \sigma_i \rangle_\rho \right|^2. \quad (11)$$

The choice of the Galerkin basis functions, the test functions and the weights enables to construct different methods.

The authors distinguished three cases:

1. **Case 1: Binning case.** Like in the finite difference approach, the eigenpairs are obtained after binning the state space to construct a transition probability and its eigendecomposition. Test functions are chosen to be integrated along the generator avoiding any numerical differentiation of the eigenfunctions. The objective function in this case is as follows:

$$\tilde{E}^b(b, a) = \sum_{k=1}^K \sum_{i=1}^N \alpha_{ki} E \left| \xi_k(X_t)(\mathcal{L}\sigma_i)(X_t) - \lambda_k \xi_k(X_t)\sigma_i(X_t) \right|^2. \quad (12)$$

2. **Case 2: Smooth Galerkin.** When the eigenfunctions are estimated with smooth Galerkin functions, the test functions are set to be the actual eigenfunctions with weights depending on the data eigenvalues. The objective function can be more easily written in terms of matrix:

$$\tilde{E}^g(b, a) = ||VQW^* - D_\lambda||_c^2. \quad (13)$$

where the weighted Frobenius norm is used with a weight denoted  $c$ .

3. **Case 3: Mixed case.** A third technique can be designed using smooth Galerkin basis functions that would also be used as test functions instead of the eigenfunctions. The objective function is

$$\tilde{E}^m(b, a) = \sum_{k=1}^K \sum_{i=1}^N \alpha_{ki} |(WQ^T - D_\lambda WW)_{ki}|^2. \quad (14)$$

---

## Bibliography

---

- [1] Y. Ait-Sahalia, *Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach*, *Econometrica* **70** (2002), 223-262.
- [2] Y. Ait-Sahalia, P.A. Mykland, and L. Zhang, *How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise*, *Review of Financial Studies* **18** (2005), 351-416.
- [3] Y. Ait-Sahalia and L. P. Hansen, *Handbook of Financial Economics, Vol.1: Tools and Techniques and Vol.2: Applications*, Elsevier (2010).
- [4] R. Azencott, A. Beri, and I. Timofeyev, *Adaptive Sub-sampling for Parametric Estimation of Gaussian Diffusions*, *Journal of Statistical Physics* **6** (2010), no. 139, 1066-1089.
- [5] R. Azencott, A. Beri, and I. Timofeyev, *Parametric Estimation of Stationary Stochastic Processes under Indirect Observability*, *Journal of Statistical Physics* **1** (2011), no. 144, 150-170.
- [6] R. Azencott, A. Beri, A. Jain, and I. Timofeyev, *Sub-sampling and Parametric Estimation for Multiscale Dynamics*, *Communications in Mathematical Sciences* **4** (2013), no. 11, 939-970.

## BIBLIOGRAPHY

---

- [7] R. Azencott, P. Ren, and I. Timofeyev, *Parametric Estimation from Approximate Data: Non-Gaussian Diffusions*, Journal of Statistical Physics **5** (2015), no. 161, 1276-1298.
- [8] T. Bjork, *Arbitrage Theory in Continuous Time*, 3rd edition, Oxford Finance Series, Oxford University Press, Oxford, 2009.
- [9] F. Black and M. Scholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy **81** (1973), no. 3, 637-654.
- [10] Z. Cai and Y. Hong, *Some recent developments in nonparametric finance*, in Qi Li, Jeffrey S. Racine (ed.) Nonparametric Econometric Methods Advances in Econometrics **25** (2009), 379-432.
- [11] J. Campbell, A.W. Lo, and A.C. MacKinlay, *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press, 1997.
- [12] J. Cha, B.R.Cho, and J.L. Sharp, *Rethinking the truncated normal distribution*, Experimental Design and Process Optimisation **3** (2014), 327-363.
- [13] J. Cha and B.R.Cho, *Classical statistical inference extended to truncated populations for continuous process improvement: Test statistics, p-values, and confidence intervals*, Quality and Reliability Engineering International **8** (2015), no. 31, 1807-1824.
- [14] R.R. Chen and C.F. Lee, *A Constant Elasticity of Variance (CEV) Family of Stock Price Distributions in Option Pricing: Review and Integration*, Journal of Financial Studies **1** (1983), 29-51.
- [15] D. Chikobvu and K. Chinhamu, *Random walk or mean reversion? Empirical evidence from the crude oil market*, Statistik Journal of the Turkish Statistical Association **6** (2013), no. 3, 1-9.
- [16] J. Cox, *Notes on Option Pricing I: Constant Elasticity of Diffusions*, The Journal of Portfolio Management **22** (1996), 15-17.
- [17] J. C. Cox, J.E. Ingersoll, and S.A. Ross, *A Theory of the Term Structure of Interest Rates*, Econometrica **53** (1985), 385-407.

## BIBLIOGRAPHY

---

- [18] D.T. Crommelin and E. Vanden-Eijnden, *Reconstruction of diffusion using spectral data from time series*, Communications in Mathematical Sciences **4** (2006), no. 3, 651-668.
- [19] D.T. Crommelin and E. Vanden-Eijnden, *Data-based inference of generators for Markov jump processes using convex optimization*, Multiscale Modeling and Simulation **7** (2009), 1751-1778.
- [20] D.T. Crommelin and E. Vanden-Eijnden, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Modeling and Simulation **9** (2011), 1588-1623.
- [21] E. Dias Frances and F.J. Rubio, *On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables*, Statistical Papers **54** (2013), 309-323.
- [22] N. Diawara, SHS. Indika, and EM. Maboudou-Tchao, *The Distribution of the Sum of Independent Product of Bernoulli and Exponential.*, American Journal of Mathematical and Management Sciences **32** (2013), no. 1, 75-89.
- [23] D. C Emanuel and J. D. MacBeth, *Further Results of the Constant Elasticity of Variance Call Option Pricing Model*, Journal of Financial and Quantitative Finance **4** (1982), 533-553.
- [24] L. C. Evans, *An Introduction to Stochastic Differential Equations*, American Mathematical Society (2014).
- [25] E. Fama, *The Behavior of Stock Market Prices*, Journal of Business **38** (1965), 34-105.
- [26] Federal reserve system Board of Governors letters, *Supervisory guidance on Model risk Management*, Division of Banking Supervision and Regulation of the Federal reserve system Board of Governors, April 4 2011, <https://www.federalreserve.gov/bankinfo/reg/srletters/sr1107.htm> (2011).
- [27] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, 2nd edition*, Springer, 1985.
- [28] H. Geman, *Energy Commodity Prices: Is Mean-Reversion Dead?*, Journal of Alternative Investments (2005), 31-45.
- [29] H. Geman, *Mean reversion versus random walk in oil and natural gas prices*, Advances in Mathematical Finance, Birkhuser, Boston (2007), 219-228.

## BIBLIOGRAPHY

---

- [30] H. Geman and YF. Shih, *Modeling Commodity Prices under the CEV Model*, The Journal of Alternative Investments **11** (2009), no. 2, 65-84.
- [31] RC. Geary, *The frequency distribution of the quotient of two normal variates*, Journal of the Royal Statistical Society B Methodology **93** (1930), 442-446.
- [32] P. Glasserman, *Monte-Carlo Methods in Financial Engineering*, Springer, New York, 2003.
- [33] E. Gobet, M. Hoffmann, and M. Reiß, *Nonparametric estimation of scalar diffusions based on low-frequency data*, The Annals of Statistics **32** (2004), 2223-2253.
- [34] C. Gouriéroux and P. Valéry, *Estimation of the Jacobi Process*, University of Montreal, D.P. (2003).
- [35] C. Gouriéroux, E. Renault, and P. Valéry, *Diffusion Processes with Polynomial Eigenfunctions*, Annales d'Économie et de Statistique **85** (2007), 115-130.
- [36] C. Gouriéroux and P. Gagliardini, *Stochastic Migration Models with Application to Corporate Risk*, Journal of Financial Econometrics **3** (2005), no. 2, 188-226.
- [37] C. Gouriéroux and J. Jasiak, *Multivariate Jacobi process and with application to smooth transitions*, Journal of Econometrics **131** (2006), 475-505.
- [38] C. Gouriéroux and J. Jasiak, *Financial Econometrics*, Princeton, NJ: Princeton University Press, 2001.
- [39] M. Grasselli, *The 4/2 stochastic volatility model: A Unified approach for the Heston and the 3/2 model*, Mathematical Finance, in press (2016), 1-22.
- [40] W. Greene, *Econometric Analysis, 5th edition*, Prentice Hall, 2003.
- [41] R. B. Israel, J. S. Rosenthal, and J. Z. Wei, *Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings*, Mathematical Finance **11** (2001), 245-265.
- [42] L. Hansen, J. Scheinkman, and N. Touzi, *Spectral methods for identifying scalar diffusions*, Journal of Econometrics **86** (1998), 1-32.

## BIBLIOGRAPHY

---

- [43] J. Hayya, D. Armstrong, and N. Gressis, *A note on the ratio of two normally distributed variables*, Management Science. **21** (1975), no. 11, 1338-1341.
- [44] S. Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, Review of Financial Studies **6** (1993), 327-343.
- [45] S. Heston, *A Simple New Formula for Options with Stochastic Volatility*, Manuscript: John M. Olin, School of Business, Washington University (1997).
- [46] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, Education **43** (2001), 525-546.
- [47] D. Hinkler, *On the ratio of two correlated normal random variables*, Biometrika **56** (1969), 635-639.
- [48] J. Hull, *Options futures and other derivatives*, Prentice Hall, 2008.
- [49] J.W. Jawitz, *Moments of truncated continuous univariate distributions*, Advances in Water Resources **27** (2004), 269-287.
- [50] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, Wiley-Interscience, 1994.
- [51] L. Kelly, E. Platen, and M. Sorensen, *Estimation for discretely observed diffusions using transform functions*, Journal of Applied Probability **41A** (2004), 99-118.
- [52] M. Kessler and M. Sorensen, *Estimating equations based on eigenfunctions for a discretely observed diffusion process*, Bernoulli **5** (1999), 299-314.
- [53] S. Kumar, *Mixed Type Distribution*, Lecture notes, Department of Statistics and Actuarial Science, University of Iowa (2004), 1-14.
- [54] D. Lando and T. Skodeberg, *Analyzing Ratings Transitions and Rating Drift with Continuous Observations*, Journal of Banking and Finance **26** (2002), 423-444.
- [55] B.G.Malkiel, *A random walk down Wall Street*, New York: Norton, 1973.

## BIBLIOGRAPHY

---

- [56] G. Marsaglia, *Evaluating the Normal Distribution*, Journal of Statistical Software **11** (2004), no. 4, 1-11.
- [57] G. Marsaglia, *Ratios of Normal Variables*, Journal of Statistical Software **16** (2006), 1-10.
- [58] N. Meddahi, *An Eigenfunction Approach for Volatility Modeling*, Manuscript, Department of Economics, University of Montreal and CIRANO (2001), 1-49.
- [59] R. C. Merton, *Theory of Rational Option Pricing*, Bell Journal of Economics and Management Science, The RAND Corporation **1** (1973), no. 4, 141-183.
- [60] P. Metzner, E. Dittmer, T. Jahnke, and Ch. Schutte, *Generator estimation of Markov jump processes*, Journal of Computational Physics **227** (2007), 353-375.
- [61] D.W.C. Miao, *Analysis of the Discrete Ornstein-Uhlenbeck Process Caused by the Tick Size Effect*, Journal of Applied Probability **50** (2013), no. 4, 1102-1116.
- [62] S. Nadarajah, *Making the Cauchy work*, Brazilian Journal of Probability and Statistics **25** (2011), no. 1, 99-120.
- [63] S. Nadarajah and S. Kotz, *A truncated Cauchy distribution*, International Journal of Mathematical Education in Science and Technology **37** (2006), 605-608.
- [64] B. Øksendal, *Stochastic Differential Equations*, Springer, 1980.
- [65] G. A. Pavliotis, *Stochastic processes and applications. Diffusion processes, the Fokker Planck and Langevin equations*, Springer, 2014.
- [66] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics, Texts in Applied Mathematics 37, 2nd Edition*, Springer, 2010.
- [67] J. Rosenthal, *A First look at Rigorous Probability Theory, 2nd edition*, World Scientific, 2013.
- [68] S. Ross, *A first course in probability, 8th edition*, Pearson, 2010.
- [69] F. Salmon, *Recipe for Disaster: The Formula That Killed Wall Street*, Wired Magazine (2009).
- [70] E.S. Schwartz, *The stochastic behavior of commodity prices: implications for valuation and hedging*, Journal of Finance **52** (1997), no. 3, 923-973.

## BIBLIOGRAPHY

---

- [71] B. Skorodumov, *Estimation of mean reversion in oil and gas markets*, Technical report, MITSUI/2008-10-14, MITSUI & CO Energy risk management LTD. (2008), 1-4.
- [72] H. Sorensen, *Parametric inference for diffusion processes observed at discrete points in time: a survey*, International Statistical Review **72** (2004), 337-354.
- [73] I. Stewart, *The mathematical equation that caused the banks to crash*, The Observer (2012).
- [74] P Sura and J Barsugli, *A note on estimating drift and diffusion parameters from time series*, Physics Letters A **5** (2002), no. 305, 304-311.
- [75] P Sura, *Stochastic Analysis of Southern and Pacific Ocean Sea Surface Winds*, Journal of Atmospheric Sciences **60** (2003), 654-666.
- [76] W. F. Thompson, A. H. Monahan, and D.T. Crommelin, *Parametric Estimation of the Stochastic Dynamics of Sea Surface Winds*, Journal of Atmospheric Sciences **71** (2014), 3465-3483.
- [77] O. Vasicek, *An equilibrium characterization of the term structure*, Journal of Financial Economics **5** (1977), 177-188.
- [78] L. Zhang, P.A. Mykland, and Y. Ait-Sahalia, *A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data*, Journal of the American Statistical Association **100** (2005), no. 472, 1394-1411.