



DOCTORAL THESIS

Bayesian Modeling and Estimation for Complex Multiparameter Problems with Real Applications

Author:

Constandina KOKI

Supervisor:

Ass. Professor

Ioannis VRONTOS

A THESIS

Submitted to the Department of Statistics of the Athens University of Economics and Business in partial fulfilment of the requirements for the degree of PhD in Statistics

February 11, 2021

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μπεϋζιανή Μοντελοποίηση και Εκτίμηση Πολύπλοκων Πολυμεταβλητών Προβλημάτων σε Πραγματικές Εφαρμογές

Συγγραφέας:

Κωνσταντίνα ΚΟΚΗ

Επιβλέπων Καθηγητής:

Αναπ. Καθηγητής

Ιωάννης ΒΡΟΝΤΟΣ

ΔΙΑΤΡΙΒΗ


*Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Διδακτορικού Διπλώματος στη Στατιστική*

11 Φεβρουαρίου 2021

Declaration of Authorship

I, Constandina KOKI, declare that this thesis titled, “Bayesian Modeling and Estimation for Complex Multiparameter Problems with Real Applications” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: 15/02/2021

Acknowledgements

I would like to express my sincere gratitude to my advisors, associate Professors Ioannis Vrontos and Loukia Meligkotsidou for the continuous support during my Ph.D studies. Their knowledge, motivation and patience have given me more power and spirit to excel in the research writing. They have always been there for me, even at very hard times, not only as academics, but as my friends. *I will always remember fondly the occasions when you opened your home to me.*

Apart from my supervisors, I am also profoundly grateful to associate Professor Costis Melolidakis. He has been my teacher since my undergraduate studies. Over the years he became my co-author and my mentor. *Costi, I could not thank you enough for your unreserved kindness and compassionate support.*

The work presented in this thesis has been critically assessed and approved by an outstanding committee to whom I am thankful: Professor Dimitrios Karlis, Professor Dellaportas Petros, Professor Dimitrios Fouskakis, Professor Theodore Kypraios and associate Professor Kostas Kalogeropoulos. I am thankful, for the committee's insightful and helpful comments.

Also, I am grateful to my parents. They have unconditionally provided everything that I needed. *Mom and dad, you where there, when I needed you the most with your selfless love and boundless support. I can not thank you enough.* I am also grateful to my siblings. They gave me enough moral and wholehearted support, encouragement and motivation to accomplish my personal goals. *Your endless love has been there all along.*

In addition, I would like to thank Sofia, who has been like a mother to me. She took care of my children in the best way that I could imagine and therefore I could focus on my studies. *It would be impossible without your help and support. I will forever be grateful to you.*

Last but foremost, I am immensely grateful to my husband, Stefanos. He has been my anchor. He was there during my best and worst periods of my life. He was always there with percipient and helpful comments on my work. He was there before all of the submissions sharing my anxiety and after the submissions, sharing my frustration or excitement. He is my person, the father of my children. *Stefane, I will always remember our endless night discussions! Your love has given me the strength and courage to came up against all obstacles.*

“To my husband Stefanos and

To my wonderful and adorable children Ioanna-Sofia and Natalia-Evelina”

Abstract

School of Information Sciences and Technology

Department of Statistics

Constandina KOKI

In the big data era, the study of complex multiparameter problems is more than necessary. The development of Machine Learning techniques enhanced the inferential ability of statistical models. In this direction, by leveraging Machine Learning techniques, we propose a new predictive Hidden Markov model with exogenous variables, within a Bayesian framework, for joint inference and variable selection. We propose a computational Markov Chain Monte Carlo algorithm that offers improved forecasting and variable selection performance, compared to existing benchmark models. Our methodology is applied in various simulated and real datasets, such as realized volatility data and cryptocurrency return series. Furthermore, we exploit the Bayesian methodology in implementing the X-ray luminosity function of the Active Galactic Nuclei under the assumption of Poisson errors in the determination of X-ray fluxes and estimation uncertainties.

Publications and Conferences

- The results of Chapter 2 are presented in the article entitled "Forecasting under model uncertainty: Non-Homogeneous Hidden Markov models with Pólya-Gamma data augmentation", co-authored by Constandina Koki, Loukia Meligkotsidou and Ioannis Vrontos and has been published in *Journal of Forecasting*. Additionally, preliminary versions of this article have been presented in various international conferences, i.e., "Computational and Methodological Statistics" (CMStatistics2017), "CRoNoS & Multivariate Data Analysis (CRoNos&MDA2017) Workshop", "13-th RCA Bayesian Econometrics Workshop", the "18-th Conference on Research on Economic Theory and Econometrics" (CRETE2019) and "Greek Stochastics 2019", among others.
- The results reported in Chapters 3 and 4 are presented in the article entitled "A Peek into the Unobservable: Hidden States and Bayesian Inference for the Bitcoin and Ether Price Series" and "Exploring the Predictability of Cryptocurrencies via Bayesian Hidden Markov Models", co-authored by Constandina Koki, Stefanos Leonardos and Georgios Piliouras. They have been presented in "Cryptocurrency Research Conference" (CRC2020). A preliminary version of the previous articles has been published in *Future Internet* journal, under the title "Do cryptocurrencies camouflage latent economic effects" and has won the Best Paper award at the 3rd International Decentralized Conference (DC2019) by Future Internet Journal.
- The work of Chapter 5 is reported in the article under the title "The X-ray luminosity function of active galactic nuclei in the redshift interval $z = 3 - 5$. It is co-authored by Constandina Koki, Antonis Georgakakis, James Aird, Johannes Buchner, Mara Salvato, Marie-Luise Menzel, William Brandt, Ian McGreer, Tom Dwelly, George Mountrichas, Ioannis Georgantopoulos, Li-Ta Hsu, Andrea Merloni, Zhu Liu, Kirpal Nandra and Nicholas Ross and it is published in the *Monthly Notices of the Royal Astronomical Society*. It has been benefited from the "THALES" project 383549 that is jointly funded by the European Union and the Greek Government in the framework of the program "Education and Lifelong Learning".

Περίληψη

Σχολή Επιστημών & Τεχνολογίας της Πληροφορίας

Τμήμα Στατιστικής

Κωνσταντίνα ΚΟΚΗ

Στην εποχή της πληροφορίας και του μεγάλου όγκου δεδομένων, η ανάπτυξη τεχνικών εκμάθησης (Machine Learning) με τη βοήθεια των ηλεκτρονικών υπολογιστών, μας έδωσε τη δυνατότητα να μελετήσουμε πολύπλοκα προβλήματα. Η μη εύρεση αναλυτικών λύσεων για εκτίμηση των παραμέτρων διάφορων υποδειγμάτων έχει οδηγήσει στην άνθιση της Μπεϋζιανής συμπερασματολογίας. Σε αυτή την κατεύθυνση στρέφεται και η παρούσα Διδακτορική διατριβή. Με εργαλεία τις τεχνικές εκμάθησης, ασχολούμαστε με τη μοντελοποίηση πολύπλοκων πολυμεταβλητών Μπεϋζιανών μοντέλων.

Η διατριβή αυτή χωρίζεται σε δύο βασικά ερευνητικά πεδία. Το πρώτο πεδίο αφορά την ανάπτυξη Κρυμμένων Μαρκοβιανών μοντέλων (Hidden Markov models) με πεπερασμένα στάδια (states/regimes) και εξωγενείς επεξηγηματικές μεταβλητές. Ιδιαίτερα, επεκτείνουμε προηγούμενα Κρυμμένα Μαρκοβιανά μοντέλα, προτείνοντας συγκεκριμένη μεθοδολογία, βασιζόμενη στην Pólya-Gamma τεχνική αύξησης δεδομένων (data augmentation) για την εκτίμηση των παραμέτρων, ενώ ταυτόχρονα επιλέγουμε στοχαστικά τις επεξηγηματικές μεταβλητές που επηρεάζουν την ανεξάρτητη μεταβλητή. Αξιοποιώντας μεθόδους βασισμένες σε Μαρκοβιανές Αλυσίδες Μοντε Καρλο (Markov Chain Monte Carlo), προτείνουμε ένα σχήμα προσομοίωσης για πιο ακριβή συμπερασματολογία και εν τέλει βελτιωμένες προβλέψεις, σε σχέση με προηγούμενες εργασίες. Αποδεικνύουμε εμπειρικά, ότι η μεθοδολογία αυτή υπερέχει — τόσο στις προβλέψεις όσο και στην επιλογή των επεξηγηματικών μεταβλητών —, έναντι άλλων προτεινόμενων μοντέλων, σε προσομοιωμένα δεδομένα. Στη συνέχεια εφαρμόζουμε το προαναφερθέν μοντέλο σε realized volatility δεδομένα.

Επιπλέον, μελετάμε χρονολογικές σειρές κρυπτονομισμάτων χρησιμοποιώντας τα Κρυμμένα Μαρκοβιανά μοντέλα, με απώτερο στόχο να κατανοήσουμε τα χαρακτηριστικά αυτών των καινούργιων επενδυτικών αγαθών (financial assets) και τη σχέση τους με τα υπάρχοντα συμβατικά οικονομικά αγαθά (traditional financial covariates). Τέλος, εφαρμόζοντας διάφορα Κρυμμένα Μαρκοβιανά μοντέλα σε

σειρές αποδόσεων κρυπτονομισμάτων, δείχνουμε ότι το προτεινόμενο μοντέλο με τέσσερα στάδια έχει μεγαλύτερη προβλεπτική ισχύ από όλα υπόλοιπα μοντέλα της μελέτης μας.

Το δεύτερο ερευνητικό πεδίο αφορά μία εφαρμογή των Μπεϋζιανών διαδικασιών Poisson για τη μοντελοποίηση της X-ray πιθανοφάνειας των Ενεργών Γαλαξιακών Νουκλεοτιδίων (Active Galactic Nuclei). Σε συνεργασία με μία ερευνητική ομάδα αποτελούμενη από φυσικούς και αστροφυσικούς, μέσω της Μπεϋζιανής προσέγγισης, λαμβάνουμε υπόψη τα Poisson σφάλματα που προκύπτουν από τις διακυμάνσεις των X-rays καθώς επίσης και την αβεβαιότητα στην εκτίμηση των φωτομετρικών ερυθρομετατοπίσεων. Έτσι, συμβάλουμε στην καλύτερη κατανόηση της δημιουργίας των υπερμαζικών μελανών οπών (super-massive black holes) στο σύμπαν.

Δημοσιεύσεις και παρουσιάσεις σε διεθνή συνέδρια

- Τα αποτελέσματα του Κεφαλαίου 2, έχουν δημοσιευτεί υπό τον τίτλο "Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Pólya-Gamma data augmentation" στο περιοδικό *Journal of Forecasting* με συν-συγγραφείς τους Ιωάννη Βρόντο και Λουκία Μελγκοτσίδου. Το άρθρο αυτό έχει παρουσιαστεί σε διεθνή συνέδρια, όπως "Computational and Methodological Statistics" (CMStatistics2017), "CRoNoS & Multivariate Data Analysis" (CRoNos&MDA2017) Workshop", "13-th RCA Bayesian Econometrics Workshop", "18-th Conference on Research on Economic Theory and Econometrics" (CRETE2019) και "Greek Stochastics 2019", μεταξύ άλλων.
- Τα αποτελέσματα που προέκυψαν κατά τη μελέτη των κρυπτονομισμάτων σε συνεργασία με τους Στέφανο Λεονάρδο και Γεώργιο Πήλιουρα, βλ. Κεφ.3, 4, συνοψίζονται στα άρθρα "A Peek into the Unobservable: Hidden States and Bayesian Inference for the Bitcoin and Ether Price Series" και "Exploring the Predictability of Cryptocurrencies via Bayesian Hidden Markov Models". Το τελευταίο άρθρο έχει παρουσιαστεί στο συνέδριο "Cryptocurrency Research Conference 2020". Προγενέστερη εργασία στα άρθρα αυτά έχει δημοσιευτεί στο περιοδικό *Future Internet* υπό τον τίτλο "Do cryptocurrency prices camouflage latent economic effects: A Bayesian Hidden Markov approach" και έχει βραβευτεί με το "Best paper award", από το προαναφερθέν περιοδικό κατά το συνέδριο "Decentralized 2019".
- Τα αποτελέσματα που βρίσκονται του Κεφαλαίου 5, έχουν δημοσιευτεί στο περιοδικό *Monthly notices of the Royal astronomical society* υπό τον τίτλο "The X-ray luminosity function of Active Galactic Nuclei in the redshift interval $z = 3 - 5$ " με συγγραφείς τους Constandina Koki, Antonis Georgakakis, James Aird, Johannes Buchner, Mara Salvato, Marie-Luise Menzel, William Brandt, Ian McGreer, Tom Dwelly, George Mountrichas, Ioannis Georgantopoulos, Li-Ta Hsu, Andrea Merloni, Zhu Liu, Kirpal Nandra και Nicholas Ross. Η εργασία αυτή καθώς και μέρος του διδακτορικού μου, είχε χρηματοδοτηθεί από το ερευνητικό σχέδιο 383549, από την Ευρωπαϊκή Ένωση και την Ελληνική Δημοκρατία στο πλαίσιο του προγράμματος "ΘΑΛΗΣ: Ενίσχυση της Διεπιστημονικής ή και Διδρυματικής έρευνας και καινοτομίας με δυνατότητα προσέλκυσης ερευνητών υψηλού επιπέδου από το

εξωτερικό μέσω της διενέργειας βασικής και εφαρμοσμένης έρευνας αρι-
στείας.

Contents

Declaration of Authorship	iii
Acknowledgements	v
Abstract	vii
Publications and Conferences	viii
Περίληψη	ix
1 Introduction	1
2 Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Pólya-Gamma data augmentation	5
2.1 Introduction	5
2.2 The Non-Homogeneous Pólya-Gamma Hidden Markov Model	9
2.2.1 Label Switching	10
2.2.2 Tackling the label switching problem	11
2.3 Bayesian Inference and Computational Strategy	12
2.3.1 The MCMC Sampling Scheme	13
2.3.2 Inference for fixed sets of predictors	13
2.3.2.1 Simulation of the logistic regression coefficients	14
2.3.3 Inference under model uncertainty	16
2.3.4 The m -state NHPG	17
2.4 Bayesian Forecasting and Scoring rules	19
2.4.1 Forecasting criteria	20
2.5 Simulation Study	21
2.5.1 Case study: The NHPG with model uncertainty	22
2.6 Empirical Application: Realized volatility data	25
2.6.1 The data	26
2.6.2 Results	27

2.6.2.1	Model Selection	29
2.6.2.2	Forecasting	30
2.7	Conclusions and Discussion	31
2.8	Appendix	32
2.8.1	The Pólya-Gamma distribution	32
2.8.2	Benchmark models	33
2.8.3	Case study for the 2-state fixed model	34
2.8.4	The Continuous Rank Probability Score	36
2.8.5	Metrics of Comparison	37
3	A Peek into the Unobservable: Hidden States and Bayesian Inference for the Bitcoin and Ether Price Series	39
3.1	Introduction	39
3.1.1	Motivation, Methodology and Main Results	39
3.1.2	Related Literature	42
3.1.3	Outline	44
3.2	Methodology & Data	44
3.2.1	The Scaled Forward-Backward algorithm	45
3.2.2	Simulation Scheme	47
3.2.3	Data	47
3.3	Results & Analysis	49
3.3.1	Hidden States: Bitcoin 2014–2019	49
3.3.2	Hidden States: Bitcoin and Ether 2017–2019	51
3.3.3	Hidden States: Financial Properties (BTC 2014-2019)	52
3.3.4	Significant Explanatory Variables: Bitcoin and Ether	53
3.4	Discussion: Limitations and Future Work	56
3.5	Appendix:	57
3.5.1	Descriptive statistics	57
3.5.2	Statistical tests	58
4	Exploring the Predictability of Cryptocurrencies via Bayesian Hidden Markov Models	61
4.1	Introduction	61
4.1.1	Other Related Literature	64
4.2	Data and Methodology	65
4.2.1	The Data	65
4.2.2	The Econometric framework	66

4.2.3	Performance Evaluation	69
4.3	Empirical Analysis	70
4.3.1	Out-of-Sample analysis	70
4.3.2	In-sample analysis of the models with the best predicting performance	73
4.3.3	Hidden States classification and interpretation	76
4.4	Discussion	78
5	The X-ray luminosity function of Active Galactic Nuclei in the redshift interval $z = 3 - 5$	81
5.1	Introduction	81
5.2	Data	84
5.2.1	Chandra survey fields	84
5.2.2	XMM-XXL survey data	90
5.3	Methodology	92
5.3.1	Statistical background	92
5.3.2	X-ray Luminosity function estimation: Chandra survey fields	102
5.3.3	X-ray Luminosity function estimation: XMM-XXL field . .	104
5.3.4	X-ray luminosity function models	106
5.3.5	Potential obscuration effects	109
5.4	Results	111
5.4.1	The non-parametric X-ray luminosity function determination	111
5.4.2	Parametric X-ray luminosity function determination	113
5.4.3	Comparison with previous studies	114
5.4.4	Contribution of AGN to ionisation of the Universe	118
5.5	Discussion	123
5.6	Conclusions	128
	Bibliography	129

List of Figures

2.1	Plots of the empirical posterior predictive distributions based on a normal kernel function for three randomly selected out-of-sample forecasts, $L = 15, 75, 85$, using the NHPG (black continuous line), M&D (gray dashed line), the HHMM (gray dotted line) and the BASAD (gray squared line). Actual out-of-sample values are marked with asterisks. These plots visualize the advantage of NHPG: global maximums of the multimodal distributions is achieved close to the actual values.	25
2.2	Observed process (black dotted line) and hidden process: the true hidden states are marked with blue x and the simulated states are marked with black dots. The true hidden process is well estimated. .	25
2.3	Time series (blue line) of the monthly realized volatility of the Standard & Poor (S&P) 500 index (in logarithmic scale, left axis) for the period 1926-2007, using the NHPG ₁ . Gray-shaded bars show a realization of the times with hidden state 1. The NHPG ₁ exploits the heteroscedasticity of the series. Red dots are the posterior mean probabilities (right axis) of staying at the same state and indicate a persistent unobserved process.	28
2.4	Thinned (1:5 observations) realized volatility time series (blue line) versus the observed process as calculated by the proposed NHPG ₁ (gray solid line).	29
2.5	Conjointly plotted empirical continuous approximations (based on a normal kernel function) of the posterior predictive distribution for out-of-sample periods $L = 15, 85, 100$, using the NHPG (black continuous line), M&D (gray dashed line) and the HHMM (gray dotted line). Actual out-of-sample values are marked with asterisks.	36

3.1	Application of the NHPG model on the BTC price series. The algorithm essentially identifies two periods, the first from 2014 (start of the dataset) to 2017 and the second from 2017 to date. This motivates separate analysis of the BTC for the latter period and comparison with the ETH price series over the same period.	49
3.2	BTC logarithmic-return series (blue line – right axis) for the period 1/2014-8/2019 with the mean smoothed marginal probabilities of state 1, i.e., $Pr(Z_t = 1)$ (gray bars – left axis).	51
3.3	BTC (left panel) and ETH (right panel) logarithmic-returns series (blue lines – right axis) for the period 1/2017-8/2019 with the mean smoothed marginal probabilities of state 1, i.e., $Pr(Z_t = 1)$, (gray bars – left axis).	51
3.4	Comparison of the USD/EUR exchange rate (blue line), S&P500 (green line) and Crude Oil Future Prices (gray line) as a percentage of price changes from the initial period. The USD/EUR exchange rate is less volatile than the other predictors.	56
4.1	Daily price series plots for the three cryptocurrencies considered in this study: Bitcoin (upper plot), Ether (middle plot) and Ripple (lower plot). We study the Bitcoin time series for the period ranging from 1/2014 until 11/2019, the Ether series for the period ranging from 9/2015 until 11/2019 and the Ripple data series from 1/2015 until 11/2019.	66
4.2	Graphical representation of the Non-Homogeneous Hidden Markov model.	68
4.3	Percentage return series (gray lines) and quantiles of the posterior sample (replicated) empirical distributions for the Bitcoin series (first row), Ether series (second row) and Ripple series (third row). Yellow lines show the 0.5% and 99.5% quantiles of the estimated in-sample distributions and define the 1% credibility region, whereas red lines show the estimated posterior median. Plots on the left are based on the estimated distributions via the 4-state Non-Homogeneous Hidden Markov (HM) model while plots on the right show the estimated distributions as derived from the Random Walk (RW) benchmark model.	73

4.4	Bitcoin (upper plot), Ether (middle plot) and Ripple (lower plot) percentage return series conditionally on a realization of the hidden process. The hidden process is estimated using the 4-state NHHM on the aforementioned cryptocurrencies' percentage return series. Red, yellow, purple and green dots indicate the time periods that the hidden process is at states 1, 2,3 and 4 respectively.	77
4.5	Smoothed probabilities of being in state m , i.e., $P(Z_t = m \mid y_1, \dots, y_T)$, $m = 1, \dots, 4$ using the 4-state Non-Homogeneous Hidden Markov model. Columns 1 to 4 correspond to states 1 to 4 while rows 1 to 3 correspond to Bitcoin, Ether and Ripple, respectively.	78
4.6	Persistence probabilities for each regime, based on the mean posterior transition probabilities $p_{ii} = P(Z_t = i \mid Z_{t-1} = i)$, $i = 1, \dots, 4$. The rows 1 to 3 corresponds to the Bitcoin, Ether and XRP return series respectively and columns 1 to 4 correspond to the states 1 to 4 respectively.	79
5.1	X-ray Sensitivity curves for the combined Chandra surveys used in the analysis (left panel) and the XMM-XXL survey (right panel). . .	86
5.2	Spectroscopic vs photometric redshift measurements for the sample of X-ray selected sources used in this paper. The data points correspond to the median value of the PDZ. The errorbars correspond to the 90% confidence interval around the median.	88
5.3	Examples of photometric redshift Probability Distribution Functions for X-ray selected AGN in the AEGIS field. The different colours correspond to different sources. There is a variety of PDZ shapes in the sample, including uni-modal and relatively narrow (blue solid), unimodal but broad (red dashed) and multi-modal (black dot-dashed).	89

- 5.4 Monochromatic 2 keV X-ray luminosity, $L_\nu(2 \text{ keV})$, plotted as a function of monochromatic 2500 Å UV luminosity, $L_\nu(2500 \text{ Å})$. The data points (cross or dots) are *XMM-XXL* X-ray selected broad line QSOs with secure spectroscopic redshifts in the interval $z = 1 - 5$. Sources at $z > 3$ are highlighted with different symbols (crosses). The red dashed line is the bisector best-fit $L_\nu(2 \text{ keV}) - L_\nu(2500 \text{ Å})$ relation determined by Lusso, E. et al. (2010); $\log L_{2 \text{ keV}} = 0.760 \log L_\nu(2500 \text{ Å}) + 3.508$. For the *XMM-XXL* QSOs the X-ray luminosity density at 2 keV is estimated from the 0.5-2 keV flux assuming a power-law X-ray spectrum with index $\Gamma = 1.9$. The 2500 Å monochromatic luminosity is determined from the SDSS photometry. For the k-corrections we adopt the simulated QSO templates of McGreer et al. (2013). For a QSO with redshift z the SDSS photometric bands with effective wavelengths that bracket the wavelength $2500 \times (1 + z) \text{ Å}$ are identified. The mean model QSO SED at that redshift is then scaled to the observed SDSS optical magnitudes in those two bands. The monochromatic luminosity at 2500 Å is then estimated from the scaled model SED. At redshifts $z \gtrsim 2.7$ the rest-frame 2500 Å lies beyond the effective wavelength of the SDSS z -band (9134 Å). For these sources the flux density at $2500 \times (1 + z) \text{ Å}$ is an extrapolation using the model SED. The results do not change if we simply linearly interpolate between the observed flux densities of the SDSS bands that bracket the rest-frame 2500 Å. This model-independent approach however, does not allow extrapolation beyond $z \gtrsim 2.7$ 92
- 5.5 Histogram of the difference $\Delta \log L_\nu(2500 \text{ Å})$ between the monochromatic 2500 Å luminosity of broad-line QSOs in the *XMM-XXL* (blue points of Figure 5.4) and the mean $\langle \log L_\nu(2500 \text{ Å}) \rangle$ value predicted by the bisector best-fit relation of Lusso et al. (2010; red dashed line of Figure 5.4) for a given $\log L_\nu(2 \text{ keV})$. The red dashed line shows the best-fit Gaussian distribution. The mean value of that distribution is consistent with zero and the standard deviation is $\sigma = 0.4$ 93

- 5.6 Fraction of *XMM*-XXL X-ray sources with successful redshift measurement as a function of the *r*-band magnitude of the optical counterpart. The fraction is defined as the ratio between the number of potential targets [$f_X(0.5 - 10 \text{ keV}, \Gamma = 1.4) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ and secure optical counterparts with $r < 22.5 \text{ mag}$] and X-ray sources with successful redshift measurements and fluxes/magnitudes within the above cuts. 93
- 5.7 Impact of different levels of obscuration on the 0.5-10 keV flux of a source as a function of redshift. The vertical axis is the ratio of the flux of an AGN obscured by column density N_H relative to the flux of the same source in the case of zero obscuration. The red curves correspond to different levels of obscuration in the range $\log N_H = 22.5 - 24.5 \text{ (cm}^{-2} \text{ units)}$. The vertical dashed lines mark the redshift interval of interest, $z = 3 - 5$. For fixed obscuration and intrinsic luminosity the flux of higher redshift AGN is less affected. The differential flux suppression between $z = 3$ and $z = 5$ is small, $\lesssim 10\%$. For the calculation of X-ray k-corrections we adopt the model X-ray spectral of Brightman and Nandra (2011b). These are based on Monte Carlo simulations of an illuminating source at the centre of a sphere with constant density and a conical region (apex at the centre of the sphere) cut-off to approximate a toroidal geometry. These simulations take into account both Compton scattering and photoelectric absorption of the X-ray photons by the obscuring medium. We adopt $\Gamma = 1.9$ for the intrinsic AGN spectrum, an opening angle of the conical region of 60 deg and a viewing angle of 45 deg, i.e. a line-of-sight intersecting the obscuring material. . . . 94

- 5.8 AGN space density as a function of 2-10 keV X-ray luminosity in the redshift intervals $z = 3-4$ and $z = 4-5$. In both panels the datapoints are the non-parametric binned X-ray luminosity function. The errors correspond to the 68th percentiles of the Probability Distribution Function. The black filled circles are the estimates for the combined *XMM-XXL* and *Chandra* deep field surveys. In the $z = 3-4$ panel (left) we also plot the constraints obtained separately from the *XMM-XXL* (red squares) and the *Chandra* deep survey fields (blue triangles). This is to demonstrate that in overlapping luminosity bins the *XMM-XXL* and *Chandra* data yield consistent results. The shaded regions are the 68% confidence intervals of the LDDE and PDE parametric models described in Section 5.3.4. The gray hatched region is for the LDDE and the pink region is for the PDE. The parametric models are estimated at the middle redshift of the redshift intervals, i.e. $z = 3.5$ and $z = 4.5$ 107
- 5.9 X-ray luminosity dependence of the ratio of the AGN space densities in the redshift intervals $z = 3-4$ and $z = 4-5$. The data points are for the non-parametric binned X-ray luminosity function. A double power-law is also fit independently to the data in the redshift intervals $z = 3-4$ and $z = 4-5$. The ratio of these parametric fits is shown by the light/dark grey shaded regions. The widths of the dark-grey and light-grey shaded regions at a given luminosity corresponds to the 68% and 95% confidence intervals. The predictions of the Gilli et al. (2007) LDDE parametrisation of the luminosity function with (dashed line) and without (dot-dashed line) an exponential cutoff at high redshift is also plotted for comparison. 110

- 5.10 Comparison of our best-fit LDDE model (pink shaded region) with previous estimates of the X-ray luminosity function in the range $z = 3 - 5$. The panels correspond to the redshift intervals of Vito et al. (2014) to allow direct comparison with their results. The black filled circles are the $1/V_{\text{max}}$ binned luminosity function estimates of Vito et al. (2014). The thin blue curves are the flexible double power-law parametric model of Aird et al. (2015) for their 0.5-2 keV band selected sample without corrections for obscuration. For the Aird et al. (2015) 0.5-2 keV sample the binned X-ray luminosity function estimates are also shown by the red crosses. The UV/optical QSO luminosity functions of Masters et al. (2012), McGreer et al. (2013), and Ross et al. (2013) are also plotted by the thick black lines at the relevant redshift intervals. 119
- 5.11 Impact of ignoring photometric redshift uncertainties on the inferred AGN space density. The red open circles correspond to the non-parametric binned X-ray luminosity function estimated by using the best-fit photometric redshift solution only and ignoring photometric redshift uncertainties (see text for details). The black data points correspond to the AGN space density estimated by using the full photometric redshift Probability Distribution Function to account for photometric redshift errors. For both set of data-points only the *Chandra* survey fields are used. 120
- 5.12 X-ray luminosity density as a function of redshift. The black dashed curve corresponds to the total X-ray luminosity density for both obscured (Compton thin and Compton thick) and unobscured AGN determined by Aird et al. (2015). The shaded blue region is the X-ray luminosity density estimated using our LDDE parametrisation for the evolution of the AGN population at $z > 3$. The inset plot shows the ratio between the blue (solid) and black (dashed) curves as a function of redshift. It represents the fraction of the luminosity density accounted by our LDDE model relative to the total luminosity density (corrected for obscuration effects) estimated by the Aird et al. (2015). 120

- 5.13 Type-I AGN fraction as a function of X-ray luminosity. The red shaded region is the ratio between the $z = 3.2$ luminosity functions of UV/optical QSOs from Masters et al. (2012) and our LDDE parametrisation for X-ray AGN. The black shaded region is $1 - \mathcal{F}_{\text{obscured}}$, where $\mathcal{F}_{\text{obscured}}$ is the obscured AGN fraction of Buchner et al. (2015) for the redshift interval $z = 2.7 - 4$ 123
- 5.14 Hydrogen Ionising photon rate density as a function of redshift. The shaded regions are the constraints from our analysis using the LDDE parametrisation for the X-ray luminosity function and under different assumptions on the escape fraction of AGN photons. The grey-shaded region assumes an escaping fraction of unity, i.e. ignoring obscuration effects close to the supermassive black hole. The pink-shaded region assumes the luminosity-dependent Type-1 AGN fraction of Merloni et al. (2014). We caution that beyond $z = 5$ the shaded curves are extrapolations. The data points correspond to results in literature. The thick black line in the plot shows the photon rate density required to keep the Universe ionised at any given redshift. The ratio between the shaded regions and the black line are presented in the inset plot. 125
- 5.15 Number of high redshift AGN as a function of 2-10 keV X-ray luminosity that the eROSITA 4-year All Sky Survey is expected to detect. The LDDE parametrisation of the X-ray luminosity function is used for the predictions. Results for 3 redshift intervals are plotted, $z = 3 - 4$ (blue solid), $z = 4 - 5$ (red dashed) and $z = 5 - 6$ (black dot-dashed). The predictions for the latter redshift bin are extrapolations of the LDDE model. The numbers are for 0.5 dex wide luminosity bins. 125

List of Tables

2.1	Median probability models using the proposed methodology (NHPG), the methodology proposed by Meligkotsidou and Dellaportas (2011) (M&D) and the model of Narisetty and He (2014) (BASAD), respectively. The first row (ME) shows the covariates used in the Mean Equation and the second row (TM) the covariates of the Transition Matrix. The proposed methodology is the only one able to identify the true data generating process.	23
2.2	Forecasting performance of the competing models. In addition to M&D and the BASAD, we include also the HHMM. The best performance (bold values) for each criterion is achieved by the proposed NHPG model.	24
2.3	Summary of runtimes (in seconds per 1000 iterations).	24
2.4	Posterior probabilities of inclusion for the competing models. Predictors with inclusion probability above 0.5 (median probability model) are marked with bold values. NHPG ₀ and NHPG ₁ denote the proposed methodology without autoregressive terms and with one autoregressive term respectively, M&D the methodology proposed by Meligkotsidou and Dellaportas (2011), HHMM the Homogeneous model with variable selection using a RJ-step and CSS the model of Christiansen, Schmeling, and Schrimpf (2012). ME stands for Mean Equation (linear relationship) and TM for Transition Matrix (non linear relationship). The HHM and CSS models included covariates only in the ME.	30
2.5	Summary of forecasting results of the five competing models, obtained from the log-realized volatility dataset. The best performance (bold values) for each criterion is achieved by the NHPG ₁ model (with one autoregressive term).	31

2.6	Summary of results: CRPS is the mean continuous rank probability score, MAFE is the mean absolute forecast error and MSFE is the mean squared forecast error. ESS is the minimum effective size of among the ESS for all parameters and MESS the multivariate effective size, for an MCMC run of 25000 iterations. mESR is the minimum effective sample rate. PSRF is the maximum potential scale reduction factor and mCM is the multivariate convergence and mixing diagnostic. In the mCM line we report the number of the components of the parameters out of the total components – in parenthesis – that fall into the 95% confidence interval of the test. NHPG is the proposed model, M&D is the model of Meligkotsidou and Dellaportas (2011), HHMM is the homogeneous model. Bold values denote the best values for the corresponding criterion among all the competing models.	35
3.1	List of variables and online resources. The Hash Rate (Hash) has been retrieved from quandl.com for Bitcoin (BTC) and from ether-scan.io for Ether (ETH).	50
3.2	Descriptive statistics (left panels) and p-values for the time series statistical tests (right panels) for the two (2) BTC price, log-price and log-return subseries – first and second line of each entry – which correspond to the two hidden states that were identified by the NHPG model for the whole 1/2014-8/2019 time period.	53
3.3	Posterior mean estimations for the BTC log-return series in the 2014-2019 (left) and 2017-2019 (right) time periods. B_1, B_2 are the mean equation coefficients and β_1, β_2 are the logistic regression coefficients for states 1,2. Statistically significant coefficients (at the 0.05 level) are marked with *.	54
3.4	Posterior mean estimations for the ETH log-return series in the 2017-2019 time period. Statistically significant coefficients (at the 0.05 level) are marked with *.	56
3.5	Descriptive statistics (left panels) and p-values of the time series statistical tests (right panels) for the BTC and ETH price, log-price and log-return series. DF denotes the Dickey-Fuller test, KPSS the Kwiatkowski-Phillips-Schmidt-Shin test, LBQ is the Ljung-Box Q test, VR the variance ratio test and JB the Jarque-Bera test.	58

4.1	Summary Statistics of the percentage logarithmic return cryptocurrency series and transformed predictors. The first column reports the transformation of each variable. The second column displays the mean. The third column reports the standard deviation. Third to seventh columns display the minimum values, the 5%,50%, 95% quantiles and the maximum values respectively. Last two columns display the kurtosis and skewness coefficients.	67
4.2	Summary of the models of this study. The first two columns show the model and its abbreviation, the third and fourth columns show the assumed relation of the studied time series and the predictors. The fifth column shows the assumed parametrization of the transition probabilities of each model. The last column shows the various number of states we considered for each model.	69
4.3	Continuous Rank Probability Score and Mean Squared Error in parenthesis for all the competing models for the Bitcoin, Ether and Ripple series. The last column reports the average CRPS (MSE) over the whole sequence of 30 one-step ahead predictions. Bold values indicate the lowest CRPS values for each out-of-sample points.	71
4.4	Empirical coverage of the empirical in-sample distributions using the 4-state NHHM and the RW benchmark for the Bitcoin, Ether and Ripple percentage return series.	74
4.5	Posterior means estimates of the 4-state Non-Homogeneous Hidden Markov model for the Bitcoin, Ether Ripple percentage return series. The first column specifies the predictors. The second, third, fourth and fifth columns report the posterior mean estimates for each predictor at the first, second and third states respectively. The last row reports the mean estimated residual variance for each state. The last column reports the posterior probabilities of inclusion for the predictors affecting the transition probabilities' multinomial regression model. These probabilities are calculated by applying a stochastic search reversible jump algorithm within the MCMC scheme. Statistical significance at the 10% level is denoted with * and posterior probabilities exceeding 0.4 are marked with bold fonts.	75

4.6	Information on the states as derived from the experiments on the BTC, ETH, XRP return series. First column reports the cryptocurrencies and the second column the different regimes. The third column reports the states' occupancies, i.e., the average time spend at each regime. The fourth column reports the average returns at each state and finally, the fifth column reports the state's estimated standard deviation.	77
5.1	Number of X-ray sources in the full-band selected sample	85
5.2	(1) Name of the X-ray survey fields used in this work. The last row lists the total number of sources for the combined fields. For that last row only the columns that correspond to the total number of sources with photometric or spectroscopic redshift in the intervals $3 < z < 4$ and $4 < z < 5$ are listed. This is because these are the relevant numbers to the analysis presented in this paper. (2) Solid angle of each sample in square arcminutes after excluding regions of poor/no photometry. (3) Total number of X-ray sources detected in the 0.5-7 keV energy band in the case of the <i>Chandra</i> , or the 0.5-8 keV band for the XMM-XXL sample. (4) Number of full-band detected X-ray sources with optical/infrared associations. (5) Number of sources with photometric redshift estimates in each survey field. Photometric redshifts for the XMM-XXL are not available. (6) Number of sources with photometric redshifts in the interval $z = 3 - 4$. This the sum of the photometric redshift Probability Distribution Functions rounded to the nearest integer. No photometric redshift estimates are available for the XMM-XXL sample. (7) Same as column (6) but for the redshift interval $z = 4 - 5$. (8) Number of sources with spectroscopic redshift estimates in each survey field. (9) Number of sources with spectroscopic redshifts in the interval $z = 3 - 4$. (10) Number of sources with spectroscopic redshifts in the interval $z = 4 - 5$	85
5.3	Power-law fits to the data in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$	111
5.4	Parametric model best-fit parameters and uncertainties	124

5.5 Listed are the best-fit parameters for each of the four parametric models considered in this paper. The listed values are the median of the probability distribution function of each parameter. The errors correspond to the 16th and 84th percentiles around the median. The columns are: (1) parametric model. Models are listed in order of decreasing Bayesian evidence, (2) X-ray luminosity function normalisation (see equation 5.5), (3) break luminosity of the X-ray luminosity function (see equation 5.5), (4) faint-end slope, (5) bright-end slope, (6) density evolution parameter (see equations 5.7, 5.8, 5.9), (7) luminosity evolution parameters (see equations 5.6, 5.8), (8) β parameter for the LDDE model, (9) base 10 logarithm of the Bayesian evidence for each model, (10) the difference between the $\log_{10} \mathcal{Z}$ of each model and the PDE that has the highest evidence. The bayes factor of the PDE model relative to each of the other three is $\exp(\Delta \ln \mathcal{Z})$.¹²⁴

Chapter 1

Introduction

The present dissertation is divided into two main areas. The first research area is presented in Chapters 2 to 4. In a nutshell, we consider a Bayesian modeling methodology for explaining and predicting financial time series. Particularly, we proposed and developed a new class of Non-Homogeneous Hidden Markov models with a variable selection scheme and applied it in realized volatility and cryptocurrency time series data. The second research area concerns a Bayesian Poisson process approach to account for the determination and estimation errors in astrophysics data, cf. Chapter 5. The structure of this dissertation is as follows.

In the second chapter of this thesis, we consider finite-state space Non-Homogeneous Hidden Markov Models for modeling and forecasting univariate time series. Given a set of predictors, the time series are modeled via predictive regressions with state dependent coefficients and time-varying transition probabilities that depend on the predictors via a logistic/multinomial function. In a hidden Markov setting, inference for logistic regression coefficients becomes complicated and in some cases impossible due to convergence issues. In this work, we aim to address this problem utilizing the recently proposed Pólya-Gamma latent variable scheme. Also, we allow for model uncertainty regarding the predictors that affect the series both linearly – in the mean – and non-linearly – in the transition matrix. Predictor selection and inference on the model parameters are based on an automatic MCMC scheme with reversible jump steps. Hence, the proposed methodology can be used as a black box for predicting time series. Using simulation experiments, we illustrate the performance of our algorithm in various setups, in terms of mixing properties, model selection and predictive ability. An empirical study on realized volatility data shows that our methodology gives improved forecasts compared to benchmark models.

Next, Chapter 3 reports an additional application of the aforementioned Non-Homogeneous Pólya-Gamma Hidden Markov Model (NHPG) on cryptocurrencies returns series. Due to the cryptocurrencies dual nature, conventional financial models

fail to explain the economic and monetary properties. Their usage as financial assets on the one side and their tight connection to the underlying blockchain structure on the other. In an effort to examine both components via a unified approach, we apply the NHPG model with an extended set of financial and blockchain specific covariates on the Bitcoin (BTC) and Ether (ETH) price data. Based on the observable series, the NHPG model offers a novel perspective on the underlying microstructure of the cryptocurrency market and provides insight on unobservable parameters such as the behavior of investors, traders and miners. The algorithm identifies two alternating periods (hidden states) of inherently different activity – fundamental versus uninformed or noise traders – in the Bitcoin ecosystem and unveils differences in both the short/long run dynamics and in the financial characteristics of the two states, such as significant explanatory variables, extreme events and varying series autocorrelation. In a somewhat unexpected result, the Bitcoin and Ether markets are found to be influenced by markedly distinct indicators despite their perceived correlation. The current approach backs earlier findings that cryptocurrencies are unlike any conventional financial asset and makes a first step towards understanding cryptocurrency markets via a more comprehensive lens.

Extending the work presented in Chapter 3, we leverage the Bayesian Hidden Markov models – including the aforementioned m-states Non-Homogeneous Hidden Markov Pólya-Gamma – in predicting cryptocurrencies in Chapter 4. In particular, we consider a variety of multi-state Hidden Markov models for predicting and explaining the Bitcoin, Ether and Ripple returns in the presence of state (regime) dynamics. In addition, we examine the effects of several financial, economic and cryptocurrency specific predictors on the cryptocurrency return series. Our preliminary results indicate that the 4-states Non-Homogeneous Hidden Markov model has the best one-step-ahead forecasting performance among all the competing models for all three series. The superiority of the predictive densities, over the single regime random walk model, relies on the fact that the states capture alternating periods with distinct returns' characteristics. In particular, we identify bull, bear and calm regimes for the Bitcoin series, and periods with different profit and risk magnitudes for the Ether and Ripple series. Finally, we observe that conditionally on the hidden states, the predictors have different linear and non-linear effects.

Lastly, the fifth chapter presents the research output of the collaboration with the National Observatory of Athens and the Max Plank Institut für Physics in Munich. Our contribution to this research lies in the Bayesian modeling and implementation of the Active Galactic Nuclei (AGN) X-ray luminosity function. In detail,

we combine deep X-ray survey data from the *Chandra* observatory and the wide-area/shallow XMM-XXL field to estimate the AGN X-ray luminosity function in the redshift range $z = 3 - 5$. The sample consists of nearly 340 sources with either photometric (212) or spectroscopic (128) redshift in the above range. The combination of deep and shallow survey fields also provides a luminosity baseline of three orders of magnitude, $L_X(2 - 10 \text{ keV}) \approx 10^{43} - 10^{46} \text{ erg s}^{-1}$ at $z > 3$. We follow a Bayesian approach to determine the binned AGN space density and explore their evolution in a model-independent way. Our methodology properly accounts for Poisson errors in the determination of X-ray fluxes and uncertainties in photometric redshift estimates. We demonstrate that the latter is essential for unbiased measurement of space densities. We find that the AGN X-ray luminosity function evolves strongly between the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. There is also suggestive evidence that the amplitude of this evolution is luminosity dependent. The space density of AGN with $L_X(2 - 10 \text{ keV}) < 10^{45} \text{ erg s}^{-1}$ drops by a factor of 5 between the redshift intervals above, while the evolution of brighter AGN appears to be milder. Comparison of our X-ray luminosity function with that of UV/optical selected QSOs at similar redshifts shows broad agreement at bright luminosities, $L_X(2 - 10 \text{ keV}) > 10^{45} \text{ erg s}^{-1}$. At fainter luminosities X-ray surveys measure higher AGN space densities. The faint-end slope of UV/optical luminosity functions however, is steeper than for X-ray selected AGN. This implies that the type-I AGN fraction increases with decreasing luminosity at $z > 3$, opposite to trends established at lower redshift. We also assess the significance of AGN in keeping the hydrogen ionised at high redshift. Our X-ray luminosity function yields ionising photon rate densities that are insufficient to keep the Universe ionised at redshift $z > 4$. A source of uncertainty in this calculation is the escape fraction of UV photons for X-ray selected AGN.

Chapter 2

Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Pólya-Gamma data augmentation

2.1 Introduction

Discrete-time finite state-space Homogeneous Hidden Markov Models (HHMMs) have been extensively studied and used to model stochastic processes that consist of an observed process and a latent (hidden) sequence of states which is assumed to affect the observation sequence, see for example Cappé, Moulines, and Ryden (2005) and Billio, Monfort, and Robert (1999). Bayesian inference, using Markov Chain Monte Carlo (MCMC) techniques, has enhanced the applicability of HHMMs and has led to the construction of more complex model specifications including Non-Homogeneous Hidden Markov Models (NHHMMs). Initially, Diebold, Lee, and Weinbach (1994) studied the 2-state Gaussian NHHMMs where the time varying transition probabilities were modeled via logistic functions. Their approach was based on the Expectation-Maximization algorithm (EM). Filardo and Gordon (1998) adopted a Bayesian perspective to overcome technical and calculation issues of classical approaches. Since then, various Bayesian methods have been proposed in the literature. For example, Spezia (2006) modeled the time-varying transition probabilities via a logistic function depending on exogenous variables and performed model selection based on the Bayes factor. In the same spirit, Meligkotsidou and Dellaportas (2011) considered an m -state ($m \geq 2$) NHHMM and assumed that the elements of

the transition matrix are linked through exogenous variables with a multinomial logistic link, whereas the observed process conditional on the unobserved process follows an autoregressive model of order p . They accommodated and exploited model uncertainty within their Bayesian model – by allowing covariate selection only on the transition matrix – to improve the predictive ability of NHHMMs on economic data series.

Based on experimental evidence, the algorithm of Meligkotsidou and Dellaportas (2011) (M&D) faces convergence issues when there exists model uncertainty, due to the data augmentation scheme of Holmes and Held (2006). Polson, Scott, and Windle (2013) confirm the efficiency issues in the Holmes and Held (2006) scheme and propose a Pólya-Gamma data augmentation strategy that significantly improves over various benchmarks, (O’Brien and Dunson, 2004; Frühwirth-Schnatter and Frühwirth, 2010; Fussl, Frühwirth-Schnatter, and Frühwirth, 2013). Furthermore, the recent work of Holsclaw et al. (2017) confirms that using Pólya-Gamma data augmentation to parametrize the transition probabilities of NHHMMs results in an algorithm that mixes well and provides adequate estimates of the model parameters.

Motivated by this, we revisit the work of Meligkotsidou and Dellaportas (2011) by employing the recent methodological advances on the Pólya-Gamma data augmentation scheme of Polson, Scott, and Windle (2013). We consider NHHMMs in which the time series are modeled via different predictive regression models for each state, whereas the transition probabilities are modeled via logistic regressions. Given an available set of predictors, we allow for model uncertainty regarding the predictors that affect the series both linearly – directly in the mean regressions – and non-linearly – in the transition probability matrix.

The resulting model is a Non-Homogeneous Pólya-Gamma Hidden Markov Model, which we will denote by NHPG. Bayesian inference is performed via a MCMC scheme which overcomes difficulties and convergence issues inherent in existing MCMC algorithms. To this end, we exploit the missing data representation of hidden Markov models and construct an MCMC algorithm based on data augmentation, consisting of several steps. First, we sample the latent sequence of states via the Scaled Forward-Backward algorithm of Scott (2002), which is a modification of the Forward-Backward algorithm of Baum et al. (1970) who used it to implement the classical EM algorithm. Then, we use a logistic regression representation of the transition probabilities and simulate the parameters of the mean predictive regression model for each state, via Gibbs sampling steps. Finally, we incorporate variable selection within our MCMC scheme by using the Reversible Jump (RJ) algorithm of

Green (1995) and Hastie and Green (2011).

Different approaches have been used in the literature to cope with the model selection problem. The use of information criteria, such as Akaike's Information Criterion (AIC), (Akaike, Petrov, and Csaki, 1973), the Bayesian Information Criterion (BIC) of Schwarz, 1978, the Deviance Information Criterion (DIC), (Spiegelhalter et al., 2002) or the Widely applicable Bayesian Information Criterion (WBIC), (Watanabe, 2013), is another approach to variable selection. A study for comparing variable selection methods is well presented in O'Hara and Sillanpää (2009) whilst Dellaportas, Forster, and Ntzoufras (2002) study the variable selection methods in the context of model choice. Holsclaw et al. (2017) consider a NHHMM similar to ours for modeling multivariate meteorological time series data. In that paper, the transition probabilities are modeled via multinomial logistic regressions affected by a specific set of exogenous variables. The authors use the BIC criterion for choosing the best model among a pre-specified class of models. We extend this work by considering the problems of statistical inference and variable selection jointly, in a purely Bayesian setting. The proposed model is flexible, since we do not decide a priori which covariates affect the observed or the unobserved process. Instead, we have a common pool of covariates $\{X\}$ and within the MCMC algorithm, we gauge which covariates are included in subset $\{X^{(1)}\}$ affecting the mean predictive equation of the observed process, and which covariates are included in subset $\{X^{(2)}\}$ affecting the time-varying transition probabilities.

Our probabilistic approach is based on the calculation of the posterior distribution of different NHPGs. Posterior probabilities can be used either for selecting the most probable model (i.e., making inference using the model with the highest posterior probability), or for Bayesian model averaging (i.e., producing inferences averaged over different NHPGs). Barbieri and Berger (2004) argue that the optimal predictive model is not necessarily the model with the highest posterior probability. Specifically, they show that the optimal predictive model for linear regression models is the median probability model, i.e. the model that is consisted of those covariates which have overall posterior inclusion probabilities greater or equal to 0.5. We calculate both the posterior probabilities of the models and the probabilities of inclusion.

We use our model for predicting realized volatility. Accurate forecasting of future volatility is important for asset allocation, portfolio construction and risk management, (Gospodinov, Gavala, and Jiang, 2006). A review on the realized volatility literature can be found in McAleer and Medeiros (2008). The relationship between the volatility and macroeconomic and/or financial variables is investigated in Paye

(2012) and Christiansen, Schmeling, and Schrimpf (2012) and Meligkotsidou et al. (2019) among others. The proposed NHPG captures not only the linear relationship between the logarithm of realized volatility and a set of predictors, as in the model of Christiansen, Schmeling, and Schrimpf (2012) (CSS), but also the nonlinear relationship, as well as other special characteristics of the analyzed series, such as heteroscedasticity and autocorrelation. NHPG outperforms the M&D, CSS models and the HHMM, in terms of forecasting ability.

The MCMC output of the predictive density of the NHPG is multimodal and thus, scoring rules that are not sensitive to distance should be avoided (Gneiting and Raftery, 2007). For instance the logarithmic scoring rule gives harsh penalty for low probability events (Boero, Smith, and Wallis, 2011; Gneiting and Raftery, 2007) and prefers the forecast density that is less informative (Machete, 2013). In this case, a better alternative not only for validating the model performance but also for assessing the quality of forecasts is the Continuous Rank Probability Score (CRPS). This proper scoring rule has gained a lot of interest in the meteorological community, see Gruit et al. (2006), and proves to be the most appropriate rule also for the NHPG model.

In summary, the main contributions of this study are the following

1. We propose a flexible model (NHPG) that can detect the linear and a non-linear relationship between the predictors and the studied time series. This results in a stable algorithm which does not need tuning and can be used as a black box for predicting time series.
2. We present experimental evidence in support of the claim that the NHPG model has improved performance in terms of variable selection and forecasting ability when compared with M&D. This is at no cost of computational complexity and running time.
3. We provide evidence that the proposed algorithm performs well also with real datasets by obtaining improved forecasts on the realized volatility data set of Christiansen, Schmeling, and Schrimpf (2012).

The paper proceeds as follows: In Section 2.2, we outline the proposed model and in Section 2.3, we describe our Bayesian computational strategy both with and without model uncertainty. Section 2.4, presents our forecasting criteria. Section 2.5 contains numerical experiments and Section 2.6 proceeds with the main application on the realized volatility data set. Finally, Section 2.7 concludes the paper. The extension of the 2-state to the general m -state, $m \geq 2$, NHPG model along with

some technical details and a case study without model uncertainty are deferred to the Appendix.

2.2 The Non-Homogeneous Pólya-Gamma Hidden Markov Model

The proposed 2-state NHPG for univariate time series is described as follows. Consider an observed random process $\{Y_t\}$ and a hidden underlying process $\{Z_t\}$ which is a 2-state non-homogeneous discrete-time Markov chain that determines the states of the observed process. Let y_t and z_t be the realizations of the random processes $\{Y_t\}$ and $\{Z_t\}$, respectively. We assume that at time t , $t = 1, \dots, T$, y_t depends on the current state z_t and not on the previous states. Consider also a set of $r - 1$ available predictors $\{X_t\}$ with realization $x_t = (1, x_{1t}, \dots, x_{r-1t})$ at time t . A subset of the predictors $X_t^{(1)} \subseteq \{X_t\}$ of length $r_1 - 1$ is used in the regression model for the observed process and a subset $X_t^{(2)} \subseteq \{X_t\}$ of length $r_2 - 1$ is used to describe the dynamics of the time-varying transition probabilities. Thus, we allow the covariates to affect the observed process $\{Y_t\}$ non-linearly.

The observed random process $\{Y_t\}$ can be written in the form

$$Y_t = g(Z_t) + \epsilon_t,$$

where $g(Z_t) = X_{t-1}^{(1)} B_{Z_t}$ is a linear function, $B_{Z_t} = (b_{0Z_t}, b_{1Z_t}, \dots, b_{r_1-1Z_t})'$ are the regression coefficients and $\epsilon_t \sim \mathcal{N}(0, \sigma_{Z_t}^2)$. We use $\mathcal{N}(\mu, \sigma^2)$ to denote the normal distribution with mean μ and variance σ^2 . In a less formal way, if s represents the hidden states, the observed series given the unobserved process has the form

$$Y_t | Z_t = s \sim \mathcal{N}(X_{t-1}^{(1)} B_s, \sigma_s^2), \quad s = 1, 2.$$

The dynamics of the unobserved process $\{Z_t\}$ can be described by the time-varying transition probabilities, which depend on the predictors $X_t^{(2)}$ and are given by the following relationship

$$P(Z_{t+1} = j | Z_t = i) = p_{ij}^{(t)} = \frac{\exp(x_t^{(2)} \beta_{ij})}{\sum_{j=1}^2 \exp(x_t^{(2)} \beta_{ij})}, \quad i, j = 1, 2,$$

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \dots, \beta_{r_2-1,ij})'$ is the vector of the logistic regression coefficients to be estimated. Note that for identifiability reasons, we adopt the convention of

setting, for each row of the transition matrix, one of the β_{ij} to be a vector of zeros. Without loss of generality, we set $\beta_{ij} = \beta_{ji} = 0$ for $i, j = 1, 2, i \neq j$. Hence, for $\beta_i = \beta_{ii}$, $i = 1, 2$ probabilities can be written in a simpler form

$$p_{ii}^{(t)} = \frac{\exp(x_t^{(2)}\beta_i)}{1 + \exp(x_t^{(2)}\beta_i)} \text{ and } p_{ij}^{(t)} = 1 - p_{ii}^{(t)}, \quad i, j = 1, 2, \quad i \neq j.$$

To model the probabilities of staying at the same state for two consecutive time periods, i.e. p_{ss}^t , we define, for $t = 1 \dots, T - 1$, the quantity $\tilde{Z}_{t+1}^s = I[Z_{t+1} = Z_t = s]$. Therefore, the sum $\sum_t \tilde{Z}_{t+1}^s$, is the number of times that the chain was at state s for two consecutive time periods. Then,

$$p(\tilde{Z}_{t+1}^s = 1 \mid x_t^{(2)}) = p_{ss}^t = \frac{\exp(x_t^{(2)}\beta_s)}{1 + \exp(x_t^{(2)}\beta_s)} \Leftrightarrow \text{logit}(p_{ss}^{(t)}) = x_t^{(2)}\beta_s, \quad s = 1, 2.$$

Summing up, the unknown quantities of the NHPG are $\{\theta_s = (B_s, \sigma_s^2), \beta_s, s = 1, 2\}$, i.e., the parameters in the mean predictive regression equation and the parameters in the logistic regression equation for the transition probabilities of the unobserved process $\{Z_t\}$, $t = 1, \dots, T$. We use conditional conjugate analysis to make inference on θ . Specifically, we assume that $\sigma_s^2 \sim \mathcal{IG}(p, q)$ and $B_s \mid \sigma_s^2 \sim \mathcal{N}(L_0, \sigma_s^2 V_0)$, $s = 1, 2$, where \mathcal{IG} denotes the Inverted-Gamma distribution. To make inference about the logistic regression coefficients, we use the auxiliary variables (data augmentation) method of Polson, Scott, and Windle, 2013 as described in Subsection 2.3.2.1. In brief, given the Pólya-Gamma auxiliary variables, a conjugate prior for the logistic regression coefficients β_s , $s = 1, 2$ is multivariate Normal distribution $\mathcal{N}(m_{\beta_s}, V_{\beta_s})$

2.2.1 Label Switching

The m -states HMMs often suffer from the so called label switching problem, see for example Spezia (2009) and Jasra, Holmes, and Stephens (2005) and Marin, Mengersen, and Robert (2005). In a nutshell, the $m!$ models that arise from the $m!$ ways to label the hidden states can be interchangeable, if no information on the parameters is available. Using uninformative priors as in Section 2.2, can lead to unidentifiable – possibly multimodal – posterior distributions.

In this study we focus in the predictive densities and in defining the predictors that affect the series, which are invariant to label switching (Dellaportas and Papageorgiou, 2006; Frühwirth-Schnatter, 2001). Nevertheless, we propose a hybrid

methodology to tackle the label switching problem if estimation of the unknown parameters – based either on the most probable model or on the median model – is desired. In short, we propose a two-layer procedure, i.e. the use of the random permutation sampler of Frühwirth-Schnatter (2001) and a modification of the online reordering algorithm of Marin, Mengersen, and Robert (2005).

2.2.2 Tackling the label switching problem

Estimation of the unknown parameters, based either on the median probability model or on the most probable model, can be challenging if label switching occurs. Using the MCMC output for inference can lead to inaccurate estimates (Frühwirth-Schnatter, 2001). Even though there have been proposed many different strategies to tackle this problem – see for example Rodríguez and Walker (2014) and references therein – in high-dimensional models (as the NHPG) finding a “fix” for the label switching problem is difficult, (Dellaportas and Papageorgiou, 2006). Our approach to this problem is a hybrid mix of two existing methodologies. In the following lines, we present in brief these methodologies. More details can be found in the referred articles.

Specifically, we propose to use the permutation sampler of Frühwirth-Schnatter (2001), which introduces online identifiability constraints via a data-driven procedure. The permutation sampler is a two-step procedure. First, we run an unconstrained sampler, followed by a randomly selected permutation at the current labeling, to explore the whole label-subspaces, e.g. if $m = 2$, change 1s to 2s with probability 0.5 and let 1s remain the same with probability 0.5. Then, we perform a graphical analysis of the unconstrained MCMC output, i.e. using the bivariate scatterplots, among all possible combinations of the mean equation parameters’ components. Based on the plots, if there are groups that correspond to different states and these groups imply a special ordering on the labels, then we define the “suitable” identification constraint. In the second step, we perform a permutation sampling under the identification constraint, as derived by the visual inspection of the unconstrained MCMC output. At each iteration of the MCMC algorithm, we permute the parameters and the hidden states until the identification constraint is fulfilled. More details on the permutation sampler for NHHMMs can be found in Spezia (2006) and Paroli and Spezia (2008), among others.

However, if no information on the constraints is available from the permutation sampler or label switching still occurs even after the constrained sampling, then we propose to use an online reordering algorithm. This algorithm is described in a

NHHMM setting in Spezia (2009) and Pinto and Spezia (2016) and it is based on the lines of Marin, Mengersen, and Robert (2005) for mixture models. Specifically, we propose to run the MCMC sampling scheme for $L + M + N$ iterations. After the L -sized burn-in period, we estimate iteratively the parameters' mode θ^* , based on an M -sized sample. Then, we store the N -sized reordered simulated parameter values and the permuted estimated hidden process. The reordering is based on the permutation η^* of the parameters that is closest to the Euclidean distance of the mode. In particular, for every iteration $i, i = 1, \dots, N$, if H is the class of $m!$ permutations, it is based on the permutation η^* defined by

$$\eta^* = \arg \min_{\eta_j \in H} \|\eta_j(\theta^{(i)}) - \theta^*\|.$$

Then, the permutation η^* is applied to the unknown parameters $\theta^{(i)} = \eta^*(\theta^{(i)})$, $\beta^{(i)} = \eta^*(\beta^{(i)})$, $i = 1, \dots, N$, and the estimated hidden process $Z^{(i)}$.

To sum up, the hybrid algorithm for inference on the parameters while dealing with the label switching problem, is consisted of five steps.

1. Run the MCMC algorithm for the NHPG with RJ-step, as in Section 2.3.1 excluding step (7).
2. Find the most probable or the median probability model. → Define the fixed model.
3. Run the *Unconstrained random permutation sampler* of Frühwirth-Schnatter (2001).
4. (a) *If* there exists a suitable constraint
→ Apply the *Permutation sampler under an identifiability constraint* of Frühwirth-Schnatter, 2001.
- (b) *Else*
→ Run the Online post-processing reordering algorithm of Marin, Mengersen, and Robert (2005).
5. Use the MCMC output to make inference on the parameters of the NHPG model.

2.3 Bayesian Inference and Computational Strategy

The key steps in our proposed framework are the following. First, for a given NHPG, we construct a Markov chain which has as stationary distribution the posterior distribution of the model parameters. Simulation of this Markov chain provides, after

some burn in period and adequately many iterations, samples from the posterior distribution of interest; see, for details, Besag et al. (1995). Second, for a given set of competing models, each including a different set of predictors in the mean regression and/or in the transition probabilities equation, we base our inference about the models on their posterior probabilities. This improves over the approach which considers the models separately and chooses the best model via significance tests or via model selection criteria.

2.3.1 The MCMC Sampling Scheme

The main steps of the proposed MCMC algorithm for joint inference on model specification and model parameters are the following.

1. Generate the initial values of $\beta, \theta = (B, \sigma^2)$ from their corresponding prior distribution.
2. Calculate the probabilities of the time-varying transition matrix, using the initial values.
3. Given the model's parameters, simulate the hidden states using the Scaled Forward-Backward algorithm of Scott (2002).
4. Simulate the mean regression parameters via Gibbs sampling.
5. Simulate the coefficients β using the Pólya-Gamma representation of Polson, Scott, and Windle (2013).
6. Use a couple of reversible jump steps to update: (i) the set of covariates that affect the transition matrix (i.e. the hidden process) and (ii) that affect the mean regression model (i.e. the observed process).
7. Make one-step-ahead predictions conditional on the simulated unknown quantities.
8. Repeat steps 3-6 until convergence and then repeat steps 3-7.

In the next subsections, we present each step in detail.

2.3.2 Inference for fixed sets of predictors

For a given NHPG, i.e., for fixed sets of predictors used in the mean equation and the transition probabilities $X^{(1)}$ and $X^{(2)}$, respectively, we update in turn (i) the latent variables z^T given the current values of the model parameters by using the scaled Forward-Backward algorithm (Scott, 2002) (ii) the logistic regression coefficients by adopting the auxiliary variables method of Polson, Scott, and Windle (2013) given

the sequence of states z^T , and (iii) the mean regression coefficients conditional on z^T by using the Gibbs sampling algorithm.

Let $y^T = (y_1, \dots, y_T)$ be the history of the observed process, $z^T = (z_1, \dots, z_T)$ the sequence of states up to time T , and let $f_s(\cdot)$ denote the normal probability density function of $Y_t \mid Z_t = s$, $s = 1, 2$ and $\pi_1(z_1)$ the initial distribution of Z_1 . The joint likelihood function of the data, y^T , and the sequence of states, z^T , is given by

$$\begin{aligned} \mathcal{L}(\theta, \beta) &= \pi(y^T, z^T \mid X, \theta) = \pi(y^T \mid z^T, X, \theta) \pi(z^T \mid X, \beta) \\ &= \pi_1(z_1) f_{z_1}(y_1) \prod_{t=1}^T p_{z_t z_{t+1}}^{(t)} f_{z_t}(y_t) \\ &= \prod_{i=1}^2 \prod_{j=1}^2 \left[\prod_{t: z_{t+1}=j} p_{ij}^{(t)} \right] \left(\frac{1}{2\pi\sigma_j^2} \right)^{N_j/2} \exp \left\{ -\frac{1}{2\sigma_j^2} (Y_j - X_j^{(1)'} B_j)' (Y_j - X_j^{(1)'} B_j) \right\}. \end{aligned}$$

We use the notation N_s , $s = 1, 2$ for the number of times the chain was in state s , that is $N_s = \sum_{t=1}^T I(Z_t = s)$, with I denoting the indicator function. If a prior distribution $\pi(\theta, \beta)$ is specified for the model parameters, then inference on all the unknown quantities in the model is based on their joint posterior distribution $\pi(\theta, \beta, z^T \mid y^T) \propto \pi(\theta, \beta) \pi(y^T, z^T \mid \theta, \beta)$.

After some straightforward algebra and using the priors described in Section 2.2, we derive the marginal posterior distribution for the state specific parameters σ_s and conditional posterior distribution for B_s ,

$$\sigma_s^2 \mid y^T, z^T \sim \mathcal{IG} \left(p + \frac{n_s}{2}, q + \frac{1}{2} (L_{0s}' V_{0s}^{-1} L_{0s} + Y_s' Y_s - L_s' V_s^{-1} L_s) \right),$$

$$B_s \mid \sigma_s^2, z^T, y^T \sim \mathcal{N}(L_s, \sigma_s^2 V_s),$$

with $V_s = (V_{0s}^{-1} + X_s^{(1)'} X_s^{(1)})^{-1}$ and $L_s = V_s (V_{0s}^{-1} L_{0s} + X_s^{(1)'} Y_s)$. Finally, the conditional posterior distribution of β_s , $s = 1, 2$, is a multivariate normal distribution, see Section 2.3.2.1.

2.3.2.1 Simulation of the logistic regression coefficients

As we described in Section 2.2, we model the two diagonal elements of the transition probability matrix by linking them to the set of covariates using a logistic link function. We use the data augmentation scheme of Polson, Scott, and Windle (2013) since, as shown in their work, the estimation of logistic regression coefficients using this scheme is superior, in terms of efficiency and simplicity.

The authors introduce a new latent variable scheme, based on the Pólya-Gamma distributions. They proved that binomial likelihoods – or Bernoulli likelihoods as in our simpler case – parametrized by log odds can be represented as mixtures of Gaussian distributions with respect to the Pólya-Gamma distribution. Their main result is that letting $p(\omega)$ be the density of a Pólya-Gamma latent variable ω , with $\omega \sim \mathcal{PG}(b, 0)$, for $b > 0$, the following identity holds for all $a \in \mathbb{R}$,

$$\frac{\exp(\psi)^a}{(1 + \exp(\psi))^b} = 2^{-b} \exp(k\psi) \int_0^\infty \exp(-\omega\psi^2/2) p(\omega) d\omega,$$

with $k = a - b/2$. Furthermore, the conditional distribution of $\omega \mid \psi$ is also Pólya-Gamma, $\mathcal{PG}(b, \psi)$.

When $\psi = x^{(2)}\beta$, the previous identity gives rise to a conditionally conjugate augmentation scheme for binomial likelihoods of logistic parameters. In particular, let \tilde{z}_{t+1}^s be the realization of $\tilde{Z}_{t+1}^s = I[Z_{t+1} = Z_t = s]$. The likelihood for each state s is given by

$$\mathcal{L}(\beta_s) = \prod_{t=1}^{N_s} \left\{ \frac{\exp(x_t^{(2)}\beta_s)}{1 + \exp(x_t^{(2)}\beta_s)} \right\}^{\tilde{z}_t^s} \left\{ \frac{1}{1 + \exp(x_t^{(2)}\beta_s)} \right\}^{1-\tilde{z}_t^s} = \prod_{t=1}^{N_s} \frac{\exp(x_t^{(2)}\beta_s)^{\tilde{z}_t^s}}{1 + \exp(x_t^{(2)}\beta_s)}.$$

Using the result of Polson, Scott, and Windle (2013) with $k_t^s = \tilde{z}_t^s - 1/2$ and setting $\Omega_s = \text{diag}\{\omega_{1,s}, \dots, \omega_{N_s,s}\}$, the augmented likelihood for each state $s = 1, 2$ is proportional to

$$\mathcal{L}(\beta_s, \omega_s) \propto \prod_{t=1}^{N_s} \frac{1}{2} \exp(k_t^s x_t^{(2)}\beta_s) \int_0^\infty \exp\left\{-\omega_{t,s} (x_t^{(2)}\beta_s)^2 / 2\right\} p(\omega_{t,s}) d\omega_{t,s}.$$

Conditioning on Ω_s , one can derive the expression for the logistic coefficients,

$$\begin{aligned} \pi(\beta \mid z^t, \omega) &\propto \prod_{t=1}^{N_s} \exp\left\{k_t x_t^{(2)}\beta_s - \frac{\omega_t}{2} (x_t^{(2)}\beta_s)^2\right\} \pi(\beta_s) \\ &\propto \pi(\beta) \prod_{t=1}^{N_s} \exp\left\{-\frac{\omega_t}{2} \left((x_t^{(2)}\beta_s)^2 - \frac{2k_t x_t^{(2)}\beta_s}{\omega_t}\right)\right\} \\ &\propto \pi(\beta) \prod_{t=1}^{N_s} \exp\left\{-\frac{\omega_t}{2} \left((x_t^{(2)}\beta_s)^2 - \frac{2k_t x_t^{(2)}\beta_s}{\omega_t} + \frac{k_t^2}{\omega_t^2}\right)\right\}. \end{aligned}$$

Assuming as prior distributions $\omega \sim \mathcal{PG}(1, 0)$ and $\beta \sim \mathcal{N}(m_{\beta_0}, V_{\beta_0})$, the posterior distributions can be derived after some matrix algebra. Simulation from the posterior

distributions can be done iteratively in two steps:

$$\begin{aligned}\omega_{t,s} \mid \tilde{z}_t^s &\sim \mathcal{PG}(1, x_t^{(2)} \beta_s), \quad t = 1 : N_s, \quad s = 1, 2, \\ \beta_s \mid \tilde{Z}, \Omega_s &\sim \mathcal{N}(m_{\beta_s}, V_{\beta_s}), \\ V_{\beta_s} &= (X^{(2)'} \Omega_s X^{(2)} + V_{\beta_0}^{-1})^{-1} \text{ and } m_{\beta_s} = V_{\beta_s} (X^{(2)'} k^s + V_{\beta_0}^{-1} m_{\beta_0}),\end{aligned}$$

where $k^s = (\tilde{z}_1^s - 1/2, \dots, \tilde{z}_{N_s}^s - 1/2)$.

We refer to Polson, Scott, and Windle (2013) for more details on the Pólya-Gamma distribution and the corresponding sampler. Additional information and a more transparent development of this data augmentation algorithm can be found in Choi and Hobert (2013).

2.3.3 Inference under model uncertainty

Here, we consider the full model comparison problem. The uncertainty about which predictors should be included in the mean regression model and in the transition probability equation is treated using a double RJMCM algorithm. In this setting, the RJMCMC does not need tuning and hence it can be used as a black box.

Suppose that a prior $\pi(k)$ is specified over k models (M_1, M_2, \dots, M_k) in a countable set \mathcal{K} and for each k we are given a prior distribution $\pi(\theta_k \mid k)$ along with a likelihood $\mathcal{L}(y \mid \theta_k, k)$ for data y . The joint prior for θ_k and k is $\pi(k, \theta_k) = \pi(\theta_k \mid k)\pi(k)$. When a move of type m from $\tilde{x} = (k, \theta_k)$ to $\tilde{x}^* = (k^*, \theta_{k^*}^*)$ is proposed from the proposal distribution g and if $j_m(\tilde{x})$ denotes the probability that move m is attempted at state \tilde{x} and $j_{m^*}(\tilde{x}^*)$ the probability of the reverse move, we accept the proposed move with probability $\alpha_m(\tilde{x}, \tilde{x}^*) = \min\{1, A_m(\tilde{x}, \tilde{x}^*)\}$, where

$$A_m(\tilde{x}, \tilde{x}^*) = \frac{\mathcal{L}(y^T \mid \tilde{x}^*) \pi(\theta_{k^*}^* \mid k^*) \pi(k^*) j_{m^*}(\tilde{x}^*) g'_m(u \mid \tilde{x}, k)}{\mathcal{L}(y^T \mid \tilde{x}) \pi(\theta_k \mid k) \pi(k) j_m(\tilde{x}) g_m(u^* \mid \tilde{x}^*, k^*)} \left| \frac{\partial(\theta_{k^*}^*, u^*)}{\partial(\theta_k, u)} \right|,$$

and $\left| \frac{\partial(\theta_{k^*}^*, u^*)}{\partial(\theta_k, u)} \right|$ is the Jacobian of the transformation.

In each step, we choose to add or remove one covariate with probability 0.5. Then, we randomly choose which covariate will be added or removed from the corresponding set of the non-included or included covariates. We propose a new value for the mean equation coefficients B^* or for the regression equation coefficients β^* from the respective full conditional posterior density, conditionally on the other coefficients. Thus, the Jacobian of the transformation is equal to unity. To be more

specific, if we want to update the covariates in the mean equation, the proposal distribution g' is just the product of the two conditional posterior distributions. With some straightforward matrix algebra, the acceptance probability for the mean equation is $\alpha_B = \min\{1, A_B\}$ and the acceptance probability for the transition matrix is $\alpha_\beta = \min\{1, A_\beta\}$ where

$$A_B = \frac{j_{m^*}(k^*)}{j_m(k)} \prod_{s=1}^2 \frac{|V_s^*|^{1/2} |V_{0s}|^{1/2}}{|V_s|^{1/2} |V_{0s}^*|^{1/2}} \\ \times \exp \left\{ -\frac{1}{2\sigma_s^2} \left(L_{0s}^{*'} V_{0s}^{*-1} L_{0s}^* - L_s^{*'} V_s^{*-1} L_s^* - L_{0s}' V_{0s}^{-1} L_{0s} + L_s' V_s^{-1} L_s \right) \right\}$$

and

$$A_\beta = \frac{j_{m^*}(k^*)}{j_m(k)} \prod_{s=1}^2 \frac{|V_{\omega s}^*|^{1/2} |V_{\beta 0s}|^{1/2}}{|V_{\omega s}|^{1/2} |V_{\beta 0s}^*|^{1/2}} \\ \times \exp \left\{ -\frac{1}{2\sigma_s^2} \left(L_{\beta 0s}^{*'} V_{\beta 0s}^{*-1} L_{\beta 0s}^* - L_{\omega s}^{*'} V_{\omega s}^{*-1} L_{\omega s}^* - L_{\beta 0s}' V_{\beta 0s}^{-1} L_{\beta 0s} + L_{\omega s}' V_{\omega s}^{-1} L_{\omega s} \right) \right\}.$$

2.3.4 The m -state NHPG

Our model and the methods described in the previous subsections can be extended into an m -state NHPG, where the rows of the transition matrix are modeled by multinomial logistic regressions. In details, let $\{Y_t\}$ be the observed process, $\{Z_t\}$ be the hidden process of an m -state NHPG and $\{X_t\}$ be the set of available predictors. In correspondence with the 2-state NHPG, conditional on the hidden process, the observed process is normally distributed using predictors $X_t^{(1)} \subseteq \{X_t\}$, i.e.

$$Y_t \mid Z_t = s \sim \mathcal{N}(X_{t-1}^{(1)} B_s, \sigma_s^2), \quad s = 1, \dots, m,$$

with $B_s = (b_{0s}, b_{1s}, \dots, b_{r_1s})$. The probabilities of the transition probability matrices $P^{(t)}$ are modeled using a multinomial (polychotomus) link with predictors $X_t^{(2)} \subseteq \{X_t\}$, i.e.

$$P(Z_{t+1} = j \mid Z_t = i) = p_{ij}^{(t)} = \frac{\exp(x_t^{(2)} \beta_{ij})}{\sum_{l=1}^m \exp(x_t^{(2)} \beta_{il})}, \quad i, j = 1, \dots, m,$$

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \dots, \beta_{r_2-1,ij})'$. For identifiability reasons, we set $\beta_{im} = 0$ for $i = 1, \dots, m$. The joint likelihood is defined as

$$\mathcal{L}(\theta, \beta) = \prod_{i=1}^m \prod_{j=1}^m \left[\prod_{t: Z_{t+1}=j} p_{ij}^{(t)} \right] \left(\frac{1}{2\pi\sigma_j^2} \right)^{N_j/2} \exp \left\{ -\frac{1}{2\sigma_j^2} (Y_j - X_j^{(1)'} B_j)' (Y_j - X_j^{(1)'} B_j) \right\}.$$

The marginal and conditional posterior distributions of the mean equation parameters $(B_s, \sigma_s^2, s = 1, \dots, m)$ are the same as in Section 2.3.2. Also, each row of the transition probability matrix corresponds to a specific multinomial (polychotomus) logistic regression. To make inference on the multinomial coefficients, we define the binary variables $\tilde{Z}_{t+1}^{ij} = I[Z_{t+1} = j, Z_t = i]$, $j = 1, \dots, m$. Following Holmes and Held, 2006; Polson, Scott, and Windle, 2013, for $i = 1, \dots, m$ and $j = 1, \dots, m - 1$, the conditional likelihood for β_{ij} is

$$\begin{aligned} \mathcal{L}(\beta_{ij} | \beta_{i(-j)}) &= \prod_{t=1}^{N_i} \{p_{ij}^{(t)}\}^{\tilde{z}_t^{ij}} = \prod_{t=1}^{N_i} \left\{ \frac{\exp(x_t^{(2)} \beta_{ij})}{\sum_{l=1}^m \exp(x_t^{(2)} \beta_{il})} \right\}^{\tilde{z}_t^{ij}} \\ &\propto \prod_{t=1}^{N_i} \frac{\{\exp(x_t^{(2)} \beta_{ij} - \log \sum_{l \neq j} \exp(x_t^{(2)} \beta_{il}))\}^{\tilde{z}_t^{ij}}}{1 + \exp(x_t^{(2)} \beta_{ij} - \log \sum_{l \neq j} \exp(x_t^{(2)} \beta_{il}))}, \end{aligned}$$

where $\beta_{i(-j)} = (\beta_{i1}, \dots, \beta_{ij-1}, \beta_{ij+1}, \dots, \beta_{im})$. Setting $k_t^{ij} = \tilde{z}_t^{ij} - 1/2$, $c_t^{ij} = \log \sum_{l \neq j} \exp(x_t^{(2)} \beta_{il})$ and augmenting the likelihood with Pólya-Gamma random variables we derive the likelihood expression,

$$\mathcal{L}(\beta_{ij} | \beta_{i(-j)}, \omega_{ij}) \propto \prod_{t=1}^{N_i} \exp\{k_t^{ij} (x_t^{(2)} \beta_{ij} - c_t^{ij})\} \int \exp\left\{-\omega_{ij}^t (x_t^{(2)} \beta_{ij} - c_t^{ij})^2 / 2\right\} p(\omega_{ij}^t) d\omega_{ij}^t.$$

Utilizing the conditionally conjugate priors on $\beta_{ij} \sim \mathcal{N}(m_{0\beta_j}, V_{0\beta_j})$ and $\omega_{ij}^t \sim \mathcal{PG}(1, 0)$, simulation of the i -th multinomial regression coefficients for the probabilities $p_{ij}^{(t)}$, $j = 1, \dots, m - 1$, is done in $m - 1$ dual steps. Particularly, let $\Omega_{ij} = \text{diag}\{\omega_{ij}^t\}$ and $K_{ij} = \text{diag}\{k_t^{ij}\}$, $t = 1, \dots, N_i$. Conditionally on Ω_{ij} and $\beta_{i(-j)}$, $j = 1, \dots, m - 1$, we use a Gibbs sampler to update the values of β_{ij} , i.e.

$$\beta_{ij} | \Omega_{ij}, \beta_{i(-j)} \sim \mathcal{N}(m_{ij}, V_{ij}),$$

with $V_{ij} = (X^{(2)'} \Omega_{ij} X^{(2)} + V_{0\beta_j})^{(-1)}$ and $m_{ij} = V_{ij} (X^{(2)'} (K_{ij} + \Omega_{ij} X^{(2)}) + V_{0\beta_j}^{-1} m_{0\beta_j})$. Then, given the updated β_{ij} , we sample the auxiliary variables from

$$\omega_{ij}^t \sim \mathcal{PG}(1, x_t^{(2)} \beta_{ij} - c_t^{ij}), \quad t = 1, \dots, N_i.$$

Regarding the variable selection, in each iteration we perform a couple of reversible jump steps, as described in Section 2.3.3. The reversible jump step for updating the covariate set that affects the mean equation parameters is analogous to the 2-state NHPG. We accept to add/remove a covariate affecting the mean equation with

probability $\alpha_B = \min \{1, A_B\}$ where

$$A_B = \frac{j_{m^*}(k^*)}{j_m(k)} \prod_{s=1}^m \frac{|V_s^*|^{1/2} |V_{0s}|^{1/2}}{|V_s|^{1/2} |V_{0s}^*|^{1/2}} \times \exp \left\{ -\frac{1}{2\sigma_s^2} \left(L_{0s}' V_{0s}^{*-1} L_{0s}^* - L_s' V_s^{*-1} L_s^* - L_{0s}' V_{0s}^{-1} L_{0s} + L_s' V_s^{-1} L_s \right) \right\}.$$

However, the reversible jump for the covariate set that affects the multinomial regressions is a bit more challenging. The acceptance probability can not be found in close form. Hence, in each iteration we calculate the acceptance probability numerically. Conditionally on Ω_s , $s = 1, \dots, m$ we simulate the candidate values β^* . Then, we compute the acceptance probability using the proposed values β and candidate values β^* . Specifically, the acceptance probability for the multinomial regression reversible jump, is $\alpha_\beta = \min \{1, A_\beta\}$ where,

$$A_\beta = \frac{\mathcal{L}(y^T | \beta^*) \pi(\beta_{k^*}^* | k^*) \pi(k^*) j_{m^*}(\beta^*) p(\beta | z, k)}{\mathcal{L}(y^T | \beta) \pi(\beta_k | k) \pi(k) j_m(\beta) p(\beta^* | z, k^*)}.$$

We calculate the likelihood numerically using the relation

$$\mathcal{L}(y^T | \beta) = \prod_{i=1}^m \mathcal{L}_i(\beta_{ij} | \beta_{i(-j)}) = \prod_{t=1}^{N_i} \{p_{ij}^{(t)}\}^{z_t^{ij}}$$

. Finally, as in Section 2.3.3 we denote with π and p the aforementioned Normal prior distributions and Normal posterior distributions, respectively.

Having described the inference and model selection for our model, we can now proceed to the description of the forecasting methodology.

2.4 Bayesian Forecasting and Scoring rules

The posterior predictive density cannot be found in closed form, but can be evaluated numerically. Given model M , the predictive distribution of y_{T+1} is

$$f_p(y_{T+1} | y^T) = \int f(y_{T+1} | y^T, z^T, M, \beta_M, \theta_M) \pi(\beta_M, \theta_M | y^T) d\beta_M d\theta_M,$$

where $f(y_{T+1} | y^T, z^T, \beta_M, \theta_M) = \sum_{s=1}^2 P(Z_{T+1} = s | Z_T = z_T) f_s(y_{T+1})$. In practice, we follow an iterative procedure within our MCMC algorithm to draw a sample from the posterior predictive distribution. At the r -th iteration of our algorithm, the algorithm chooses model M_r . Furthermore, the hidden states and the unknown

parameters $\beta_{M_r}, \theta_{M_r}$ are simulated as described in Subsection 2.3.2. To make an one-step-ahead prediction (i.e., simulate y_{T+1}), we first simulate the hidden state for time $T + 1$ from the discrete distribution based on the transition probabilities $P(Z_{T+1}^{(r)} = s \mid Z_T = z_T^{(r)})$, $s = 1, 2$, and then, conditional on the hidden state, we draw a value y_{T+1}^r from $\mathcal{N}(X_T^{(1)} B_{s,M_r}, \sigma_{s,M_r}^2)$, $s = 1, 2$. Given y_{T+l} , Z_{T+l} and the covariates X_{T+l-1} , for $l=1, \dots, L$, we may also update the transition matrix P^{T+l} , simulate Z_{T+l+1} and finally simulate the prediction y_{T+l+1} from its respective predictive distribution. In this way, at each iteration we obtain sequentially a sample of L one-step-subsequent predictions.

2.4.1 Forecasting criteria

In our model, the predictive distributions are multimodal. Hence, to evaluate the quality of the obtained forecasts or to compare with benchmark models, the selection of the right scoring rule is integral, (Gelman, Hwang, and Vehtari, 2014a). In the same manner, Geweke and Whiteman (2006) observe that the predictive accuracy is valued not only for its own sake, but it can be used as a metric to evaluate the model's performance.

Advances in numerical integration via MCMC algorithms made probabilistic forecasts possible. Besides, having the posterior predictive distribution, one can obtain point forecasts using suitable scoring functions (Gneiting, 2011). Scoring rules provide summary measures for the evaluation of probabilistic forecasts by assigning a numerical score based on the forecast and on the event or value that it materializes. We refer to Gneiting and Raftery (2007) and Machete (2013) for reviews on the theory and properties of scoring rules. A widely used, extensively studied and quite powerful criterion is the Logarithmic Score (LS), see Gelman, Hwang, and Vehtari (2014a) and Gschlößl and Czado (2007) and references therein. It is based on the logarithm of the posterior predictive density evaluated at the observed value. However, LS lacks robustness as it involves harsh penalty for low probability events and thus is sensitive to extreme cases (Boero, Smith, and Wallis, 2011). Besides, comparing the entropies of the forecasts, Machete (2013) showed that LS prefers the forecast density that is less informative. In the same spirit Gneiting and Raftery (2007) noticed that measures which are not sensitive to distance give no credit for assigning high probabilities to values near but not identical to the one materializing. Sensitivity to distance seems desirable when predictive distributions tend to be multimodal, which is the case in our setting. To deal with this issue, one could calculate the Continuous Ranked Probability Score (CRPS) which is based on the cumulative predictive

distribution, see Appendix 2.8.4 for the definition. Boero, Smith, and Wallis (2011) argued that when density forecasts are collected in histogram format, then the ranked probability score has advantages over the other studied scoring rules.

To compute the CRPS for the forecast y_l , $l = T + 1, \dots, T + L$ we use the identity of Székely and Rizzo (2005a),

$$CRPS(F_p, y_l) = \frac{1}{2} E_F |Y - Y'| - E_F |Y - y_l|,$$

where Y, Y' are independent copies of a random variable with the posterior predictive distribution function F_p (see also Gschlößl and Czado, 2007).

Finally, along with the CRPS, we also use two standard point forecasting criteria: the Mean Squared Forecast Error, $MSFE_i = \frac{1}{L} \sum_{l=T+1}^{T+L} (y_l - \hat{y}_l)^2$ and the Mean Absolute Forecast Error, $MAFE_i = \frac{1}{L} \sum_{l=T+1}^{T+L} |y_l - \hat{y}_l|$, where \hat{y}_l is the estimation of the forecast y_l at the i -th MCMC iteration. The values for CRPS, MSFE and MAFE are computed in every iteration of the MCMC algorithm. In the end, we compute and report the average CRPS, MSFE and MAFE over all MCMC iterations.

2.5 Simulation Study

We have conducted a series of simulation experiments to assess the performance of the proposed approach in terms of inference, model selection and predictive ability. We have scrutinized our algorithms, using different sample sizes and assigning various values to the parameters. Our experiments have been carried out using MATLAB 2018b on a Windows 10 system with 32GB of RAM and Intel Core i7 8-core processor.

To assess its inferential ability we benchmarked our model with the M&D model (without model uncertainty) and with a Homogeneous Hidden Markov Model (HHMM), see Appendix 2.8.3. The NHPG is at least as good as the M&D model – in forecasting ability, and sample quality – but is faster and more efficient (as reported in Table 2.6, Effective Sample Rate). In Table 2.6, we present a summary of the case study of the fixed model.

Our model shines when there is model uncertainty, Section 2.5.1. We compare NHPG with existing variable selection schemes, i.e., the M&D, a HHMM with RJ step and a model using the spike and slab prior for variable selection as studied in Narisetty and He (2014) and referred to as BAyesian Shrinking And Diffusing priors (BASAD), see Appendix 2.8.2..

The data were generated either from a HHMM or from a NHHMM with covariates simulated from independent normal distributions. We found that the mean equation coefficients converged rapidly, whereas the logistic regression coefficients converged only after a short burn in period. The hidden chain Z^T was well estimated. For each iteration, we kept a replication of the hidden chain and compared it with the real simulated hidden chain, using a 1-0 loss function (see Figure 2.2). For instance, in the presented simulation study of Section 2.5.1, the NHPG model failed to correctly identify only approximately 2 states out of the 1104 hidden states per MCMC iteration and in the simulation study of Appendix 2.8.3 the NHPG model failed to correctly identify < 1 states per MCMC iteration.

Furthermore, to test the predictive ability of our model, we used L out-of-sample observations. We calculated, for all the competing models, the CRPS, the MSFE and the MAFE. However, we note that due to the large out-of-sample period, we only report the averages (for all the draws) of the aforementioned forecasting criteria. In all the experiments, we found that our model outperforms all competing models in forecasting the observed process.

2.5.1 Case study: The NHPG with model uncertainty

The main applications in which our algorithm considerably improves over the benchmark models – M&D, BASAD – is when there exists model uncertainty. We simulated data from a NHHMM of size $T = 1200$. From a common pool of independently normally distributed covariates $X = \{1, X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9\}$ with means $\mu_x = [4, 3, -2, -5, 2.5, -4, -6, 7, 1]$ and variances $\sigma_x^2 = [1, 1, 0.5, 1, 1, 1, 0.5, 2, 1.5]$, we used three covariates $X^{(1)} = \{1, X_1, X_2, X_3\}$ affecting the mean equation and another three covariates $X^{(2)} = \{1, X_1, X_2, X_4\}$ the transition matrix. The mean equation parameters were $B_1 = [2, -0.3, 2, 2]'$, $\sigma_1^2 = 1.5$ and $B_2 = [1, 3, 4, 3]'$, $\sigma_2^2 = 0.8$ whereas the logistic regression coefficients where $\beta_1 = [1.5, 1, 2, 3]'$ and $\beta_2 = [3, -2.5, 4, 1]'$, for the two states respectively.

Our results are based on a sample of 15000 predictions after discarding an initial burn in period of 10000 iterations. We kept $L = 96$ out-of-sample observations and we computed a sequence of one-step-ahead forecasts of the real observed process. In this forecasting analysis, we also included the HHMM with variable selection, in the mean equation. We used non-informative priors for the unknown parameters σ_s^2, B_s, β_s , $s = 1, 2$, that is $\sigma_s^2 \sim \mathcal{IG}(0.1, 0.1)$, $B_s \mid \sigma_s^2 \sim \mathcal{N}(0, 100\sigma_s^2 \times I)$ and finally $\beta_s \sim \mathcal{N}(0, 100 \times I)$. Also, as suggested by Narisetty and He (2014) and Narisetty, Shen, and He (2018), we used as hyperparameter values $\tau_{0B,n}^2 = \frac{\hat{\sigma}^2}{10T}$, $\tau_{1B,n}^2 =$

$\hat{\sigma}^2 \max\left(\frac{r_1^{2,1}}{100T}, \log(T)\right)$ and $\tau_{0\beta,n}^2 = \frac{1}{T}$, $\tau_{1\beta,n}^2 = \max\left(\frac{r_2^{2,1}}{100T}, 1\right)$, where $\hat{\sigma}^2$ is the estimated variance of the data Y .

Our approach was able to identify – as the most probable or the median probability model – the correct data generating process. This was in contrast to the competing methodologies, as can be seen in Table 2.1. Results from further simulation studies (not reported here), imply that the performance of our method to identify the true data generating process remains robust in the choice of parameters. In terms of comparison, the competing algorithms could perform at most equally well.

Median probability model				
	True Model	NHPG	M&D	BASAD
ME	X_1, X_2, X_3	X_1, X_2, X_3	X_1, X_2, X_3	X_2, X_4, X_9
TM	X_1, X_2, X_4	X_1, X_2, X_4	$X_1, X_2, \dots, X_7, X_9$	X_1, X_2, X_4

TABLE 2.1: Median probability models using the proposed methodology (NHPG), the methodology proposed by Meligkotsidou and Dellaportas (2011) (M&D) and the model of Narisetty and He (2014) (BASAD), respectively. The first row (ME) shows the covariates used in the Mean Equation and the second row (TM) the covariates of the Transition Matrix. The proposed methodology is the only one able to identify the true data generating process.

In Table 2.2, we report the forecasting criteria scores. The NHPG had the best performance according to all forecasting criteria. Supplementary to Table 2.2 are the plots in Figure 2.1. This figure shows the approximation of the empirical posterior predictive distributions (based on a normal kernel) of the four competing models, for three randomly selected out-of-sample periods and the actual observed values in the same graph. Figure 2.2 gives a graphical indication of the improved forecasting performance of NHPG.

Forecasting Criteria				
	NHPG	M&D	HHMM	BASAD
CRPS	-1.9526	-3.6829	-2.6597	-2.4952
MAFE	3.9271	4.3911	5.4101	5.0611
MSFE	32.8856	39.4958	53.8280	49.1432

TABLE 2.2: Forecasting performance of the competing models. In addition to M&D and the BASAD, we include also the HHMM. The best performance (bold values) for each criterion is achieved by the proposed NHPG model.

Finally, for each MCMC iteration we kept a replicated chain of the hidden process and we compared it with the true simulated chain. Using the 0-1 Loss function, we computed the average number of misestimated states in each chain. All three approaches had similar performance according to this criterion. Specifically, from the chain with 1104 hidden states, NHPG failed to recognize 2 states per iteration, M&D 3 states per iteration and BASAD methodology 1 state per iteration. Concerning the misclassifications, we observed that the algorithms classified as state 1 the state 2, i.e. the state with the larger variance, in all instances. A virtualization of the estimation of the hidden process against the true hidden process is shown in Figure 2.2. This figure presents the thinned version (1:2 observations) of the simulated time series along with the true hidden process and an estimate of the hidden process using the proposed methodology.

	NHPG	M&D	BASAD
Mean runtimes per 1000 iterations (seconds)	310	693	160

TABLE 2.3: Summary of runtimes (in seconds per 1000 iterations).

In Table 2.3, we report the runtimes for every methodology. The trade-off for the better forecasts of NHPG is 150 seconds per 1000 iterations in comparison to the BASAD. However, the NHPG is more than two times faster than the M&D.

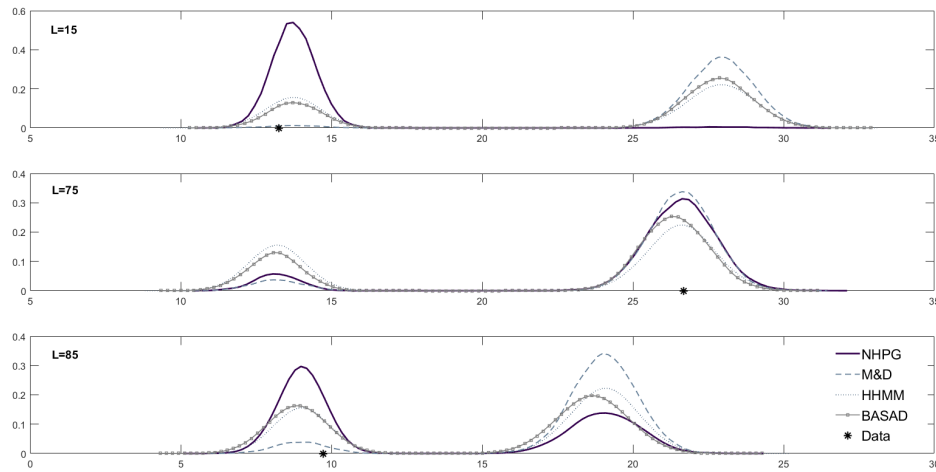


FIGURE 2.1: Plots of the empirical posterior predictive distributions based on a normal kernel function for three randomly selected out-of-sample forecasts, $L = 15, 75, 85$, using the NHPG (black continuous line), M&D (gray dashed line), the HHMM (gray dotted line) and the BASAD (gray squared line). Actual out-of-sample values are marked with asterisks. These plots visualize the advantage of NHPG: global maximums of the multimodal distributions is achieved close to the actual values.

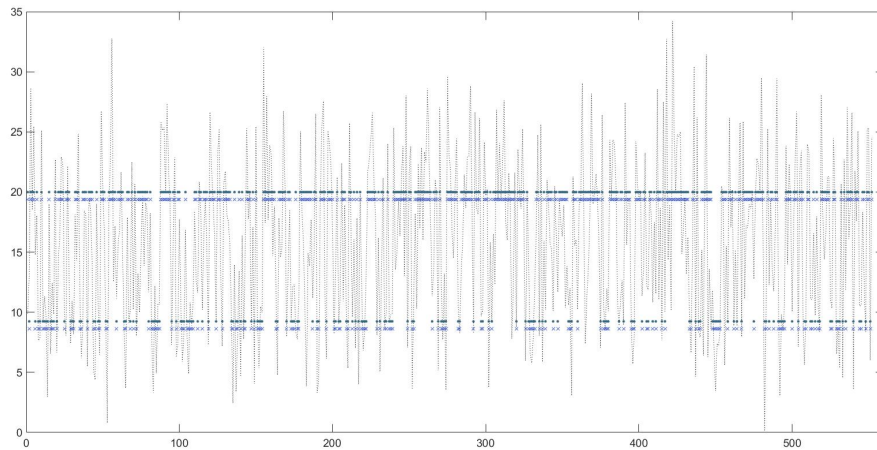


FIGURE 2.2: Observed process (black dotted line) and hidden process: the true hidden states are marked with blue x and the simulated states are marked with black dots. The true hidden process is well estimated.

2.6 Empirical Application: Realized volatility data

We use the NHPG to assess the predictive ability of 13 financial variables in forecasting future volatility. Financial volatility has been extensively studied in the literature due to its crucial role in various financial fields, such as asset pricing, risk management, investment and asset allocation among others, see Gospodinov, Gavala, and Jiang (2006). Several studies have considered predicting realized stock volatility

using various financial and/or economic predictors (see for example, Mittnik, Robinson, and Spindler (2015), Meligkotsidou et al. (2019), Christiansen, Schmeling, and Schrimpf (2012), and Paye (2012)).

2.6.1 The data

We used the realized stock market volatility data and more precisely the “long” sample of the U.S. equity market, S&P500, as described in Christiansen, Schmeling, and Schrimpf (2012). The realized volatility is the squared root of the realized variance for asset class i in month t expressed as the sum of squared intra-period (daily) returns

$$RV_{i,t} = \sqrt{\sum_{\tau=1}^{u_t} r_{i,t,\tau}^2}, \quad t = 1 \dots, T,$$

where $r_{i,t,\tau}$ is the r -th daily continuously compounded return of month t for asset i with u_t the trading days. Thus $\sum_{\tau=1}^{u_t} r_{i,t,\tau}^2$ is the realized variance for asset class i in month t . The distribution of the realized daily variances are highly non-normal and skewed to the right, but the logarithms of the realized variances are approximately normal and thus, they have better behavior (Andersen et al., 2003). Hence, in the following analysis, we study the natural logarithm of the realized volatility series, $\ln(RV_{i,t})$ $t = 1 \dots, T$.

The data are observed in a monthly basis, from December 1926 to December 2015. We used a five-years extended dataset compared to the dataset of Christiansen, Schmeling, and Schrimpf, 2012. The out-of-sample forecast evaluation period was set to eight years, i.e., 96 observations from December 2007 until December 2015. We kept 40000 iterations as a burn-in period and we generated an output of 40000 MCMC draws. We used non-informative priors for the unknown parameters $\sigma_s^2, B_s, \beta_s, s = 1, \dots, m$, that is $\sigma_s^2 \sim \mathcal{IG}(0.15, 0.15)$, $B_s \mid \sigma_s^2 \sim \mathcal{N}(0, 100\sigma_s^2 \times I)$ and finally $\beta_s \sim \mathcal{N}(0, 100 \times I)$. At first, we modeled this dataset as an m -state NHPG model with $m = 2, 3, 4$, to explore the cardinality of the state-space of the hidden Markov chain. Examining the output of the hidden process, we observed that there were only a few or none time periods corresponding to the third or the fourth state. Based on these results we conclude that the m -state, $m = 3, 4$, NHPG models, are not consistent with this dataset. Hence, from this point forward, we only report the results of the 2-state NHPG model.

Following Christiansen, Schmeling, and Schrimpf (2012) and Meligkotsidou et

al. (2019), we took into account 13 macroeconomic and financial standardized predictive covariates. Particularly, from a list of equity market variables and risk factors, we considered the dividend price ratio (DP) and the earnings price ratio (EP) (Welch and Goyal, 2008). To capture the leverage effect, i.e. the asymmetric response of volatility to positive and negative returns (Nelson, 1991) we included the lagged equity market returns (MKT). We also used the risk factors of Fama and French (1993), that is, the size factor (SMB), value factor (HML) and a short-term reversal factor (STR). From the set of interest rates, spreads and bond market factors, we included the treasury bill rate (TBL), i.e., the interest rate on a three-month Treasury bill, the long-term return (LTR) on long-term government bonds, the term spread (TMS), i.e., the difference between the long-term yield and treasury bill rate, the relative T-bill rate (RTB) as the difference between T-bill rate and its 12-month moving average and the relative bond rate (RBR), as the difference between LTR and its 12 month moving average (Welch and Goyal, 2008). To proxy for weighted credit risk, we also used the default spread (DEF) defined as the yield spread between BAA and AAA rated bonds. Lastly, we considered the macroeconomic variable inflation rate (INF), which is the monthly growth rate of CPI.

The strong contemporaneous relation between the volatility and the business conditions implies that lagged volatility plays an important role in forecasting (see Paye (2012), Baillie et al. (2019), and Andersen et al. (2003)). Besides, quoting Christiansen, Schmeling, and Schrimpf (2012), we include at least one autoregressive term, “since volatility is fairly persistent, it is important to include autoregressive terms in the predictive regression to investigate whether there is additional predictive content of the macroeconomic and financial variables that goes beyond the information contained in lagged volatility”. We ran a series of experiments for this data. Specifically, we performed our analysis using the predictors described and then we repeated the analysis adding to the common –for the mean equation and the logistic regression equations – predictor set, the autoregressive terms (AR) of lag 1, 2 and 3.

2.6.2 Results

Based on the posterior probabilities of inclusion, we see that if we do not include any Auto-Regressive (AR) terms in the predictors’ pool, then the median NHPG₀ model has three predictors affecting only the mean equation of the series (Table 2.4). Thus, based on the median probability model, the realized volatility series is considered to be a HHMM. The probabilities of staying at the same state are in this case high, concluding that the states are highly persistent. When we add the AR(1) term, the

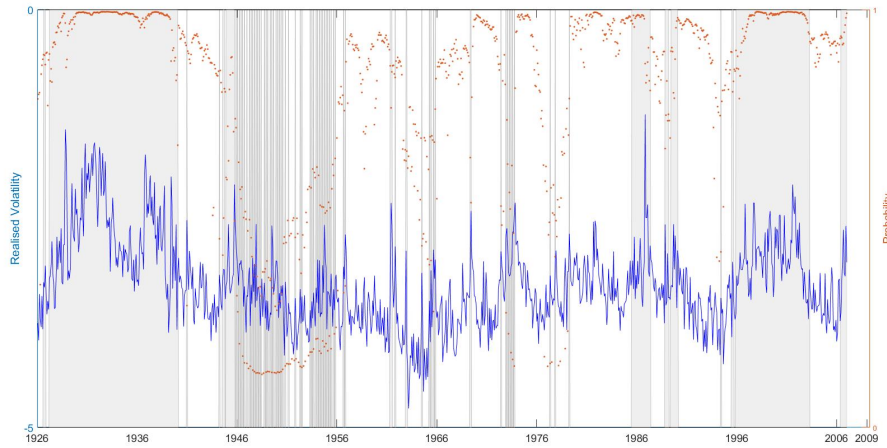


FIGURE 2.3: Time series (blue line) of the monthly realized volatility of the Standard & Poor (S&P) 500 index (in logarithmic scale, left axis) for the period 1926-2007, using the NHPG₁. Gray-shaded bars show a realization of the times with hidden state 1. The NHPG₁ exploits the heteroscedasticity of the series. Red dots are the posterior mean probabilities (right axis) of staying at the same state and indicate a persistent unobserved process.

included predictors in the median probability model (NHPG₁) are also three but they affect the series both linearly and not linearly. We observe that an autoregressive term explains a big fraction of the variance of the realized volatility, even though the AR(1) term (based on the median probability model) is affecting the series only linearly, i.e. in the mean equation. Adding more AR terms (of lag 2 and lag 3), the median probability model remained almost the same as in the case of the model with one AR term and hence, we only report the NHPG₁ model. Furthermore, in our out-of-sample analysis, we did not encounter any significant improvement in the forecasting ability of the models with AR(2) and AR(3) terms. We note that this result confirms the findings of the model of Christiansen, Schmeling, and Schrimpf (2012), hereafter CSS, who also used only an AR(1) term in their analysis.

Even though – based on the CRPS – the model with the best performance was the one with the autoregressive term (NHPG₁), we present the results of both the model with no AR terms (NHPG₀) and the NHPG₁ model, for the sake of completeness. Also, we compare our results with those of the CSS model – which is a linear model with one autoregressive term and a Markov Chain Monte Carlo model compositions algorithm (MC³) with a Bayesian Model Averaging (BMA) approach. For the CSS model, we allowed for a much longer burn in period, as suggested by the authors, of 500000 draws. Moreover, we included in our comparative analysis the M&D model and HHMM with one AR term.

Figure 2.3 shows a plot of the realized volatility data (blue line) together with

the probabilities of staying at the same state (e.g. if at time t we are at state 1 then the red dot at time t shows the probability of staying at state 1 at time $t + 1$, that is the transition probability p_{11}^t). The high probabilities of staying at the same state indicate that the unobserved process is persistent. The shaded bars represent the time period that the chain was in state 1, based on one realization of the hidden process. Additionally, by visual inspection of this plot, we observe that the realized volatility series is fragmented into two subchains according to the hidden states: hidden state 1 defines periods with high variability of the studied series and hidden state 2 defines periods with low variability. Furthermore, in Figure 2.4 we present a thinned (1:5) in-sample realization of the observed process inferred by our algorithm, i.e., using the in-sample estimations of the parameters and the states to reproduce the realized volatility series, along with the real data. The in-sample evaluation of the observed process gives an indication of the good performance of the estimation procedure.

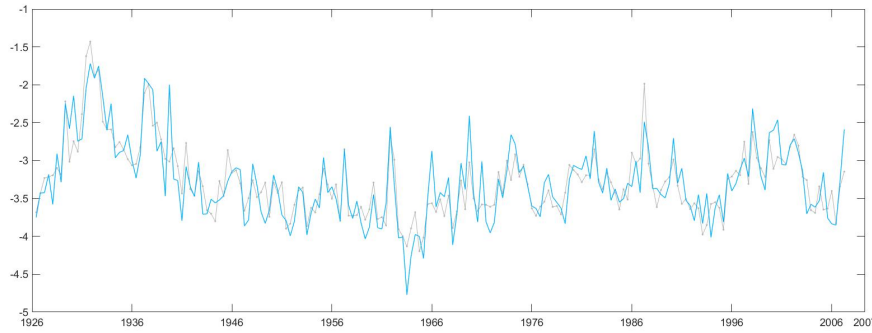


FIGURE 2.4: Thinned (1:5 observations) realized volatility time series (blue line) versus the observed process as calculated by the proposed NHPG₁ (gray solid line).

2.6.2.1 Model Selection

Our model selection algorithm did not assign high probability to any specific model indicating that there exists model uncertainty. In Table 2.4, we summarize the posterior probabilities of inclusion for each predictor, both for the mean equation and for the transition matrix for the NHPG₀, NHPG₁, M&D and only for the mean equation for CSS and HHMM. Our methodology – when the autoregressive term of lag 1 was included – was not only able to identify which covariates affect the realized volatility series but also to decide how the covariates affect the series, i.e., linearly or non-linearly. The number of the predictors defining the median probability NHPG₁ has diminished to three (instead of thirteen). Specifically, we found that the MKT affects the series linearly, the SMB affects the series non-linearly and the DEF both linearly and non linearly. The predictors that were included in NHPG₁ are in common

with the predictors included in the HHMM. The CSS model identifies four predictors with probability at least 0.5, three of them being the same with NHPG₁ and HHMM, that is the MKT, the DEF and the EP plus the predictor STR. However, the M&D algorithm includes all the predictors in the mean equation model, while it includes the predictors DP, MKT, TBL and DEF in the logistic regression for the transition probabilities.

Covariates	Posterior probabilities of inclusion							
	NHPG ₀		NHPG ₁		M&D		HHMM	CSS
	ME	TM	ME	TM	ME	TM	ME	ME
DP	0.01	0.04	0.08	0	1	0.77	0.3	0.38
EP	0.98	0.12	0.06	0.89	1	0.37	0.78	0.50
MKT	1	0.04	0.98	0.03	0.93	1	1	0.97
SMB	0	0.02	0	0.04	0.90	0.28	0	0.05
HML	0.01	0.03	0	0.02	0.92	0.09	0	0.06
STR	0	0.05	0	0.06	0.92	0.09	0	0.53
TBL	0	0.02	0	0.10	0.99	0.71	0.03	0.10
RTB	0	0.02	0	0.09	0.99	0.02	0	0.04
LTR	0.01	0.03	0	0.03	1	0.01	0	0.05
RBR	0	0.03	0	0.03	1	0.01	0	0.05
TMS	0.01	0.06	0.17	0.26	0.89	0.23	0	0.05
DEF	1	0.02	1	0.79	1	1	1	1
INF	0	0.02	0.11	0.03	0.92	0.47	0	0.04

TABLE 2.4: Posterior probabilities of inclusion for the competing models. Predictors with inclusion probability above 0.5 (median probability model) are marked with bold values. NHPG₀ and NHPG₁ denote the proposed methodology without autoregressive terms and with one autoregressive term respectively, M&D the methodology proposed by Meligkotsidou and Dellaportas (2011), HHMM the Homogeneous model with variable selection using a RJ-step and CSS the model of Christiansen, Schmeling, and Schrimpf (2012). ME stands for Mean Equation (linear relationship) and TM for Transition Matrix (non linear relationship). The HHM and CSS models included covariates only in the ME.

2.6.2.2 Forecasting

The values of the forecasting criteria that we used for all competing models are reported in Table 2.5. We conclude that NHPG₁ performs better than all the other models, since it has the best scores in all forecasting criteria: the mean Continuous Ranked Probability Score (CPRS), MAFE and MSFE.

Forecasting Criteria					
	NHPG ₁	NHPG ₀	M&D	HHMM	CSS
CRPS	-0.1971	-0.2175	-0.2191	-0.2118	-0.2238
MAFE	0.3821	0.4643	0.4172	0.4534	0.4787
MSFE	0.2467	0.3426	0.2678	0.3449	0.3813

TABLE 2.5: Summary of forecasting results of the five competing models, obtained from the log-realized volatility dataset. The best performance (bold values) for each criterion is achieved by the NHPG₁ model (with one autoregressive term).

2.7 Conclusions and Discussion

In this paper, we considered inference on predictive Non - Homogeneous Hidden Markov Models with Pólya-Gamma data augmentation. Given a common pool of predictors, we allowed for different sets of covariates to affect the mean equation and the time-varying transition probabilities. To determine which covariates affect the series linearly and/or non-linearly, we performed stochastic variable selection using a couple of reversible jump steps. Additionally, we modeled the probabilities of the transition probability matrix via a logistic/multinomial link function. Bayesian inference for the logistic regression model has been recognized as a hard problem – many of the proposed methodologies face efficiency and convergence issues – due to the analytically inconvenient form of the model’s likelihood function. To account for these issues, which are amplified in the more complex setting of NHHMMs, we developed an accurate MCMC algorithm based on the recently proposed Pólya-Gamma data augmentation scheme of Polson, Scott, and Windle (2013).

In each MCMC iteration, we simulated the hidden states using the scaled Forward-Backward algorithm of Scott (2002), the mean equation parameters using a Gibbs step, and the logistic regression coefficients using the Pólya-Gamma augmentation scheme. Finally, we performed a couple of reversible jump steps to choose the covariates that affect the mean equation and the transition probabilities. Using the most probable model, the median probability model or Bayesian Model Averaging, we make one-step-ahead predictions, within the Bayesian framework.

To assess the performance of the proposed algorithm and the predictive ability of our model, we conducted an extensive number of simulation experiments. The results showed that our algorithm mixes and converges well and provides accurate

estimates of the model's parameters. Moreover, they exhibited that our model outperforms benchmark models, such as the approach of Meligkotsidou and Dellaportas (2011), the BASAD model of Narisetty and He (2014) and the Homogeneous hidden Markov model, in terms of both variable selection and forecasting ability according to the continuous ranked probability score, the mean absolute forecasting error and the mean squared forecasting error. The currently proposed methodology was applied to a realized volatility dataset – detailed in Christiansen, Schmeling, and Schrimpf (2012) – for predicting future observations and for predictor selection. The median probability model identified three predictors, one affecting the analyzed series linearly, one non-linearly and one both linearly and non-linearly. Using the proposed methodology we obtained improved forecasts, compared to Christiansen, Schmeling, and Schrimpf (2012).

The findings of the present study indicate that complex NHHMMs are promising for predicting univariate financial and economic time series. More accurate forecasts can be derived without the need of tuning (black box functionality) and at a low trade-off in terms of computational complexity.

The efficiency of the proposed model can be further improved by refining the model selection process via a more elaborated or simpler variable selection schemes or by simultaneously selecting the predictors affecting the mean equation and the transition probabilities equation in a block-based algorithm. Furthermore, more informative priors, such as g-priors, various spike and slab or horseshoe priors, can be exploited. It is still an open question of whether the aforementioned priors can more effectively address the issue of collinearity. Moreover, using standard methods, it can be extended to the prediction of multivariate time series that arise in many economic and non-economic applications. In this way, the proposed methodology may be of interest not only to the econometric but also to the broader forecasting community.

2.8 Appendix

2.8.1 The Pólya-Gamma distribution

The definition and more details on the properties of the new family of Pólya-Gamma distribution can be found in Polson, Scott, and Windle (2013). In this section we only give the definition of the aforementioned family of distributions.

A random variable X has a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$, $X \sim \mathcal{PG}(b, c)$ if

$$X \stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2 / (4\pi^2)}$$

where the $g_k \sim \mathcal{Ga}(b, 1)$ are independent gamma random variables and $\stackrel{D}{=}$ indicates equality in distribution.

2.8.2 Benchmark models

We give the definitions of the M&D model, BASAD model, (Narisetty and He, 2014), and the standard HHMM that we use in Section 2.5.1.

The model M&D of Meligkotsidou and Dellaportas (2011) is essentially a NHHMM with the auxiliary variable representation for the transition probabilities of Holmes and Held (2006).

$$P(\tilde{Z}_t^s = 1) = P(u_t > 0),$$

$$u_t = x_t \beta_{ss} + e_t$$

$$e_t \sim \mathcal{N}(0, \lambda_t),$$

$$\lambda_t = (2\psi_t)^2,$$

$$\psi_t \sim \mathcal{KS},$$

where \mathcal{KS} denotes the Kolmogorov-Smirnov distribution. The variable selection scheme is close to the variable selection scheme of this work and it is based on the joint updates of β_{ss} and covariate set. A detailed description of this scheme can be found in the lines of Meligkotsidou and Dellaportas (2011) and Holmes and Held (2006)

In the BASAD model (Narisetty and He, 2014), the authors introduce shrinking and diffusing priors as a spike and slab priors model, with prior parameters depending on the sample size to achieve appropriate shrinkage. They work with orthogonal design matrices and use binary latent variables U_i to indicate if a covariate is active or not. In our setting, the BASAD model for the mean equation is defined as:

$$Y_t \mid (X_{t-1} B_s, \sigma_s^2) \sim \mathcal{N}(X_{t-1} B_s, \sigma_s^2), \quad s = 1, 2, \quad t = 1, \dots, T,$$

$$B_{k,s} \mid (\sigma_s^2, U_{k,s} = 0) \sim \mathcal{N}(0, \sigma_s^2 \tau_{0B,n}^2)$$

$$B_{k,s} \mid (\sigma_s^2, U_{k,s} = 1) \sim \mathcal{N}(0, \sigma_s^2 \tau_{1B,n}^2), \quad k = 1, \dots, r,$$

$$P(U_{k,s} = 1) = 1 - P(U_{k,s} = 0) = q_n, \quad k = 1, \dots, r,$$

and

$$\sigma_s^2 \sim \mathcal{IG}(\alpha_1, \alpha_2).$$

The transition probabilities are parametrized as:

$$\tilde{Z}_{t+1}^s \sim \text{Bin}\left(1, \frac{\exp(x_t \beta_s)}{1 + \exp(x_t \beta_s)}\right),$$

$$\beta_k^s \mid (U_{k,s} = 0) \sim \mathcal{N}(0, \sigma_s^2 \tau_{0\beta,n}^2)$$

$$\beta_i^s \mid (U_{k,s} = 1) \sim \mathcal{N}(0, \sigma_s^2 \tau_{1\beta,n}^2), \quad k = 1, \dots, r,$$

$$\omega_s \sim \mathcal{PG}(b_\omega, 0)$$

and

$$P(U_{k,s} = 1) = 1 - P(U_{k,s} = 0) = q_n, \quad k = 1, \dots, r.$$

In contrast, in the HHMM, covariates affect only the mean equation and the transition probability matrix is constant,

$$Y_t \mid Z_t = s \sim \mathcal{N}(X_{t-1}^{(1)} B_s, \sigma_s^2), \quad s = 1, 2, \quad t = 1, \dots, T,$$

$$P(z_t = j \mid z_{t-1} = i) = p_{ij}, \quad i, j = 1, 2 \quad \forall t = 1, \dots, T.$$

2.8.3 Case study for the 2-state fixed model

We present the results of a case study without model uncertainty. This study shows empirically that our algorithm converges, mixes well and is effective. In this case, the results are marginally better than the M&D model and the HHMM. However, together with the results of the case with model uncertainty, they demonstrate that the proposed algorithm provides an overall improvement over M&D.

We simulated data from a NHHMM of size $T = 1500$. We used three covariates $X^{(1)} = \{1, X_1, X_2, X_3\}$ affecting the mean equation and three covariates $X^{(2)} = \{1, X_1, X_2, X_4\}$ affecting the transition matrix, with X being a set of independently normally distributed covariates with means $\mu_x = [4, 3, -2, -5]$ and variances $\sigma_x^2 = [1, 1, 0.5, 1]$. The mean equation parameters were $B_1 = [2, -0.3, 2, 2]'$, $\sigma_1^2 = 1.5$ and $B_2 = [1, 3, 4, 3]'$, $\sigma_2^2 = 0.8$ whereas the logistic regression coefficients where

$\beta_1 = [1.5, 1, 2, 3]'$ and $\beta_2 = [3, -2.5, 4, 1]'$ for the two states, respectively. We kept $L = 100$ out-of-sample observations and we computed a sequence of one-step-ahead forecasts of the real observed process. We used non-informative priors for the unknown parameters σ_s, B_s, β_s , $s = 1, 2$, that is $\sigma_s^2 \sim \mathcal{IG}(0.1, 0.1)$, $B_s | \sigma_s^2 \sim \mathcal{N}(0, 100\sigma_s^2 \times I)$ and finally $\beta_s \sim \mathcal{N}(0, 100 \times I)$.

		NHPG	M&D	HHMM
Forecasting criteria	CRPS	-2.0794	-2.1920	-4.4323
	MAFE	4.2661	4.3907	8.6745
	MSFE	60.9978	63.1617	143.1512
Sample Quality	ESS	11936	11934	12153
	MESS	24391	24343	24578
Efficiency	mESR	3.2130	0.0120	110
Convergence & Mixing	PSRF	1	1	1.0005
	mCM	18(18)	18(18)	8(10)

TABLE 2.6: Summary of results: CRPS is the mean continuous rank probability score, MAFE is the mean absolute forecast error and MSFE is the mean squared forecast error. ESS is the minimum effective size of among the ESS for all parameters and MESS the multivariate effective size, for an MCMC run of 25000 iterations. mESR is the minimum effective sample rate. PSRF is the maximum potential scale reduction factor and mCM is the multivariate convergence and mixing diagnostic. In the mCM line we report the number of the components of the parameters out of the total components – in parenthesis – that fall into the 95% confidence interval of the test. NHPG is the proposed model, M&D is the model of Meligkotsidou and Dellaportas (2011), HHMM is the homogeneous model. Bold values denote the best values for the corresponding criterion among all the competing models.

Inferences are based on an MCMC sample of 25000 iterations after a burn-in period of 10000 iterations. A summary of the results of this experiment is reported in Table 2.6. We used several metrics for assessing the efficiency of our algorithm (see Appendix 2.8.5). The quality of the sample is measured with the Effective Sample Size (ESS), Multivariate Effective Sample Size (MESS). We also use the minimum Effective Sample Rate (mESR) as a measure of the efficiency of the algorithm. To assess the convergence and mixing of the algorithm we use the Potential Scale Reduction factor (PSRF) of Brooks and Gelman (1998) and Gelman et al. (2013) and the multivariate mixing diagnostic of Paye (2012). Specifically, we show that our algorithm converges to the stationary distribution and has good mixing properties,

using the aforementioned diagnostic criteria. Also, the univariate and multivariate effective sample sizes for all the methodologies/models are large, implying an efficient algorithm. The Effective Sample Rate of NHPG is 3.213 whilst M&D's ESR is significantly lower with a score of 0.012. In addition, the NHPG has best forecasting performance, since it has the best score among the benchmarks' scores in all forecasting criteria. In Figure 2.5 we visualize the empirical continuous approximation of the posterior predictive densities of NHPG, M&D, HHMM, for the three randomly selected out-of-sample periods, $L = 15, 85, 100$. These plots provide additional evidence that the NHPG gives at least good predictions as the M&D model.

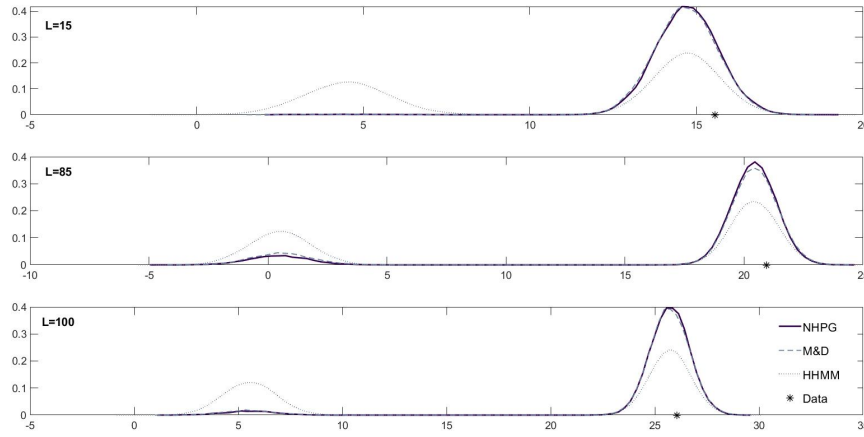


FIGURE 2.5: Conjointly plotted empirical continuous approximations (based on a normal kernel function) of the posterior predictive distribution for out-of-sample periods $L = 15, 85, 100$, using the NHPG (black continuous line), M&D (gray dashed line) and the HHMM (gray dotted line). Actual out-of-sample values are marked with asterisks.

2.8.4 The Continuous Rank Probability Score

Let y_l be the real observed values of the forecasts, $y^L = (y_1, \dots, y_L)$ the history of the predictive quantity and \hat{y}_l the estimated forecasts. Using the notation f for the distribution of the true model, $f_p(\hat{y}_l)$ for the posterior predictive density of the new data, and $F_p(x) = \int_{-\infty}^x f_p(\hat{y} | y^L) d\hat{y}$ for the posterior predictive cumulative density function. The CRPS for y_i is defined as,

$$CRPS(F_{p,l}, y_l) = - \int_{-\infty}^{\infty} \left(F_{p,l}(\hat{y}_l) - I(\hat{y}_l \geq y_l) \right)^2 d\hat{y}_l = - \int_{-\infty}^{\infty} \left(F_{p,l}(\hat{y}_l) - F_{y_l}(\hat{y}_l) \right)^2 d\hat{y}_l,$$

where $I(x \geq y)$ denotes a step function along the real line that attains the value 1 if $x \geq y$ and the value 0 otherwise, $F_{y_l} = H(\hat{y}_l - y_l)$ is the cumulative distribution of the real value y_l and H is the Heaviside function (Hersbach, 2000), $H(x) = 0$, if $x \leq 0$ and 1 otherwise.

2.8.5 Metrics of Comparison

We briefly present the convergence diagnostics, mixing criteria and metrics of effectiveness that we used to measure the performance of our algorithm.

As a primary metric of comparison, following Holmes and Held (2006) and Polson, Scott, and Windle (2013), we calculated the effective sample size (ESS). For each dimension of the parameter vector, ESS_i is the number of independent samples needed to obtain a parameter estimate with the same standard error as the MCMC estimate based on M dependent samples (Neal, 1993; Kass et al., 1998). If θ is the p -dimensional parameter of interest, and $\theta_n = 1/n \sum_{t=1}^n g(x_t)$ is an estimate of θ based on a Markov chain $\{X_t\}$, with $\theta_n \rightarrow \theta$ the Monte Carlo error, $\theta_n - \theta$ is described asymptotically by the Central Limit Theorem (CLT), $\sqrt{n}(\theta_n - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \Sigma_p)$. The idea of the ESS lies on the univariate CLT for each component of θ and it is defined, for $i = 1, \dots, p$, as

$$ESS_i = \frac{M}{1 + 2 \sum_{j=1}^k \rho(j)} = M \frac{\lambda_i}{\sigma_i},$$

where $\rho(k)$ is the sample autocorrelation of lag k of the parameter θ_i , λ_i the diagonal element of the sample covariance matrix Λ , σ_i the diagonal element of Σ_p and M the number of post-burn in samples. We also report the minimum Effective Sample Rate (mESR) to compare a slow sampler with a fast sampler, as in Polson, Scott, and Windle (2013) and Frühwirth-Schnatter and Frühwirth (2010). The mESR is defined as the minimum ESS per second of running time, i.e., $mESR = (\min(ESS)) / t_{\text{cpu}}$. It quantifies how rapidly a Markov-chain sampler can produce independent draws from the posterior distribution.

Vats, Flegal, and Jones (2019) argue that a univariate approach ignores cross-correlation across components, leading to an inaccurate picture of the quality of the sample. Thus, they define a multivariate version of the ESS. Specifically,

$$MESS = M \left(\frac{|\Lambda|}{|\Sigma_p|} \right)^{1/p}.$$

When there is no correlation, then $\Sigma = \Lambda$ and $MESS = M$.

To assess the convergence and mixing of our algorithm we use the Potential Scale Reduction factor (PSRF) of Brooks and Gelman (1998) and Gelman et al. (2013) and the multivariate convergence and mixing diagnostic proposed by Paul, MacEachern, and Berliner (2012). In brief, implementation of PSRF requires sample runs from

multiple chains (alternatively a very long chain can be divided into two or more subchains). The key quantity is the ratio of the resulting between- and within-chain variances. If the within-chain variance dominates the between-chain variance, the ratio approaches 1, which suggests that the chains have approximately reached stationarity. Desirable values for PSRF are the values below 1.1 for every component of the parameters. In short, Paul, MacEachern, and Berliner (2012) obtain MCMC-based estimators of posterior expectations by combining different subgroup (subchain) estimators using stratification and post-stratification methods. They develop variance estimates of the limiting distributions of these estimators. Based on these variance estimates, they propose a statistic test to aid in the assessment of convergence and mixing of chains.

Chapter 3

A Peek into the Unobservable: Hidden States and Bayesian Inference for the Bitcoin and Ether Price Series

3.1 Introduction

3.1.1 Motivation, Methodology and Main Results

The present study is motivated by the still limited understanding of the economic and financial properties of cryptocurrencies. Shedding light on such properties constitutes a necessary step for their wider public adoption and is fundamental for blockchain stakeholders, investors, interested authorities and regulators (*Crypto.com* 2019; *Paper*, 2019). More importantly, it may provide hints about market manipulation and fraud detection.

Unfortunately, existing financial models that are used to study fiat currency exchange rates fail to capture the convoluted nature of cryptocurrencies (Catania, Grassi, and Ravazzolo (2019)). The additional challenge that they face is the tight connection between cryptocurrency prices and the underlying blockchain technology which drives the dynamics of the observable market. To some extent, this is expressed via the particular market microstructure of cryptocurrencies: the market depth which depends on the exchange and the market maker, the functionality of exchanges as custodians (unique property among financial assets) and the absence of stocks, equities or other financial investment instruments (with the exception of Bitcoin futures, (Kapar and Olmo, 2019)) which render acquiring and/or trading the cryptocurrency the main way of investing in this new technology, (Koutmos, 2018). The miners and/or stakers emerge as the main actors who drive the creation and distribution of the currency whereas the cheap and immediate transactions essentially obviate the

need for conventional brokers. All these features (among many others), starkly distinguish cryptocurrencies from conventional financial assets or fiat money. However, a precise understanding of their defining financial and economic properties is still elusive, (Brauneis and Mestel, 2018; Corbet et al., 2018; Urquhart, 2018). With this in mind, the concrete research questions that we set out to understand are the following:

- How do cryptocurrencies compare – in terms of their economic and financial properties – to well understood financial assets like commodities, precious metals, equities and fiat currencies (Bouri et al., 2020; Baur, Dimpfl, and Kuck, 2018; Kang, McIver, and Hernandez, 2019)? How do they relate to traditional financial markets and global macroeconomic indicators?
- What are the defining microstructure characteristics of the cryptocurrency market and which are the distinguishing features (if any) between different coins (Aste, 2019; Katsiampa, 2019a)?

To address these questions, we use a recently developed instance of Non-Homogeneous Hidden Markov (NHHM) modeling, namely the Non-Homogeneous Pólya Gamma Hidden Markov model (NHPG) of Koki, Leonardos, and Piliouras (2019), which has been shown to outperform similar models in conventional financial data (Meligkotsidou and Dellaportas, 2011). Using financial and blockchain specific covariates on the Bitcoin (Nakamoto, 2008) and Ether (Buterin, 2014; Buterin et al., 2020) log-return series (henceforth BTC and ETH, respectively), the NHHM methodology aims not only to capture dynamic patterns and statistical properties of the observable data but more importantly, to shed some light on the unobservable financial characteristics of the series, such as the activity of investors, traders and miners.

The present model falls into the Markov-switching or regime-switching literature with two possible states that is the benchmark for predicting exchange rates (Engel, 1994a; Lee and Chen, 2006a; Frömmel, MacDonald, and Menkhoff, 2005; Beckmann and Schüssler, 2016; Groen, Paap, and Ravazzolo, 2013; Wright, 2009). This linear model was first introduced by Hamilton (1989) as an alternative approach to model non-linear and non-stationary data. It involves switches between multiple structures (equations) that can characterize the time series behavior in different regimes (states). The switching mechanism is governed by an unobservable state

variable that follows a first-order Markov chain¹. Therefore, the NHMM is suitable for describing correlated and heteroskedastic data with distinct dynamic patterns during different time periods, as are precisely cryptocurrency prices (Aggarwal, 2019; Bouri et al., 2019a; Katsiampa, 2019a).

Although standard in financial applications (Mamon and Elliott, 2014), Hidden Markov models have only been applied in the cryptocurrency context by Poyser (2019) as state space models, by Koutmos and Payne (2020) in an asset pricing model for BTC prices under heterogenous trading patterns, by Koutmos (2019) to study the relation of BTC with traditional market risk indices, by Koutmos (2018) to capture the liquidity uncertainty and Phillips and Gorse (2017) in the context of price bubbles. Yet, their more extensive use is supported by the specific characteristics of cryptocurrency data that have been identified by earlier research. Balcilar et al. (2017) and Katsiampa (2017a) and Demir et al. (2018) demonstrate the non-stationarity of BTC prices and volume and underline the importance of modeling non-linearities in Bitcoin prediction models. This is further elaborated by Beckmann and Schüssler (2016), Pichl and Kaizoji (2017), and Phillip, Chan, and Peiris (2018a) and Yu (2019) who suggest that model selection and the use of averaging criteria are necessary to avoid poor forecasting results in view of the cryptocurrencies' extreme and non-constant volatility. Along these lines, Ciaian, Rajcaniova, and Kancs (2016) show that the Bitcoin price series exhibits structural breaks and suggest that significant price predictors may vary over time. Additional motivation for the analysis of cryptocurrency data with regime-switching models as the one employed here, is provided by Katsiampa (2017a) who demonstrate the heteroskedasticity of BTC prices and Baur and Dimpfl (2018) who identify periods of different trading activity. Our main findings can be summarized as follows

- The NHPG algorithm identifies two hidden states with frequent alternations for the BTC log-return series, cf. Figure 3.2. State 1 corresponds to periods with higher volatility and returns and accounts for roughly one third of the sample period (2014-2019). By contrast, state 2 marks periods with lower volatility, series autocorrelation (long memory), trend stationarity and random walk properties, cf. Table 3.2. At the more variable state 1, the BTC data series is influenced by miners' activity and more volatile covariates (stock indices) in comparison to more stable indicators (exchange rates) in state 2, cf. Table 3.3.

¹For example, in the seminal paper of Hamilton (1989), the author used the underlying hidden process to define the business cycles (recession periods). More recent examples and a comprehensive theory about NHMM in finance can be found in Mamon and Elliott (2014).

- The results for the hidden process are the same for both the long run (2014-2019) and the short run (2017-2019) BTC data, cf. Figures 3.2 and 3.3a. However, differences in the significant predictors indicate more speculative activity in the short run compared to more fundamental investor behavior in the long run, cf. Table 3.3. In sum, speculative activity (noise traders) is identified in the less frequent state 1 and in the short run whereas increased activity of fundamental investors is seen in state 2 and in the long run.
- The algorithm does not mark a well defined hidden process with clear transitions for the ETH series, cf. Figure 3.3b. This is further supported by the low number and the small values of significant predictors from the current set, cf. Table 3.4. These results imply that ETH prices are still driven by variables beyond the currently selected set of predictors, showing characteristics of an emerging market that is more isolated than BTC from global financial and macroeconomic indicators.

More details are presented in Section 3.3. Overall, the outcome of the NHPG model can be useful for investors and blockchain stakeholders by providing hints on periods of differentiating activities and effects in the cryptocurrency markets. From a theoretical perspective, it backs earlier findings that cryptocurrencies are unlike any other financial asset and suggests that their understanding requires not only the integration of existing financial tools but also a more refined framework to account for their bundled technological and financial features (Shorish, 2018).

3.1.2 Related Literature

The literature on the financial properties of cryptocurrencies is expanding at an exponential rate and an exhaustive review is not possible (see Corbet et al. (2019) and references therein for a more comprehensive reference list). More relevant to the current context is the scarcity (to the best of our knowledge) of papers that address the bundled nature of cryptocurrencies as both blockchain applications and financial assets. Existing studies focus either on the underlying blockchain technology/consensus mechanism or on the observable financial market but not on both. By contrast, the current NHPG model parses the observable financial information to recover the underlying structure of cryptocurrency markets and hence makes a first step towards a unified approach to fill this gap. Its limitations are discussed in Section 3.4. In the remaining part of this section, we provide a (non-exhaustive) list of studies that focus on the financial part.

Early research, mainly focusing on BTC has provided mixed insights on the properties of cryptocurrencies. Klein, Thu, and Walther (2018) claim that BTC is fundamentally different from valuable metals like gold due to its shortage in stable hedging capabilities. Along with Cheah and Fry (2015), Ciaian, Rajcaniova, and Kancs (2016) also argue that standard economic theories cannot explain BTC price formation and using data up to 2015, they provide evidence that BTC lacks the necessary qualities to be qualified as money. However, Dyhrberg (2016) demonstrate that BTC has similarities to both gold and the US dollar (USD) and somewhat surprisingly, that it may be ideal for risk-averse investors. Bouri et al. (2020) and Bouri, Azzi, and Dyhrberg (2017) and Bouri et al. (2017) also explore BTC's characteristics as a financial asset and find that while BTC is useful to diversify financial portfolios – due to its negative correlation to the US implied volatility index (VIX) – it otherwise has limited *safe haven* properties. Using data from a longer period (between 2010 and 2017), Demir et al. (2018) conclude the opposite, namely that BTC may indeed serve as a hedging tool, due to its relationship to the Economic Policy Uncertainty Index (EUI). In comparative studies, Fry (2018) and Corbet, Lucey, and Yarovaya (2018) provide empirical evidence of bubbles in both BTC and ETH and Gkillas and Katsiampa (2018) suggest that BTC is less risky than ETH, i.e., that it exhibits less fat tailed behavior. Phillip, Chan, and Peiris (2018b) confirm that Bitcoin exhibits long memory and heteroskedasticity and argue that cryptocurrencies display mild leverage effects, predictable patterns with mostly oscillating persistence, varied kurtosis and volatility clustering. Comparing BTC with ETH, they argue that kurtosis is lower for ETH being easier to transact than BTC. Along this line, the findings of Mensi, Al-Yahyaee, and Kang (2019) and Katsiampa (2019a) further motivate the use of non-homogeneous and regime-switching modeling for both the BTC and ETH log-returns series.

The differences between cryptocurrencies and conventional financial markets are further elaborated by Katsiampa (2017a), Hayes (2017), and Phillip, Chan, and Peiris (2018a). High volatility, speculative forces and large dependence on social sentiment at least during its earlier stages are shown by some as the main determinants of BTC prices (Garay, Kiayias, and Leonardos, 2015; Georgoula et al., 2015; Colianni, Rosales, and Signorotti, 2015; Yi, Xu, and Wang, 2018). Yet, a large amount of price variability remains unaccounted for (Hotz-Behofsits, Huber, and Zörner, 2018; McNally, Roche, and Caton, 2018). Moreover, the proliferation of cryptocurrencies on different blockchain technologies suggests that their current correlation may be discontinued in the near future and calls for comparative studies as the one conducted

here (Borri, 2019).

3.1.3 Outline

The rest of the chapter is structured as follows. In Section 3.2, we describe the NHPG model and simulation scheme and present the set of variables that have been used (some preliminary descriptive statistics and tests about this data are relegated to Section 3.5). Section 3.3 contains the main results and their analysis. In the first part (Sections 3.3.1 to 3.3.3), we present the outcome of the algorithm and discuss the statistical findings for the hidden states and the generated subseries. In the second part, Section 3.3.4, we focus on the significant explanatory variables for the BTC data series in both the short and long run and the ETH data series. We conclude the paper with a discussion of the limitations of the present model and directions for future work in Section 3.4.

3.2 Methodology & Data

Given a time horizon $T \geq 0$ and discrete observation times $t = 1, 2, \dots, T$, we consider an observed random process $\{Y_t\}_{t \leq T}$ and a hidden underlying process $\{Z_t\}_{t \leq T}$. The hidden process $\{Z_t\}$ is assumed to be a two-state non-homogeneous discrete-time Markov chain that determines the states (s) of the observed process. In our setting, the observed process is either the BTC or the ETH log-return series. Importantly, the description of the hidden states is not pre-determined and is subject to the outcome of the algorithm and interpretation of the results.

Let y_t and z_t be the realizations of the random processes $\{Y_t\}$ and $\{Z_t\}$, respectively. We assume that at time t , $t = 1, \dots, T$, y_t depends on the current state z_t and not on the previous states. Consider also a set of $r - 1$ available predictors $\{X_t\}$ with realization $x_t = (1, x_{1t}, \dots, x_{r-1t})$ at time t . The explanatory variables (covariates) $\{X_t\}$ that are used in the present analysis are described in Table 3.1. The effect of the covariates on the cryptocurrency price series $\{Y_t\}$ is twofold: first, linear, on the mean equation and second non-linear, on the dynamics of the time-varying transition probabilities, i.e., the probabilities of moving from hidden state $s = 1$ to the hidden state $s = 2$ and vice versa. Given the above, the cryptocurrency price series $\{Y_t\}$ can be modeled as

$$Y_t \mid Z_t = s \sim \mathcal{N}(x_{t-1}B_s, \sigma_s^2), \quad s = 1, 2,$$

where $B_s = (b_{0s}, b_{1s}, \dots, b_{r-1s})'$ are the regression coefficients and $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 . The dynamics of the unobserved process $\{Z_t\}$ can be described by the time-varying (non-homogeneous) transition probabilities, which depend on the predictors and are given by the following relationship

$$P(Z_{t+1} = j \mid Z_t = i) = p_{ij}^{(t)} = \frac{\exp(x_t \beta_{ij})}{\sum_{j=1}^2 \exp(x_t \beta_{ij})}, \quad i, j = 1, 2,$$

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \dots, \beta_{r-1,ij})'$ is the vector of the logistic regression coefficients to be estimated. Note that for identifiability reasons, we adopt the convention of setting, for each row of the transition matrix, one of the β_{ij} to be a vector of zeros. Without loss of generality, we set $\beta_{ij} = \beta_{ji} = \mathbf{0}$ for $i, j = 1, 2, i \neq j$. Hence, for $\beta_i := \beta_{ii}$, $i = 1, 2$, the probabilities can be written in a simpler form

$$p_{ii}^{(t)} = \frac{\exp(x_t \beta_i)}{1 + \exp(x_t \beta_i)} \text{ and } p_{ij}^{(t)} = 1 - p_{ii}^{(t)}, \quad i, j = 1, 2, \quad i \neq j.$$

To make inference on the hidden process, we use the smoothed marginal probabilities $P(Z_t = i \mid Y_{1:T}, z_{t+1}, \theta)$ which are the probabilities of the hidden state conditional on the full observed process, (Hamilton, 1989) as derived from the Forward-Backward algorithm. Details on the estimation of the hidden process via the Scaled Forward-Backward which is explained in detail in Section 3.2.1. In the rest of the paper, we use the notation $P(Z_t = i)$ for $i = 1, 2$ to denote the smoothed probabilities for convenience.

3.2.1 The Scaled Forward-Backward algorithm

We leverage the scaled stochastic version of the Forward-Backward algorithm of Scott (2002) for estimating the hidden Markov process. This algorithm is consisted of a set of filtering recursions for calculating the likelihood and simulating realizations of the hidden process. Firstly, the forward variable is defined as

$$\alpha_t(Z_t) = P(y^t, Z_t = s \mid \theta) = P(Z_t = s) P(y^t \mid Z_t = s, \theta),$$

where y^t is the vector with all the observations up to time t , i.e., $y^t = (y_1, \dots, y_t)$. The forward variable is basically the joint likelihood contribution of the data y up to time t and the event $Z_t = s$, averaging over the events up to time $t - 1$. Similarly, $\alpha_t^* = \sum_{s=1}^2 \alpha_t(s)$ is the likelihood contribution from y^t . Traditionally, the forward

variables are computed recursively as,

$$\begin{aligned}\alpha_t(s) &= \sum_{j=1}^2 P(y_t, y^{t-1}, Z_{t-1} = j, Z_t = s \mid \theta) \\ &= \left[\sum_j \alpha_{t-1}(j) p_{js}^{(t-1)} \right] f_s(y_t).\end{aligned}$$

However, using likelihood instead of logarithmic likelihood, is unstable and can cause overflow problems during the computer calculations. Scott (2002) proposed the following remedy. Define the scaled forward probabilities,

$$\pi_t(s \mid \theta) = \frac{\alpha_t(s)}{\alpha_t^*} = P(Z_t = s \mid y^t, \theta),$$

and the quantity

$$M_t = \max_s \log \left\{ \left[\sum_j \alpha_{t-1}(j) p_{js}^{(t-1)} \right] f_s(y_t) \right\}.$$

Then with some straight forward algebra, we derive the recursive relationship

$$\log(\alpha_t^*) = \log(\alpha_{t-1}^*) + M_t + \log \left[\sum_j \exp \left[\log(f_s(y_t \mid \theta)) + \log \left(\sum_j p_{js} \pi_{t-1}(j \mid \theta) \right) - M_t \right] \right].$$

Hence, the more stable scaled forward probability can be calculated recursively using the following,

$$\log(\pi_t(s \mid \theta)) = \log(\alpha_t(s)) - \log(\alpha_t^*) = Q_t(s) - M_t - \log \left(\sum_j \exp [Q_t(s) - M_t] \right),$$

where $Q_t(s) = \log(f_s(y_t \mid \theta)) + \log(\sum_j p_{js}^{(t-1)} \pi_{t-1}(j \mid \theta))$. Under the assumption that the dependent variable follows a Normal distribution with regime-switching mean and standard deviation, $f_s(y_t)$ is just the probability density function of the Normal distribution with parameters the simulated values of θ in regime s .

The hidden states can be simulated from the conditional or marginal distribution via the backward variable. However, we use a stochastic version of the Forward-Backward algorithm that mixes rapidly and is less computational demanding. First, we begin by drawing Z_T from,

$$P(Z_T = s \mid y^T) = \frac{P(y^T, Z_T = s)}{P(y^T)} = \frac{\alpha_T(s)}{\sum_j \alpha_T(j)} = \pi_T(j).$$

Then, due to the Markov property and the assumed dependencies of the HHMs, applying the Bayes rule will lead to the following relationship,

$$\begin{aligned} P(Z_t = i \mid y^T, Z_{t+1}) &= \frac{P(Z_t = i, y^T, Z_{t+1})}{P(y^T, Z_{t+1})} \\ &\propto P(Z_t = i, y_1, \dots, y_t) P(Z_{t+1} \mid Z_t) P(y_{t+1}, \dots, y_T \mid Z_{t+1}) \\ &\propto p_{iZ_{t+1}} \alpha_t(i). \end{aligned}$$

Thus, the given the realization of the hidden state Z_{t+1} , Z_t is drawn from

$$P(Z_t = i \mid y^T, Z_{t+1}) = \frac{p_{iZ_{t+1}}^{(t)} \pi_t(i)}{\sum_j p_{jZ_{t+1}}^{(t)} \pi_t(j)}.$$

Having described the statistical methodology we are now ready to describe our simulation scheme.

3.2.2 Simulation Scheme

The unknown quantities of the NHPG are $\{\theta_s = (B_s, \sigma_s^2), \beta_s, s = 1, 2\}$, i.e., the parameters in the mean predictive regression equation and the parameters in the logistic regression equation for the transition probabilities. We follow the methodology of Koki, Meligkotsidou, and Vrontos (2020). In brief, the authors propose the following MCMC sampling scheme for joint inference on model specification and model parameters.

1. Given the model's parameters, the hidden states are simulated using the Scaled Forward-Backward of algorithm of Scott (2002).
2. The posterior mean regression parameters are simulated using the standard conjugate analysis, via a Gibbs sampler method.
3. The logistic regression coefficients are simulated using the Pólya-Gamma data augmentation scheme Polson, Scott, and Windle (2013), as a better and more accurate sampling methodology compared to the existing schemes.

The steps 1-3 of the MCMC algorithm are detailed in Algorithm 1.

3.2.3 Data

We assess the ability of 11 financial–macroeconomic and 3 cryptocurrency specific variables, outlined in Table 3.1, in explaining and forecasting the prices of BTC and

Algorithm 1 MCMC Sampling Scheme for Inference on Model Specification and Parameters

```

1: % After each procedure the parameters and model space are updated condition-
   % ally on the previous quantities
2: procedure SCALED_FORWARD_BACKWARD( $(z_{1:T})$ )
3:   %Simulation of a realization of the hidden states  $z_t$ 
4:   for  $t = 1, \dots, T$  and  $i = 1, 2$  do
5:      $\pi_t(i | \theta) \leftarrow \frac{\alpha_t(s)}{\sum_{j=1}^2 \alpha_t(j)} = P(z_t = i | \theta, y^t)$ 
6:     (▷) Simulation of the scaled forward variables
7:   end for
8:   for  $t = T, T - 1, \dots, 1$  do
9:      $z_t \leftarrow P(z_t | y^T, z_{t+1}) = \frac{p_{iz_{t+1}} \pi_t(i | \theta)}{\sum_{j=1}^m p_{jz_{t+1}} \pi_t(j | \theta)}$ 
10:    (▷) Backwards simulation of  $z_t$  using the smoothed probabilities
11:   end for
12: end procedure
13: procedure MEAN_REGRES_PARAM( $B_s, \sigma_s^2, s = 1, 2$ )
14:   %Simulation of the mean regression parameters
15:   for  $s = 1, 2$  do (▷) Conjugate analysis with Gibbs sampler
16:      $B | \sigma^2 \sim f_B, \sigma^2 \sim \mathcal{IG}$  (▷)  $f_B \equiv$  Normal and  $\mathcal{IG} \equiv$  Inverse-Gamma
17:   end for
18: end procedure
19: procedure LOG_REGRES_COEF( $(\beta_s, \omega_s)$ )
20:   %Simulation of the logistic regression coefficients
21:   for  $s = 1, 2$  do (▷) Pólya-Gamma data augmentation scheme
22:     → augment the model space with  $\omega_s$ 
23:     (▷) Conjugate analysis on the augmented space
24:     → sample from  $\beta_s \sim f_{\beta_s|\omega}$  and  $\omega_s | \beta_s \sim \mathcal{PG}$ 
25:     → posteriors  $f_{\beta_s|\omega} \equiv$  Normal and  $\mathcal{PG} \equiv$  Pólya-Gamma
26:   end for
27: end procedure

```

ETH via the NHPG model. In the related cryptocurrency literature these indices are commonly studied under various settings (Van Wijk, 2013; Yermack, 2015; Bouri et al., 2017; Estrada, 2017; Pichl and Kaizoji, 2017; Hotz-Behofsits, Huber, and Zörner, 2018). The findings of the descriptive statistics and preliminary stationarity tests, cf. Section 3.5, indicate that the logarithmic return (log-return), i.e., the change in log price, $r_t = \log(y_t) - \log(y_{t-1})$, series of BTC and ETH exhibit trend non-stationarity, non-linearities, rich (i.e., non-random) underlying information structure and non-normalities. Based on these properties, the NHPG model seems appropriate for the study of the log-return data series. Accordingly, we apply the NHPG algorithm on daily log-returns of BTC and ETH, with normalized explanatory variables.

We perform two experiments over two different time frames: in the first, we study the BTC series between 1/2014 and 8/2019 and in the second, we study both the BTC and ETH series between 1/2017 and 8/2019. The second time frame has been selected to allow reasonable comparisons between the BTC and ETH prices after eliminating an initial period following the launch of the ETH currency. It is further motivated by the outcome of a test-run of the NHPG model on BTC prices, cf. Figure 3.1, which indicates a transition point to a different period for the BTC price series in January 2017.

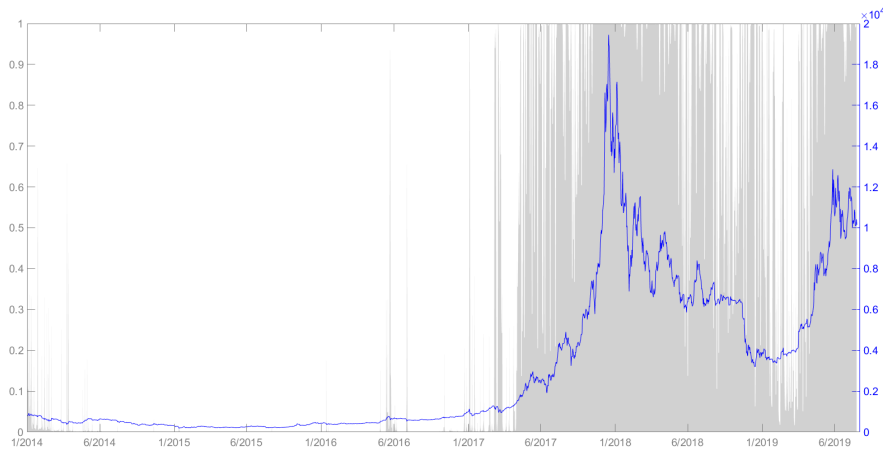


FIGURE 3.1: Application of the NHPG model on the BTC price series. The algorithm essentially identifies two periods, the first from 2014 (start of the dataset) to 2017 and the second from 2017 to date. This motivates separate analysis of the BTC for the latter period and comparison with the ETH price series over the same period.

3.3 Results & Analysis

In this section, we discuss the findings from the NHPG model on the BTC and ETH log-return series. We first present the graphics with the output of the algorithm for the whole 2014-2019 period on BTC log-returns (Section 3.3.1) and the shorter 2017-2019 period on both BTC and ETH log-returns (Section 3.3.2). Then, we interpret the results and compare the statistical properties and the significant covariates between the two hidden states of both the BTC and ETH series and between the short and long run BTC series (Sections 3.3.3 and 3.3.4).

3.3.1 Hidden States: Bitcoin 2014–2019

Figure 3.2 displays the BTC log-return series (blue line) along with the smoothed marginal probabilities (gray bars) of the hidden process being at state 1. Using as a threshold the probability $P(Z_t = 1) > 0.5$, we estimate the hidden states for each

Explanatory Variables		
Description	Symbol	Retrieved from
US dollars to Euros exchange rate	USD/EUR	investing.com
US dollars to GBP exchange rate	USD/GBP	investing.com
US dollars to Japanese Yen exchange rate	USD/JPY	investing.com
US dollars to Chinese Yuan exchange rate	USD/CNY	investing.com
Standard & Poor's 500 index	SP500	finance.yahoo.com
NASDAQ Composite index	NASDAQ	finance.yahoo.com
Silver Futures price	Silver	investing.com
Gold Futures price	Gold	investing.com
Crude Oil Futures price	Oil	investing.com
CBOE Volatility index	VIX	finance.yahoo.com
Equity market related Economic Uncertainty index	EUI	fred.stlouisfed.org
Daily Block counts	Blocks	coinmetrics.io
Hash Rate	Hash	quandl.com, etherscan.io
Transfers of native units	Tx-Units	coinmetrics.io

TABLE 3.1: List of variables and online resources. The Hash Rate (Hash) has been retrieved from quandl.com for Bitcoin (BTC) and from etherscan.io for Ether (ETH).

time period. The NHPG model identifies a subseries of 667 observations in state 1 and a subseries of 1388 observations in state 2. The description of the hidden states is not predetermined by the model and is done a posteriori, by comparison of the statistical properties of the two subseries that have been generated. As it is obvious from Figure 3.2, state 1 corresponds to periods of larger log-returns and increased volatility in comparison to state 2. The frequent changes are in line with previous studies on the heteroskedasticity and on the regime switches (structural breaks) of the Bitcoin time series (Phillips and Gorse (2017) and Katsiampa (2017a) and Thies and Molnár (2018), Koutmos (2018), Ardia, Bluteau, and Rüede (2019), and Bouri et al. (2019b), respectively). Yet, the refined outcome of the NHPG model, which determines the time periods that the series spends in each state, allows for a more granular approach. Specifically, it adds information about the significant covariates that affect both the observable and the unobservable process and on the financial properties of each state. This is done in Section 3.3.4 below.

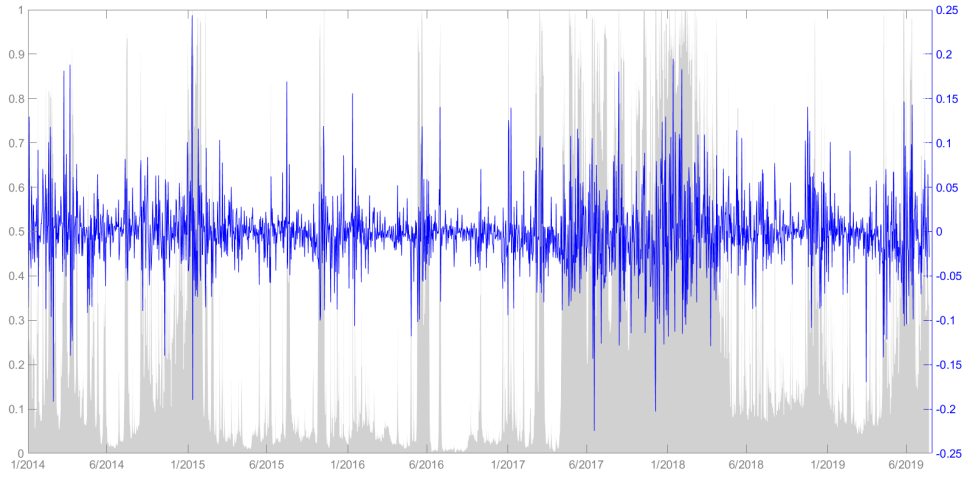


FIGURE 3.2: BTC logarithmic-return series (blue line – right axis) for the period 1/2014–8/2019 with the mean smoothed marginal probabilities of state 1, i.e., $Pr(Z_t = 1)$ (gray bars – left axis).

3.3.2 Hidden States: Bitcoin and Ether 2017–2019

Figure 3.3 shows the results of the NHPG model for both the BTC (left panel) and ETH (right panel) log-return series over the shorter 1/2017–8/2019 period. The algorithm has again identified two states in the BTC series, Figure 3.3a, as indicated by the clear distinction between high-low marginal probabilities of state 1, i.e., $P(Z_t = 1)$, that are given by the gray bars. Moreover, a comparison with the same period in Figure 3.2 demonstrates that the NHPG has produced the same result (zoom in) – in terms of statistical quality – even over this smaller period, i.e., the algorithm has converged and returns essentially the same probabilities for the underlying process. However, as we will see below, cf. Section 3.3.4, the statistical analysis unveils differences in the significant predictors and financial properties between the short and long run.

The picture is different for the ETH series, cf. Figure 3.3b. Here, the hidden process is not well defined since the probabilities of state 1 at each time period are

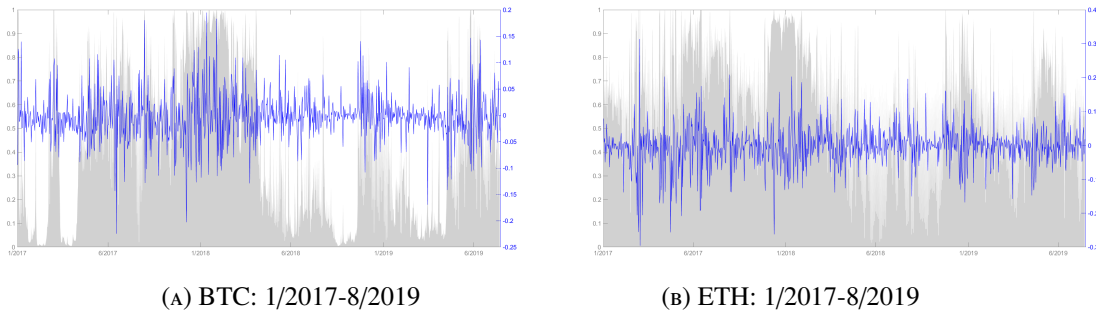


FIGURE 3.3: BTC (left panel) and ETH (right panel) logarithmic-returns series (blue lines – right axis) for the period 1/2017–8/2019 with the mean smoothed marginal probabilities of state 1, i.e., $Pr(Z_t = 1)$, (gray bars – left axis).

mostly close to 0.5. This indicates high degree of randomness in the transitions of the algorithm and along with the low number of significant covariates that have been identified for ETH (cf. Table 3.4 below), it suggests that ETH prices are still influenced by forces which are beyond the current set of financial and blockchain indicators (Katsiampa, 2019b; Phillip, Chan, and Peiris, 2018a; Phillip, Chan, and Peiris, 2018b). This implies that ETH – when viewed as a financial asset – shows characteristics of an evolving, non-static and still emerging market. However, the relative isolation of ETH from other financial assets agrees with earlier findings in the literature (Phillip, Chan, and Peiris, 2018a; Corbet et al., 2018).

Our next task is to provide additional insight on the structural financial and economic attributes that differentiate these two states for all experiments. Based on the similarities between the short and long run BTC time frames and the poor convergence of the algorithm for the ETH series, we focus on the long-run BTC series.

3.3.3 Hidden States: Financial Properties (BTC 2014-2019)

The results of both the descriptive statistics and the relevant statistical tests are summarized in Table 3.2. Each entry – BTC price, log-price and log-return series – consists of two rows that correspond to the subseries of state 1 (upper row) and state 2 (lower row), respectively. The first two columns of Table 3.2 verify that the estimated hidden process segments the series into two subseries with high/low mean and variance values for all the examined data series. Log-returns exhibit increased kurtosis in comparison to the initial estimates, cf. Table 3.5, for both subseries (in particular for state 2). Similarly, the skewness of both subseries has increased and has turned positive with the skewness of the second subseries being again much higher than that of the first. These distributional properties lead to rejection of normality for either subseries and suggest the presence of heavy-tailed data (phenomena in which extreme events are likely, ²).

The identification of two subchains with different kurtosis and skewness can be a useful tool to investors (Jondeau and Rockinger, 2003; Konno, Shirakawa, and Yamazaki, 1993; Dittmar, 2002). As risk measures, kurtosis and skewness cause major changes to the construction of the optimal portfolio (Chunhachinda et al., 1997; Conrad, Dittmar, and Ghysels, 2013), especially in emerging and highly volatile markets (Canela and Collazo, 2007). The asymmetry on the distributions and the difference of volatility between the two subchains can be related to the activity of informed

²Inclusion of a third hidden state could potentially lead to smoothing of these measurements, cf. Section 3.4.

		Descriptive statistics				Tests				
		Mean	Variance	Kurtosis	Skewness	DF	LBQ	KPSS	VR	JB
BTC	Price	4920	2.1×10^7	2.79	0.81	0.58	0	0.01	0.86	0.00
		2133	7.8×10^6	3.99	1.48	0.79	0.00	0.01	0.14	0.00
	Log-Price	7.83	1.84	1.71	-0.41	0.95	0	0.01	0.86	0.00
		6.87	1.49	1.98	0.68	0.97	0.00	0.01	0.45	0.00
	Log-Return	0.0039	0.0050	7.85	0.76	0.00	0.77	0.02	0	0.00
		0.0018	0.0023	45.48	2.68	0.00	0.00	0.10	0.08	0.00

TABLE 3.2: Descriptive statistics (left panels) and p-values for the time series statistical tests (right panels) for the two (2) BTC price, log-price and log-return **subseries** – first and second line of each entry – which correspond to the two hidden states that were identified by the NHPG model for the whole 1/2014-8/2019 time period.

or fundamental vs uninformed, noise or non-fundamental investors (or traders). Intuitively, the activity of uninformed investors leads to periods with higher volatility (cf. Baur and Dimpfl, 2018 and references therein). This is true for state 1 and refines the findings of Zargar and Kumar (2019a) and Baur and Dimpfl (2018) who attribute the informational inefficiency of BTC not only to its endogenous factors of an emerging, non-mature market but also to the non-existence of fundamental traders.

The differences between the two states are further explained by the statistical tests. While the p-values of the Dickey-Fuller (DF) and Jarque-Bera (JB) remain the same as for the combined data series, cf. Table 3.5, the results for the Ljung-Box-Q (LBQ), KPSS and Variance Ratio (VR) tests unveil different characteristics of the two subseries. In state 2 of the log-return series, the zero hypothesis is rejected for the LBQ test but not for the KPSS and VR tests. This suggests that the subseries defined by state 2 is a random walk with trend stationarity and long memory. These findings are related to (and to some extent refine) the results of Jiang, Nie, and Ruan (2018), Khuntia and Pattanayak (2020), Mensi, Al-Yahyaee, and Kang (2019), and Zargar and Kumar (2019b) by determining periods with (state 2) and without (state 1) permanent effects (long memory). The subchain of state 1 stills exhibits richer structure which can be potentially attributed to the combined activity and herding behavior of the non-fundamental traders (Bouri, Gupta, and Roubaud, 2019; Stavroyiannis and Babalos, 2019).

3.3.4 Significant Explanatory Variables: Bitcoin and Ether

The second functionality of the NHPG model is to identify the significant explanatory variables from the set of available predictors that affect the underlying series both linearly, i.e., in the mean equation (observable process), and non-linearly, i.e.,

in the non-stationary transition probabilities (unobservable process). The algorithm also distinguishes between the variables that are significant in each state. The corresponding results for the BTC log-return series over both the 2014-2019 and 2017-2019 time periods are given in Table 3.3 and the results for the ETH log-return series over the 2017-2019 time period are given in Table 3.4. We use B_i to denote the posterior mean equation coefficients and β_i the posterior mean logistic regression coefficients for states $i = 1, 2$, as described in Section 3.2. The predictors that have

Estimations BTC								
Variables	2014-2019				2017-2019			
	B_1	B_2	β_1	β_2	B_1	B_2	β_1	β_2
USD/EUR	0.00	0.00	0.19	1.82*	-0.01	0.00	0.48	3.97*
USD/GBP	0.02*	≈ 0	-1.35*	-1.68*	-0.01	0.00	-1.82	4.34*
USD/JPY	0.00	0.00	0.52	-0.77	0.00	-0.00	-0.53	-0.77
USD/CNY	-0.01	0.00	0.90	0.57	≈ 0	0.00	1.98	1.65
SP500	0.04*	-0.01	3.90	-1.87	0.04	0.01	8.62*	1.23
NASDAQ	-0.04	0.00	-1.65	-2.24	-0.00	-0.02	-8.26*	-2.04
Silver	0.00	≈ 0	0.22	0.57	0.01	-0.00	-0.42	1.15
Gold	0.00	≈ 0	1.40*	-0.18	0.00	0.00	1.19*	-0.35
Oil	≈ 0	≈ 0	-0.14	0.00	0.00	0.00	1.27*	1.98*
VIX	≈ 0	≈ 0	0.47	-0.18	≈ 0	≈ 0	0.53	0.18
EUI	≈ 0	≈ 0	0.00	-0.00	≈ 0	≈ 0	0.00	0.00
Blocks	0.00	≈ 0	-0.26	-0.07	-0.01	-0.00	-0.40	0.06
Hash	0.01*	0.00	-1.84*	-0.03	0.01*	0.01	-2.48*	1.60
Tx-Units	≈ 0	0.00	0.51	-0.13	≈ 0	0.00	0.53	-0.50

TABLE 3.3: Posterior mean estimations for the BTC log-return series in the 2014-2019 (left) and 2017-2019 (right) time periods. B_1, B_2 are the mean equation coefficients and β_1, β_2 are the logistic regression coefficients for states 1, 2. Statistically significant coefficients (at the 0.05 level) are marked with *.

been found significant at the 0.05 level are marked with bold font and *. The main findings are the following:

BTC: state 1 vs state 2. The significant predictors (covariates) that dominate both the observable and the unobservable processes in the more volatile state 1 (cf. Section 3.3.3), correspond to more volatile financial instruments such as stock markets (S&P500 and NASDAQ). By contrast, state 2 is mostly influenced by the more stable exchange rates, cf. Figure 3.4. These findings suggest

increased speculative activity in state 1 in comparison to fundamental investors in state 2.

BTC: short vs long run. While the algorithm has identified essentially the same hidden process for both the short and long run windows, cf. Figures 3.2 and 3.3a, the significant predictors that affect both the observable and unobservable processes are remarkably different: more volatile for the short run versus more fundamental (monetary) for the long run. In line with de la Horra, de la Fuente, and Perote (2019), these findings provide evidence for increased speculative behavior in the short run. They also extend BTC's financial and safe haven properties to more recent windows (Poyser, 2019; Balcilar et al., 2017; Bouri, Azzi, and Dyhrberg, 2017). Additionally, they refine the results of Corbet and Katsiampa (2018) and Chaim and Laurini (2019) who argue about the differences in the short and long run BTC markets and the hedging properties of BTC against volatile stock indices in time varying periods, respectively.

ETH vs BTC: short run. The lower number of significant predictors in the ETH log-return series reflects the inability of the NHPG model to parse the underlying process, cf. Figure 3.3b. This differentiates the ETH from the BTC market and provides evidence that ETH is still at its infancy, evolving independently from established economic indicators and fundamentals. Yet, the main – and somewhat unexpected – conclusion is that, despite the evident correlation between the prices of BTC and ETH (Pearsons serial correlation 0.62), the two cryptocurrencies are affected by different fundamental financial and macroeconomic indicators over the same time period.

Finally, an observation that applies to all series is that the current set of predictors cannot fully explain the data volatility. Excluding the miners' activity (as expressed by the Hash Rate) which appears significant in state 1 for all series (both for the observable and the unobservable processes), this observation follows from the small values of the predictors in the mean equation of state 1 (cf. columns B_1 in Table 3.3) and the absence of predictors in the mean equation (observable process) of state 2 (cf. columns B_2 in Table 3.3).

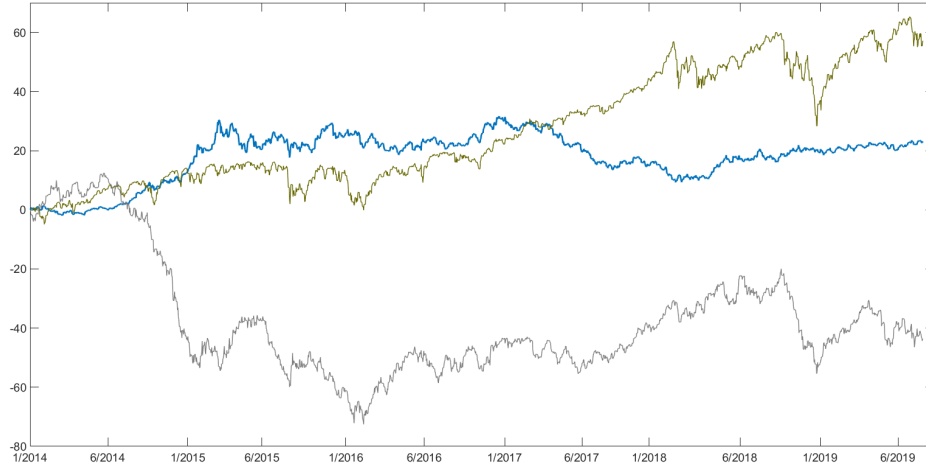


FIGURE 3.4: Comparison of the USD/EUR exchange rate (blue line), S&P500 (green line) and Crude Oil Future Prices (gray line) as a percentage of price changes from the initial period. The USD/EUR exchange rate is less volatile than the other predictors.

Estimations ETH 2017-2019									
Variables	B_1	B_2	β_1	β_2	Variables	B_1	B_2	β_1	β_2
USD/EUR	-0.01	-0.00	-0.36	-0.38	USD/GBP	0.01	-0.06	0.21	0.59
USD/JPY	-0.01	-0.00	-0.69	0.68	USD/CNY	-0.02	0.01	-0.14	-0.01
SP500	0.06*	-0.00	2.70	0.13	VIX	0.01	0.00	-0.19	0.26
NASDAQ	-0.06*	0.00	-4.11*	0.40	EUI	≈ 0	≈ 0	-0.00	-0.71*
Silver	-0.01	0.01	-0.60	0.07	Blocks	-0.01	≈ 0	0.36	2.20
Gold	0.00	-0.01	-0.28	0.21	Hash	0.04*	0.00	-0.46*	-1.41
Oil	-0.01	-0.01	-0.64	-1.92	Tx-Units	0.00	0.00	1.03	1.16

TABLE 3.4: Posterior mean estimations for the ETH log-return series in the 2017-2019 time period. Statistically significant coefficients (at the 0.05 level) are marked with *.

3.4 Discussion: Limitations and Future Work

The application of NHHM modeling in cryptocurrency markets comes with its own limitations. From a methodological perspective, the main concerns stem from the decision rule for each state which is probabilistic and the exogenously given number of hidden states. In the present study, we used the threshold of 0.5 to decide transitions from state 1 to state 2 and vice versa. However, in the related financial literature, there are many different approaches even with lower thresholds. Moreover, while two hidden states are generally considered the norm in most financial applications, the current results suggest that it may be worth exploring the possibility of a third hidden state. Alternatively, one may define a *gray zone* for time periods in which

the algorithm returns probabilities around 0.5 for both states. This will allow for the identification of periods with high uncertainty about the underlying process and hence, will lead to more scarce, yet more uniform (in terms of distributional properties) subseries.

From a contextual perspective, the present approach does not account for qualitative attributes of the predictive variables. For instance, it does not measure centralization of the transactions or alleged fake volumes among different exchanges (Gandal et al., 2018; Bouri et al., 2019a). Coupling the present approach with transaction graph analysis, Dixon et al. (2019), and user metrics to capture potential market manipulation and the purpose of usage such as speculative trading or exchange of goods and services (Cheah and Fry, 2015; Blau, 2017; Baur and Dimpfl, 2018) will lead to improved results. Lastly, as more blockchains transition to alternative consensus mechanisms such as Proof of Stake, further iterations of the current model should also include the underlying technology (e.g., staking versus mining) as a determining factor. At the current stage, such a comparative study is not possible from a statistical perspective, since the market capitalization and trading volume of conventional Proof of Work cryptocurrencies is still not comparable to that of coins with alternative consensus mechanisms. The long-anticipated transition of the Ethereum blockchain to Proof of Stake consensus may define such an opportunity in the near future.

Along these lines, extensions of the current model may enrich the set of covariates (explanatory variables) to capture technological features and/or advancements of various cryptocurrencies, refine the NHPG model with potentially three hidden states and finally, couple the statistical/economic findings with transaction graph analysis. The expected outcome is a more detailed understanding of the financial properties of cryptocurrencies and the assembly of a model with improved explanatory and predictive ability for cryptocurrency markets.

3.5 Appendix:

3.5.1 Descriptive statistics

In Table 3.5, we summarize the descriptive statistics for the BTC and ETH data series, log-prices and the p-values of the necessary preliminary statistical tests that assess the properties of the given data series prior to the application of the NHPG model. In detail:

Mean & variance: We report the mean and variance of prices, log-prices and log-returns of BTC and ETH. As expected, all series exhibit very high (to extreme) volatility.

Kurtosis: Based on the kurtosis values, the distributions of all series – except the log-price BTC series – are leptokurtic, i.e., they exhibit tail data exceeding the tails of the normal distribution (values above 3), indicating the large number of outliers (extreme values).

Skewness: Additionally, we report the skewness values, as measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean and the opposite if skewness is positive. We observe that the price series are highly right skewed, whereas the skewness of the log-returns for both coins are close to 0, indicating an approximately symmetrical, around the mean, series.

		Descriptive statistics				Tests				
		Mean	Variance	Kurtosis	Skewness	DF	LBQ	KPSS	VR	JB
BTC	Price	3057	1.37×10^7	4.35	1.40	0.54	0	0.01	0.80	0.00
	Log-Price	7.18	1.82	1.56	0.32	0.96	0	0.01	0.54	0.00
	Log-Return	0.0012	0.0016	7.78	-0.27	0	0.31	0.03	0	0.00
ETH	Price	311	6.41×10^4	4.95	1.43	0.31	0	0.01	0.43	0.00
	Log-Price	5.34	1.14	4.42	-1.17	0.91	0	0.01	0.90	0.00
	Log-Return	0.0032	0.0037	6.14	0.24	0	0.88	0.01	0	0.00

TABLE 3.5: Descriptive statistics (left panels) and p-values of the time series statistical tests (right panels) for the BTC and ETH price, log-price and log-return series. DF denotes the Dickey-Fuller test, KPSS the Kwiatkowski-Phillips-Schmidt-Shin test, LBQ is the Ljung-Box Q test, VR the variance ratio test and JB the Jarque-Bera test.

3.5.2 Statistical tests

Stationarity captures the intuitive idea that certain properties of a (data generating) process are unchanging. This means that if the process does not change at all over time, it does not matter which sample portion of observations we use to estimate the parameters of the process, cf. Sections 3.3.1 and 3.3.2.

DF-ADF: First, we report the p-values of the Dickey-Fuller (DF) unit root test³. This test assesses the null hypothesis of a unit root using the model $y_t = \phi y_{t-1} + \epsilon_t$. The null hypothesis is $H_0 : \phi = 1$ under the alternative $H_1 : \phi < 1$. The H_0 was rejected only in the log-return series. The existence of the unit root is one of the common causes of non-stationarity. Intuitively, if a series is unit root nonstationary then the impact of the previous shock ϵ_{t-1} on the series has a permanent effect on the series.

LBQ: To test for serial autocorrelation on the long-run, i.e., to detect if the observations are random and independent over time, we used the Ljung-Box-Q (LBQ) test which assesses the presence of autocorrelations (ρ) at lags p in one hypothesis, jointly. The null hypothesis of the LBQ test is $H_0 : \rho_1 = \dots = \rho_p = 0$, under every possible alternative. The null hypothesis was not rejected only for the log-return series and for lags up to $p_{BTC} = 10$ and $p_{ETH} = 6$, for BTC and ETH respectively. However, when $p_{BTC} > 10$ and $p_{ETH} > 6$ the null hypothesis was rejected, indicating long-memory (persistent) log-return series.

KPSS: The next column presents the p-values of the Kwiatkowski, Phillips Schimdt, Shin (KPSS) test. The KPSS test assesses the null hypothesis that a univariate time series is trend stationary against the alternative that it is a non stationary unit root process. The test uses the structural model: $y_t = c_t + \delta_t + u_{1t}$, $c_t = c_{t-1} + u_{2t}$ where δ_t is the trend coefficient at time t , u_{1t} is a stationary process and u_{2t} is an independent and identically distributed process with mean 0 and variance σ^2 . The null hypothesis is that $\sigma^2 = 0$, which implies that the random walk term (c_t) is constant and acts as the model intercept. The alternative hypothesis is that $\sigma^2 > 0$, which introduces the unit root in the random walk. Based on the p-values, we reject all the hypothesis of trend stationarity of the series.

VR: Additionally, we report the p-values of the Variance Ratio (VR) test which assesses the hypothesis of a random walk. The random walk hypothesis provides a mean to test the weak-form efficiency – and hence, non-predictability – of financial markets, and to measure the long run effects of shocks on the path of real output in macroeconomics, see Charles and Darné (2009) and references therein. The model under the H_0 is $y_t = c + y_{t-1} + \epsilon_t$, where c is a drift constant and ϵ_t are uncorrelated

³We also performed the Augmented Dickey-Fuller test with drift c , which assesses the null hypothesis of a unit root using the model $y_t = c + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p} + \epsilon_t$ where $\Delta y_t = y_t - y_{t-1}$ and lagged operator $p = 7$. The results were similar as the DF test. The null hypothesis was rejected only in the log-returns series for both cryptocurrencies.

innovations with zero mean. The random walk hypothesis is rejected only in the log-return series for both coins. Essentially, the rejection of the random walk hypothesis shows that there exists information that can be used in explaining the movement of the returns.

JB: Lastly, we report the Jarque-Bera (JB) test, as a normality test. Based on these results, all the series are not normally distributed.

Chapter 4

Exploring the Predictability of Cryptocurrencies via Bayesian Hidden Markov Models

4.1 Introduction

The growth of cryptocurrency markets made a splash in the world. At present, there are more than one thousand cryptocurrencies that constitute a multi-billion market (Hu, Parlour, and Rajan, 2019). Due to their increasing popularity and intriguing financial and econometric properties (Dyhrberg, Foley, and Svec, 2018), they have attracted the interest of investors, regulatory authorities, policymakers, tech-savvy entrepreneurs and academics (Corbet et al., 2019). Their documented safe haven and hedge properties, most relevant in periods with volatile stock markets and inflationary pressures in fiat currencies, render cryptocurrencies increasingly important in portfolio optimization and risk diversification (Urquhart and Zhang, 2019; Bouri et al., 2020; Platanakis and Urquhart, 2020). In turn, informed decisions concerning optimal portfolio allocation and asset management require models with good predicting ability (Chen, Li, and Sun, 2020).

Similar to forecasting studies about conventional financial assets and exchange rates (McMillan, 2020; Panopoulou and Souropanis, 2019), this has prompted a growing literature on the predictability of cryptocurrency returns. Existing results range from the identification of significant explanatory variables ((Aalborg, Molnár, and Vries, 2019; Bleher and Dimpfl, 2019; Kurka, 2019) and (Corbet et al., 2019; Katsiampa, 2019a) for comprehensive surveys of earlier models) to price prediction with elaborate machine and deep learning models (Chen, Li, and Sun, 2020; Chen et al., 2020). Under the Bayesian framework, main efforts involve continuous

state space models (Hotz-Behofsits, Huber, and Zörner, 2018) univariate and multivariate dynamic linear models, model averaging and time-varying vector autoregression models (Catania, Grassi, and Ravazzolo, 2019). These articles show that time-varying models give significantly improved point and density forecasts when compared to various benchmarks such as the random walk model.

The improved performance of the state space and time varying models is not a surprise, since these models accommodate various characteristics of the cryptocurrency series, such as time varying volatility and time varying mean returns. Accumulating evidence points to the existence of structural breaks (Meligkotsidou et al., 2019; Bouri et al., 2019b; Katsiampa, 2019a; Thies and Molnár, 2018) return and volatility jumps (Shen, Urquhart, and Wang, n.d.; Chaim and Laurini, 2018) and regime/state switches (Ardia, Bluteau, and Rüede, 2019; Koutmos, 2019; Koutmos, 2018) in cryptocurrency returns. However, while regime switching models have been shown to deliver improved forecasting results in exchange rates series and stock market returns (see e.g., Panopoulou and Pantelidis, 2015; Dias, Vermunt, and Ramos, 2015; Yuan, 2011 among others), their application in the cryptocurrency context is still limited (cf. Section 4.1.1).

Stimulated by the above and aiming to contribute to the growing literature on the predictability of cryptocurrencies, we perform a systematic analysis of various multi-state (regime-switching) Hidden Markov (HM) models on the return series of the three largest (in terms of market capitalization) cryptocurrencies, Bitcoin (BTC), Ether (ETH) and Ripple (XRP). Our goal is to examine the impact of regime switches in predicting the return series and the state-dependent (time-varying) effects of several financial, economic and cryptocurrency specific exogenous predictors. In total, we consider eight discrete state space HM models with exogenous predictors, cf. Table 4.2. The models include between 2 and 5 hidden states,¹ and either Homogeneous (HHM) or Non-Homogeneous (NHHM) transition probabilities. We also consider the standard 2-state Markov Switching Random Walk (MS-RW) model without exogenous predictors. We benchmark the aforementioned HM models against three single regime models: the Random Walk (RW) model that is commonly used (as a benchmark) in predicting exchange rates (Panopoulou and Pantelidis, 2015; Frömmel, MacDonald, and Menkhoff, 2005; Yuan, 2011; Cheung and Erlandsson, 2005), the linear random walk model with all the predictors, often referred to as

¹To determine the number of states we undertook an extensive specification test. Experiments with more than 5 states (not presented here) exhibit worst performance. Even though adding more states may improve the in-sample fit, the decreased parsimony leads to worse predictions (Guidolin and Timmermann, 2006).

the *Kitchen Sink* (KS) model, (Catania, Grassi, and Ravazzolo, 2019), and the linear Auto-Regressive (AR(5)) model with lagged values up to lag 5. All models are estimated using Bayesian MCMC methods.

The predictor set includes exchange rates of various fiat currencies, stock and volatility indices, commodities, and cryptocurrency specific variables. Following the tradition in the literature (Gelman, Hwang, and Vehtari, 2014b; Geweke and Amisano, 2010; Bergman and Hansson, 2005), we use the out-of-sample forecasting performance of the aforementioned models to discriminate between the different empirical models. The statistical evaluation of the models is based on the Continuous Rank Probability Score (CRPS) and Mean Squared Forecast Error (MSFE). Finally, to examine if there is an underlying non-linear correlation between the predictors and the return series through the transition probabilities, we add a stochastic search reversible jump step in the NHHM model with the best forecasting performance. We report the posterior probabilities of inclusion in the hidden states transition equations for each predictor.

Our results demonstrate that the 4-states NHHM model has the best forecasting performance for all three series with significant improvements over the single regime models. Even though the optimal model (in terms of predicting accuracy) is common for all three series, we find that the predictors affecting the observed and unobserved processes are significantly different. From the complete set of twelve predictors, the Bitcoin series is affected linearly and non-linearly by five predictors, whereas the Ether and Ripple series are affected by seven predictors. In addition, only the US Treasury Yield and the CBOE stock market volatility index, VIX, have predictive power on all three series. Turning to the in-sample analysis of the 4-state NHHM model, we observe that the returns of each state present distinct characteristics. Unlike conventional exchange rates, see for example Yuan (2011), we observe that the hidden states for all cryptocurrencies are not persistent, but they present frequent alternations. In particular, concerning the Bitcoin return series, we find that state 1, the most frequently occurring state, corresponds to a bear regime (i.e., negative returns and high volatility), states 2 and 3 correspond to a bull regime (positive returns and low volatility) but with different kurtosis and state 4 corresponds to a calm regime (returns close to 0 and low volatility). The relation of the BTC returns and volatility within the hidden states is consistent with the asymmetric volatility theory. Regarding the ETH return series, we observe frequent alternations between state 1, the high volatility state, and the low volatility state 2, while states 3 and 4 serve as auxiliary states with low occupancies. Lastly, state 1 of the Ripple series corresponds to

periods with extremely high average returns but, as a trade-off, also with high risk. States 2 and 3 are the states with the highest occupancies, while state 4 serves as an auxiliary state.

The remainder of this section is structured as follows. Section 4.2 provides an overview of the data and methodology. Specifically, Section 4.2.1 describes the data along with their transformations and descriptive statistics and Section 4.2.2 presents the Hidden Markov models and forecasting evaluation criteria. Section 4.3 presents the empirical findings of the forecasting exercise, i.e., the out-of-sample results and the in-sample analysis of the model with the best performance. Finally, Section 4.4 summarizes and discusses our results.

4.1.1 Other Related Literature

Our model relates to two strands of literature. From a methodological perspective, our model draws from the econometric literature of HM models. Since the seminal work of Hamilton (1989), HM models have been fruitfully applied in diverse areas such as communications engineering and bioinformatics (Cappé, Moulines, and Rydén, 2006). In finance, they have been extensively used in predicting and explaining exchange rates (Panopoulou and Pantelidis, 2015; Lee and Chen, 2006b; Frömmel, MacDonald, and Menkhoff, 2005; Bollen, Gray, and Whaley, 2000), stock market returns (Dias, Vermunt, and Ramos, 2015; Angelidis and Tassaromatis, 2009; Guidolin and Timmermann, 2006), business cycles (Tian and Shen, 2019; Chauvet and Hamilton, 2006), realized volatility (Koki, Meligkotsidou, and Vrontos, 2020; Liu and Maheu, 2018), the behavior of commodities (Pereira, Ramos, and Dias, 2017) and in portfolio allocation (Platanakis, Sakkas, and Sutcliffe, 2019). The reason for their increased popularity is that they present various attractive features. In particular, the time-varying parameters which are driven by the state variable of the presumed underlying Markov process, lead to models that can accommodate both non-linearities and mean reversions (Wu and Zeng, 2014; Guidolin and Timmermann, 2008). In addition, HM models can act as filtering processes that account for outliers and abrupt changes in financial market behavior (Persio] and Frigo, 2016; Ang and Timmermann, 2012) and flexibly approximate general classes of density functions (Timmermann, 2000).

In the cryptocurrency context, HM models have been applied by Hotz-Behofsits, Huber, and Zörner (2018) and Catania, Grassi, and Ravazzolo (2019) as a state space model, by Phillips and Gorse (2017) in the understanding of price bubbles and by

Koutmos (2018) and Koutmos (2019) in examining the relation of BTC with conventional financial assets. Bouri et al., 2019b; Caporale and Zekokh, 2019; Ardia, Bluteau, and Rüede, 2019 use the HM setting to study volatility of cryptocurrencies with GARCH models. The motivation of these studies lies in the observed features of the cryptocurrencies return and volatility series. In particular, cryptocurrencies series are non-stationary and present non-normalities, heteroskedasticity, volatility clustering, heavy tails and excess kurtosis (Katsiampa, 2017b; Katsiampa, 2019a). Chaim and Laurini (2018) and Thies and Molnár (2018) (among others) document the existence of abrupt price changes and outliers, while Corbet and Katsiampa, 2018 show that BTC returns are characterized by an asymmetric mean reverting property. With the aforementioned attractive features of HM models and the characteristics of the cryptocurrency series, it is only natural to ask: do HM models offer improved predictive performance of cryptocurrency returns?

4.2 Data and Methodology

4.2.1 The Data

We use the percentage logarithmic end-of-the-day returns, defined as $y_t = 100 \times (\log(p_t) - \log(p_{t-1}))$, with p_t be the prices of BTC, ETH and XRP. For each coin, we excluded an initial adjustment market period. In particular, we study the BTC time series for the period ranging from 1/2014 until 11/2019, the ETH series for the period ranging from 9/2015 until 11/2019 and the XRP data series from 1/2015 until 11/2019. Figure 4.1 displays the series under study. The covariate set is consisted of normalized fiat currencies, i.e., Euros to US Dollars (EUR/USD), Great Britain Pounds to US Dollars (GBP/USD), Chinese Yuan to US Dollars (CNY/USD) and Japanese Yen to US Dollars (JPY/USD), commodities, i.e., Gold and crude Oil normalized future prices, stock indices, i.e. Standard and Poor's 500 logarithmic returns (SP500), CBOE volatility logarithmic index (VIX), interest rates, i.e., US 10-year Treasury Yield (TY) logarithmic returns and cryptocurrency specific variables, i.e., the blockchain block size (SIZE) as percentage of difference between two consecutive days and the percentage of difference between two consecutive days of Hash Rate (HR)². We report some illustrative descriptive statistics of our covariate set in Table 4.1. Finally, we also include the the lagged 1 autoregressive term of the studied

²The Hash rate is used in BTC and ETH series only

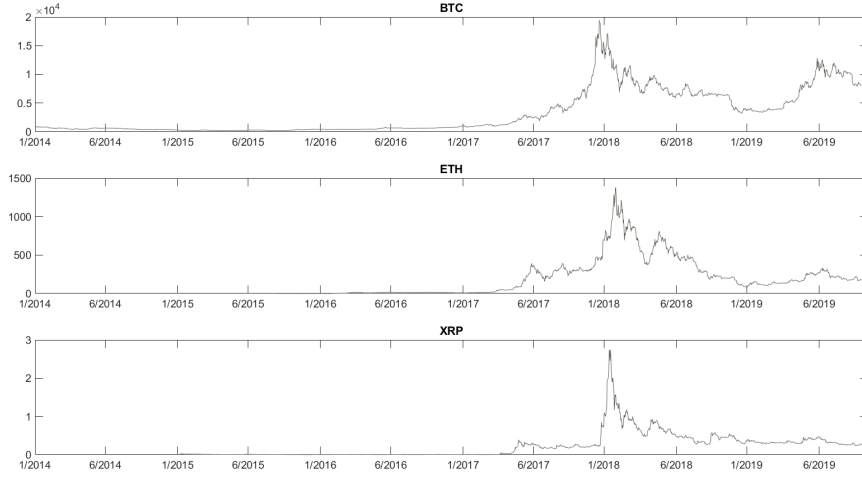


FIGURE 4.1: Daily price series plots for the three cryptocurrencies considered in this study: Bitcoin (upper plot), Ether (middle plot) and Ripple (lower plot). We study the Bitcoin time series for the period ranging from 1/2014 until 11/2019, the Ether series for the period ranging from 9/2015 until 11/2019 and the Ripple data series from 1/2015 until 11/2019.

series as a predictive variables.^{3 4}. In our extensive experimental study, we added up to 5 lagged autoregressive terms as predictive variables. We did not observed any improvement in the performance and hence we omit the results of these experiments.

4.2.2 The Econometric framework

In this study, we focus on a widely used class of econometric models, the HM models. In a HM setting, the probability distribution of the studied series Y_t depends on the state of an unobserved (hidden) discrete Markov process, Z_t . Let (Y_t, X_t) be pair of a the random process of the assumed cryptocurrency return series Y_t , with realization y_t and the set of explanatory variables (predictors) X_t with realization $x_t = (x_{1t}, \dots, x_{kt})$. Then, given the state z_t the observed process is modeled as $y_t = g(z_t)$, with g a predetermined function. The hidden process Z_t follows a first order finite Markov process with $m < \infty$ states and transition probabilities $P(Z_{t+1} = j | Z_t = i) = p_{ij}$, $i, j = 1, \dots, m$. If $m = 1$, then the model is the standard linear regression model.

³The importance of including autoregressive terms is highlighted in Timmermann (2000) who proves that including autoregressive parameters gives rise to cross-product terms that enhance the set of third- and fourth-order moments and the patterns in serial correlation and volatility dynamics that these models can generate and hence provide the basis for very flexible econometric models.

⁴The cryptocurrency price series and blockchain size were downloaded from coinmetrics.io, the exchange rates and commodities prices were downloaded from investing.com, the S&P 500 index, VIX and Treasury Yield from Yahoo Finance and lastly the Hash Rate from quandl.com for BTC and from etherscan.io for ETH.

Variables	Transf	Mean	Std.	Min.	q05	p50	q95	Max.	Kurt	Skew
Bitcoin	% log returns	0.11	3.95	-24.37	-6.47	0.17	6.13	22.47	7.84	-0.27
Ether	% log returns	0.33	6.54	-31.67	-9.76	-0.01	11.61	30.06	6.64	0.07
Ripple	% log returns	0.14	7.02	-63.65	-8.46	-0.31	10.25	100.85	38.99	2.55
EUR/USD	normalized	0	1	-1.42	-1.20	-2.24	2.05	6.30	6.16	1.50
GBP/USD	normalized	0	1	-5.57	-1.65	0.03	1.64	7.23	8.02	0.15
CNY/USD	normalized	0	1	-16.60	-1.52	0.02	1.57	6.27	41.26	-2.00
JPY/USD	normalized	0	1	-9.70	-1.52	0.04	1.44	6.42	15.13	-0.59
Gold	normalized	0	1	-5.32	-1.60	0	1.61	9.32	9.80	0.29
Oil	normalized	0	1	-7.93	-1.22	0	1.29	10.79	23.92	1.04
SP500	log returns	0	0.01	-0.05	-0.01	0	0.01	0.05	9.47	-0.56
VIX	log prices	2.68	0.25	2.21	2.31	2.63	3.15	3.71	3.37	0.70
TY	log returns	0	0.01	-0.10	-0.03	0	0.03	0.11	7.18	0.13
BTC Hash	% of change	-1.12	14.77	-138.86	-25.42	-0.72	20.81	66.80	10.58	-0.88
ETH Hash	% of change	0.42	4.02	-25.50	-4.21	0.37	4.82	99.90	248.92	9.50
BTC size	% of change	-0.82	13.61	-81.26	-23.04	-0.27	20.84	45.08	5.10	-0.46
ETH size	% of change	-0.52	14.63	-359.96	-16.87	-0.15	15.81	55.28	243.51	-10.24
XRP size	% of change	-0.83	13.76	-90.94	-24.63	-0.06	20.43	51.02	6.61	-0.73

TABLE 4.1: Summary Statistics of the percentage logarithmic return cryptocurrency series and transformed predictors. The first column reports the transformation of each variable. The second column displays the mean. The third column reports the standard deviation. Third to seventh columns display the minimum values, the 5%, 50%, 95% quantiles and the maximum values respectively. Last two columns display the kurtosis and skewness coefficients.

We consider the Normal HM models, i.e., conditional on the hidden process marginal distribution of Y_t is \mathcal{N} ormal,

$$y_t = B_{z_t} X_t + \epsilon_{z_t},$$

with $B_{z_t} = (b_{0z_t}, b_{1z_t}, \dots, b_{kz_t})$ be the regression coefficients when the latent process at time t is at state $z_t = s$, $s = 2, \dots, m$, and ϵ_{z_t} be the normally distributed random shocks, $\epsilon_{z_t} \sim \mathcal{N}(0, \sigma_{z_t}^2)$. The hidden process is determined by the transition probability matrix

$$P^{(t)} = \begin{bmatrix} p_{11}^{(t)} & p_{12}^{(t)} & \cdots & p_{1m}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} & \cdots & p_{2m}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}^{(t)} & p_{m2}^{(t)} & \cdots & p_{mm}^{(t)} \end{bmatrix}$$

where $p_{ij}^{(t)} = P(Z_{t+1} = j | Z_t = i)$ is the probability that at time t the hidden state is j given that at time $t - 1$ the hidden state was i . If the transition probabilities are time-constant, then the resulting model is a *Homogeneous Hidden Markov* (HHM) model. However, relaxing the hypothesis of constant probabilities, then the resulting model is the more flexible *Non-Homogeneous Hidden Markov* (NHHM) model. A graphical representation of the NHHM is shown in Figure 4.2.

Among the various NHHM models, we use recently proposed NHHM of Koki,

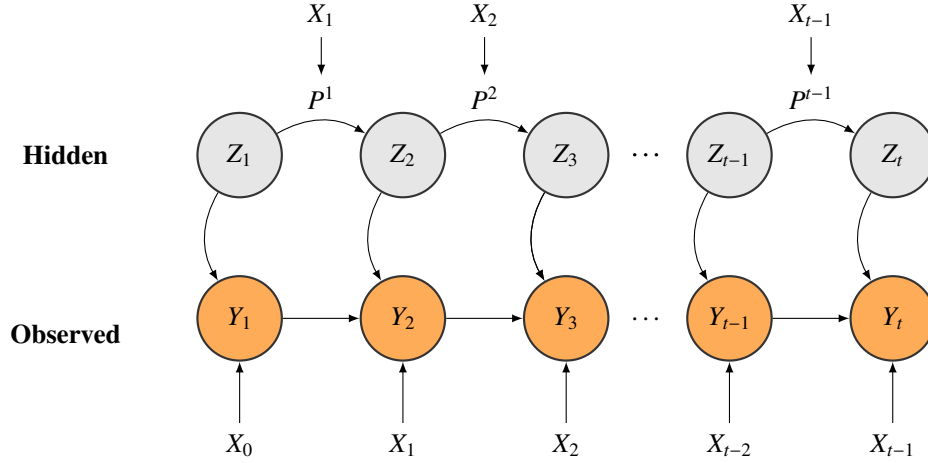


FIGURE 4.2: Graphical representation of the Non-Homogeneous Hidden Markov model.

Meligkotsidou, and Vrontos (2020), namely the Non-Homogeneous Pólya-Gamma HM model. In this model, the transition probabilities are modeled using the multinomial link with predictors X_t and multinomial regression coefficients $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \dots, \beta_{k,ij})'$, i.e.

$$p_{ij}^{(t)} = \frac{\exp(x_t \beta_{ij})}{\sum_{l=1}^m \exp(x_t \beta_{il})}, \quad i, j = 1, \dots, m.$$

In the proposed model, the authors use a further latent variable scheme to make inference on the multinomial regression coefficients, which is based on a Pólya-Gamma data augmentation scheme (Polson, Scott, and Windle, 2013), leading to more accurate and robust inferences.

To make inference on the models' parameters we use a Bayesian Markov Chain Monte Carlo (MCMC) algorithm, which is consisted of the following steps: (a) A FB algorithm for simulating the hidden states, (b) a Gibbs sampling step for estimating the linear regression coefficients using conditional conjugate analysis, (c) a Gibbs sampling step with a Pólya-Gamma data augmentation scheme for estimating the multinomial regression coefficients and (d) simulation of L one-step-look-ahead forecasts. To study the effects of the predictors on the non-homogeneous transition probabilities, we intercalate between the fourth (c) and fifth (d) steps a stochastic variable search (via reversible jump) step. We refer to Koki, Meligkotsidou, and Vrontos (2020) for a detailed description of these steps.

In addition, we include in our analysis the 2-state Markov-Switching Random Walk (MS-RW) model with drift (Engel, 1994b; Nikolsko-Rzhevskyy and Prodan, 2012) and Normal errors, as a simpler and parsimonious regime switching model. By leveraging this model we allow for both the drift term μ and variance σ^2 to take

two distinct values, i.e.,

$$y_t \sim \mathcal{N}(\mu_{z_t}, \sigma_{z_t}^2), z_t = 1, 2,$$

where the variable Z_t is governed by the constant transition probabilities

$$P(Z_t = 1 | Z_{t-1} = 1) = p_{11} \text{ and } P(Z_t = 2 | Z_{t-1} = 2) = p_{22}.$$

Summing up, our methodology is the following. First, we study the performance of various HM settings with fixed covariate set, in explaining and predicting the cryptocurrency log-return series. In particular, we consider 9 HM models, i.e., NHHM models with $m = 2, \dots, 5$ states, HHMs with $m = 2, \dots, 5$ state and the 2-state MS-RW. Following the standard practice, we also implement the Random Walk model (RW) (i.e., a linear model with no covariates), the linear regression model with all the covariates and the autoregressive term, often referred as Kitchen Sink (KS) model and a autoregressive model lagged endogenous variables up to lag 5, AR(5), as single regimes models, leading to 12 total models for each coin, cf. Table 4.2.

Model	Abbreviation	Predictors		Probabilities	States
		Mean equation	Transition Probabilities		
Non-Homogeneous Hidden Markov	(NHHM)	X, AR	X, AR	multinomial	2-5
Homogeneous Hidden Markov	(HHM)	X, AR	—	constant	2-5
Markov Switching Random Walk	(MS-RW)	—	—	constant	2
Kitchen Sink	(KS)	X, AR	—	—	1
Linear Regression	(AR(5))	AR	—	—	1
Random Walk	(RW)	—	—	—	1

TABLE 4.2: Summary of the models of this study. The first two columns show the model and its abbreviation, the third and fourth columns show the assumed relation of the studied time series and the predictors. The fifth column shows the assumed parametrization of the transition probabilities of each model. The last column shows the various number of states we considered for each model.

Then, we choose the model with the out-of-sample (predicting) performance based on the CRPS and MSE. Finally, we focus on the possibly missed hidden effects on the transition probabilities. To this end, we apply a reversible jump stochastic search algorithm on the multinomial regression predictors of the NHHM model with the best predicting performance.

4.2.3 Performance Evaluation

We assess the performance of the studied models based on their predicting ability. Reflecting the logical positivism of the Bayesian approach stating that *a model is as*

good as its predictions Geweke and Amisano (2010) and Guidolin (2011), the predictive accuracy is valued not only for its own sake but rather for comparing different models within the Bayesian framework. Focusing on the accuracy of the predictive density, we rely on two distance-based metrics: the Continuous Rank Probability Score (CRPS) and the Mean Square Error (MSE).

Let y_p be the actual forecasting values with distribution F_p . Utilizing the MCMC output, we obtain a sample of the L one-step-look ahead predictions, $\hat{Y}_l, l = 1, \dots, L$, from the empirical posterior predictive distribution. For every out-of-sample observation, the CRPS is defined as,

$$\text{CRPS}(F_{l,p}, y_{l,p}) = \int_{-\infty}^{\infty} \left(F(\hat{y}_p) - I_{\{\hat{y}_l \geq y_{l,p}\}} \right)^2 d\hat{y}_l, \quad l = 1, \dots, L.$$

We compute the CRPS numerically, using the identity of Székely and Rizzo (2005b)

$$\text{CRPS}(F_{l,p}, y_{l,p}) = -\frac{1}{2} \mathbb{E} |\hat{Y}_l - \hat{Y}'_{l,p}| - \mathbb{E} |\hat{Y}_l - y_{l,p}|,$$

where \hat{Y}_l, \hat{Y}'_l are independent replicates from the estimated (empirical) posterior predictive distribution. The MSE for the l -th, $l = 1, \dots, L$, out-of-sample observation is defined as

$$\text{MSE}_l = \frac{1}{N} \sum_{i=1}^N (y_{p,l} - \hat{y}_{l,i})^2,$$

where N is the MCMC sample size. We report the the CRPS and MSE for every prediction over all MCMC iterations and the average CRPS and MSE over all observations. The best model among its counterparts, is the one with the lowest CRPS and MSE.

4.3 Empirical Analysis

4.3.1 Out-of-Sample analysis

We asses the forecasting performance of the models under scrutiny, i.e., the m -states, $m = 2, \dots, 5$ NHHMs and HHMs, 2-states MS-RW and the benchmarks RW, KS and AR(5). The out-of-sample accuracy is assessed using a sequence of $L = 30$ one-step-ahead predictive densities. In Table 4.3, we report the CRPS and MSE in parenthesis for 5 randomly chosen out-of-sample points, i.e., $L = 1, 2, 7, 15, 30$. The last column reports the average scores over all the out-of-sample points. Through this exercise, parameter estimates are held fixed.

Horizon	Bitcoin					
	1	2	7	15	30	Average
NHHM ₂	0.91 (8.79)	0.45 (8.20)	0.65 (13.30)	0.82 (17.07)	0.64 (14.80)	1.86 (28.80)
NHHM ₃	1.00 (8.47)	0.40 (8.82)	0.68 (14.13)	0.88 (19.42)	0.61 (16.04)	1.85 (29.26)
NHHM ₄	0.71 (7.89)	0.32 (6.37)	0.58 (13.58)	0.78 (15.62)	0.60 (15.54)	1.78 (28.00)
NHHM ₅	0.96 (15.33)	0.60 (14.27)	0.91 (25.92)	0.90 (18.97)	0.62 (15.72)	1.85 (30.41)
HHM ₂	0.91 (9.67)	0.52 (9.33)	0.66 (14.61)	0.86 (17.19)	0.67 (16.00)	1.87 (29.36)
HHM ₃	0.93 (15.60)	0.57 (15.67)	0.78 (16.85)	0.83 (17.73)	0.56 (15.72)	1.87 (30.02)
HHM ₄	0.97 (14.61)	0.62 (13.86)	0.85 (23.06)	0.81 (16.73)	0.53 (14.85)	1.87 (29.43)
HHM ₅	0.91 (14.26)	0.49 (12.92)	0.75 (20.28)	0.86 (15.13)	0.50 (12.73)	1.83 (29.20)
MS-RW	0.95 (9.80)	0.53 (8.93)	0.60 (14.00)	0.81 (16.05)	0.62 (15.50)	1.85 (29.50)
KS	1.10 (17.05)	0.95 (15.67)	0.90 (15.52)	1.09 (19.19)	0.92 (17.90)	1.92 (29.72)
AR(5)	1.12 (17.50)	0.92 (15.35)	0.95 (15.69)	0.99 (16.20)	0.97 (18.12)	1.95 (30.09)
RW	1.20 (19.50)	1.00 (15.85)	1.05 (14.89)	1.00 (16.15)	0.96 (18.85)	1.98 (31.12)
Ether						
NHHM ₂	1.60 (27.82)	1.21 (27.83)	1.11 (29.22)	0.89 (27.71)	0.87 (27.09)	1.62 (31.51)
NHHM ₃	1.59 (27.13)	1.15 (25.40)	1.15 (26.82)	0.85 (25.80)	0.93 (24.74)	1.60 (25.09)
NHHM ₄	1.45 (20.45)	1.04 (18.66)	1.05 (19.21)	0.69 (15.55)	0.70 (15.95)	1.56 (24.87)
NHHM ₅	1.68 (31.57)	1.42 (30.12)	1.26 (31.59)	1.24 (28.08)	1.16 (27.43)	1.83 (37.12)
HHM ₂	1.60 (30.30)	1.29 (33.10)	1.15 (40.16)	1.10 (41.60)	1.07 (43.28)	1.75 (49.25)
HHM ₃	1.67 (44.88)	1.48 (45.83)	1.24 (45.29)	1.06 (41.83)	1.14 (41.41)	1.80 (51.47)
HHM ₄	1.71 (39.94)	1.35 (35.37)	1.22 (34.48)	1.21 (34.20)	1.14 (29.25)	1.81 (40.55)
HHM ₅	1.97 (50.69)	1.74 (46.94)	1.57 (43.16)	1.58 (43.17)	1.56 (44.20)	1.80 (43.42)
MS-RW	1.36 (28.40)	1.20 (28.38)	1.11 (35.90)	1.07 (42.36)	1.22 (43.40)	1.73 (48.86)
KS	1.97 (50.69)	1.74 (46.94)	1.57 (43.16)	1.58 (43.17)	1.56 (44.20)	2.03 (52.00)
AR(5)	1.95 (50.45)	1.82 (47.89)	1.61 (42.79)	1.49 (42.32)	1.62 (43.92)	2.04 (51.92)
RW	2.00 (51.65)	1.86 (48.03)	1.66 (43.87)	1.48 (42.25)	1.66 (45.12)	2.15 (53.03)
Ripple						
NHHM ₂	0.98 (27.38)	1.24 (25.71)	0.81 (22.00)	0.87 (23.57)	0.75 (19.67)	1.54 (29.69)
NHHM ₃	0.90 (26.45)	1.05 (23.91)	0.80 (23.86)	0.69 (22.34)	0.63 (16.92)	1.46 (28.67)
NHHM ₄	0.82 (24.50)	0.99 (22.14)	0.77 (21.70)	0.60 (21.19)	0.62 (18.69)	1.39 (27.38)
NHHM ₅	1.00 (31.50)	1.15 (27.87)	0.95 (25.34)	0.71 (22.38)	0.73 (22.39)	1.60 (30.32)
HHM ₂	0.99 (28.00)	1.22 (24.40)	0.82 (21.88)	0.94 (22.80)	0.81 (23.37)	1.54 (29.16)
HHM ₃	1.22 (37.94)	1.24 (33.45)	0.99 (30.69)	0.89 (29.15)	0.73 (28.56)	1.54 (36.40)
HHM ₄	1.06 (36.42)	1.27 (26.97)	0.86 (27.55)	0.92 (26.35)	0.74 (24.18)	1.52 (32.96)
HHM ₅	1.28 (33.41)	1.10 (29.07)	1.19 (38.27)	0.77 (27.34)	0.85 (27.16)	1.52 (33.68)
MS-RW	1.25 (40.57)	1.09 (34.78)	1.38 (41.25)	0.97 (39.33)	1.01 (39.05)	1.55 (44.93)
KS	2.08 (56.26)	2.00 (54.89)	1.64 (50.50)	1.85 (52.36)	1.66 (48.83)	2.09 (57.64)
AR(5)	1.96 (51.97)	2.05 (56.14)	1.74 (49.13)	1.72 (50.37)	1.67 (47.97)	2.10 (56.75)
RW	1.97 (52.02)	2.06 (56.08)	1.73 (48.45)	1.76 (52.01)	1.72 (50.11)	2.21 (58.73)

TABLE 4.3: Continuous Rank Probability Score and Mean Squared Error in parenthesis for all the competing models for the Bitcoin, Ether and Ripple series. The last column reports the average CRPS (MSE) over the whole sequence of 30 one-step ahead predictions. Bold values indicate the lowest CRPS values for each out-of-sample points.

As a preliminary point, we observe that all the HM models significantly surpass the single regime RW, KS, AR(5) models, while all the single regime models have

similar forecasting performance. Even the KS model, which includes all the predictors besides the autoregressive terms, does not improve the forecasting performance over the AR or RW models. These results suggest that the HM models can identify time-varying parametrizations leading to improved forecasting performance, relatively to the forecasting performance of the single regime benchmarks. This finding is in line with previous studies arguing on the necessity of incorporating the structural breaks and regime switches in modeling the BTC return series, see for example Thies and Molnár (2018). In addition, we also confirm the argument on the existence of time-varying effects on BTC cryptocurrencies series, see for example Meligkotsidou et al. (2019) and expand it to the ETH and XRP series.

As far as the predicting accuracy among the various HM models, we observe that – based on the average CRPS and MSE scores (last column of Table 4.3) – the model with the best forecasting performance is the 4-state NHHM for all coins. More specifically, we observe that the 4-state NHHM model delivers the best predicting performance, since it has the lowest CRPS for all the randomly chosen out-of-sample points, with the exception of the 30-th point in the BTC forecasting exercise and the 1-st point of the ETH forecasting exercise where the 5-state HHM and the MS-RW models have better forecasting accuracy for the two coins respectively. However, by collating the resulting MSEs for each point individually with the CRPS, we observe that in some forecasting horizons different models are found to outperform the best models derived by the CRPS. Even though this might seem contradictory, these differences are expected since the CRPS is more robust to outliers and more reliable when assessing the density forecasts, see also Koki, Meligkotsidou, and Vrontos, 2020 and references therein. Over all coins, the lowest CRPS and the lowest average MSE (best predicting accuracy) are achieved when predicting the BTC return series and the highest CRPS and the highest average MSE (worst predictive accuracy) are achieved when predicting the ETH series.

This forecasting exercise provides empirical evidence that relaxing the hypothesis of constant transition probabilities by allowing the predictors to affect the series non-linearly — through the latent process — improves the forecasting accuracy of the HM models. This is an indication that, besides the conditional time-varying linear correlations between the cryptocurrency return series and predictor set, there exist more complex correlations, such as the underlying non-linear multinomial logistic relationship which can lead to better forecasts.

4.3.2 In-sample analysis of the models with the best predicting performance

Based on the forecasting accuracy, we treat the 4-state NHHMs as our final model for further analysis for the BTC, ETH and XRP return series. Within our MCMC algorithm and at each iteration, we estimate an in-sample realization of the observed data, i.e., we use the in-sample estimations of the parameters and the states to reproduce the cryptocurrency percentage log return series, often referred to as *replicated data* or *within-sample predictions* (Gelman, 2003).

The derived realized distributions along with the observed series for each coin are shown in Figure 4.3. The left column shows the in-sample replicated distributions

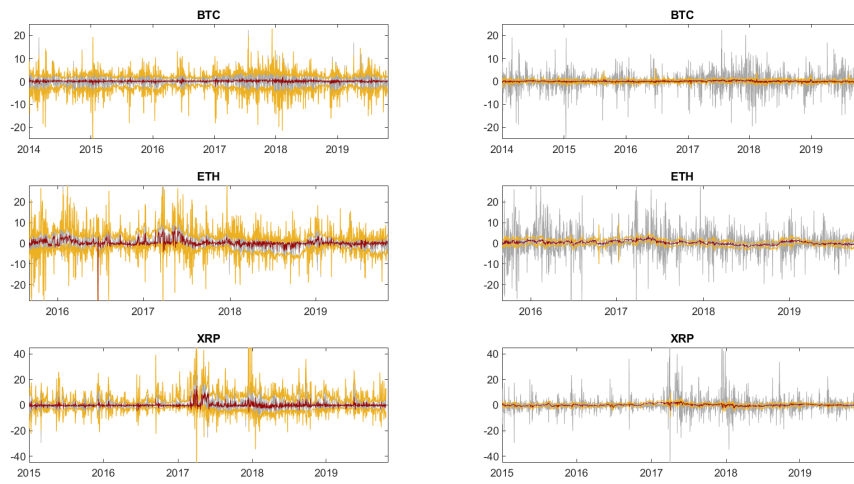


FIGURE 4.3: Percentage return series (gray lines) and quantiles of the posterior sample (replicated) empirical distributions for the Bitcoin series (first row), Ether series (second row) and Ripple series (third row). Yellow lines show the 0.5% and 99.5% quantiles of the estimated in-sample distributions and define the 1% credibility region, whereas red lines show the estimated posterior median. Plots on the left are based on the estimated distributions via the 4-state Non-Homogeneous Hidden Markov (HM) model while plots on the right show the estimated distributions as derived from the Random Walk (RW) benchmark model.

derive using the aforementioned 4-state NHHM and the right column shows the replicated distributions derived using the RW benchmark for all coins. Gray lines show the observed percentage log-return series, yellow lines show the fitted 0.5%-th and 99.5%-th quantiles of the estimated in-sample distribution and the red line shows the 50%-th quantile (median). By visual inspection, we observe that, by identifying the various volatility clusters, the 4-state NHHM models offer substantially improved in-sample performance compared to the in-sample performance of the RW model. The graphical proof of the good in-sample performance of the 4-state NHHM compared

to the RW model is substantiated by Table 4.4 which shows the overall empirical coverage of the estimated quantiles curves. The first and second rows report the proportion of the observed percentage log-returns that fall out of the empirical quantiles curves for the 4-state NHHM, and the in-sample MSE, respectively, for the 4-state NHHM, while the third and fourth rows report the proportion and in-sample MSE for the RW model.

		Bitcoin	Ether	Ripple
NHHM ₄	Proportion	0.05 (121/2114)	0.06 (95/1506)	0.04 (79/1748)
	MSE	15.18	37.13	43.22
RW	Proportion	0.75 (1601/2114)	0.74 (1114/1506)	0.72 (1255/1748)
	MSE	17.03	42.28	50.01

TABLE 4.4: Empirical coverage of the empirical in-sample distributions using the 4-state NHHM and the RW benchmark for the Bitcoin, Ether and Ripple percentage return series.

Table 4.5 provides a gauge of what drives the documented predictability by showing the posterior mean estimates for the linear regression predictors for each state for BTC, ETH and XRP respectively. Predictors that fall into the 10% credibility intervals are marked with asterisk. In addition to the linear regression estimates, the last column of Table 4.5 shows the posterior probabilities of inclusion for the predictors affecting the transition probabilities, as derived from the stochastic search algorithm on the multinomial regression coefficients. Posterior probabilities of inclusion exceeding 0.4 are marked with bold fonts.

We observe that the majority of the predictors are not statistically significant in the linear regression parametrization, especially for the BTC return series. At this point, it is important to stress that even when the coefficient of an explanatory variable is not statistically different from zero, this does not necessarily mean that the variable has no predictive power for return series, see also Panopoulou and Pantelidis (2015). It is often the case that a variable that is insignificant in-sample has predictive out-of-sample power and vice versa. This argument is strengthened by our extensive experimental study — results not reported here — which shows that if we remove the insignificant predictors, the forecasting accuracy deteriorates. Furthermore, we observe that depending on the hidden state the mean posterior estimates can be markedly different, even change their sign. Regarding the predictors affecting the transition probabilities, we observe that there exist predictors that affect the transition probabilities with high posterior probabilities of inclusion, with the volatility

BTC					
Predictors	State 1	State 2	State 3	State 4	Probabilities
Intercept	11.47	6.65	−0.80	0.02	0.00
EUR/USD	1.02	0.43	−0.15	−1.59	0.00
GBP/USD	0.06	0.17	−0.04*	−1.26	0.00
CNY/USD	−0.14	−0.12	−0.15	−1.30	0.00
JPY/USD	−0.37*	0.08	−0.16	−0.78	0.00
Gold	−0.09	−0.12	0.07	0.01	0.00
Oil	0.02	0.14	0.01	0.52	0.00
SP500	−6.73	−0.85	0.32	0.08	0.00
VIX	−4.37	−2.34	0.30	0.03	1.00
TY	−2.07	−2.21	1.49	0.15	0.90
Size	−0.01	−0.01	0.01	0.13	0.00
Hash	−0.01	−0.01	0.00	0.26	0.00
AR(1)	−0.04	0.00	0.03	0.94*	0.00
ETH					
Intercept	7.68	1.93	0.85	0.80	1.00
EUR/USD	−0.14	0.10	0.64	0.81	1.00
GBP/USD	1.78	0.51*	2.35	2.14	0.00
CNY/USD	−0.40	−0.24	−0.18	0.31	0.00
JPY/USD	−0.14	0.31	1.49	1.54	0.00
Gold	−0.12	0.32*	1.17	1.26	0.82
Oil	−2.02*	0.26*	0.24	0.42	1.00
SP500	−9.14	−3.52	1.62	1.91	0.11
VIX	−2.54*	−0.78*	−1.01	−0.78	1.00
TY	−13.45	3.75	1.45	0.31	0.84
Size	0.07	0.21	0.10	0.03	0.00
Hash	0.03	0.00	0.17	0.12	0.00
AR(1)	−0.06	−0.17*	0.74	0.82	0.00
XRP					
Intercept	9.30	0.25	0.87	0.96	1.00
EUR/USD	−1.88*	−0.19	0.25*	0.07	1.00
GBP/USD	−4.03	−0.04	−0.25	−0.85	0.01
CNY/USD	3.31	0.26	0.13	0.48	0.02
JPY/USD	1.36	1.13	0.02	2.12	0.70
Gold	−1.44	−0.10	−0.01	−0.62	0.10
Oil	0.44	−0.07	−0.09	0.17*	0.06
SP500	2.00	−7.06	−0.87	0.13	0.40
VIX	−1.88	−0.19	0.25	−0.85	1.00
TY	−22.46	1.29	3.68	−0.08	0.52
Size	0.07	0.21	0.10	0.03	0.00
AR(1)	0.44	−0.08	−0.10*	0.18	0.00

TABLE 4.5: Posterior means estimates of the 4-state Non-Homogeneous Hidden Markov model for the Bitcoin, Ether Ripple percentage return series. The first column specifies the predictors. The second, third, fourth and fifth columns report the posterior mean estimates for each predictor at the first, second and third states respectively. The last row reports the mean estimated residual variance for each state. The last column reports the posterior probabilities of inclusion for the predictors affecting the transition probabilities' multinomial regression model. These probabilities are calculated by applying a stochastic search reversible jump algorithm within the MCMC scheme. Statistical significance at the 10% level is denoted with * and posterior probabilities exceeding 0.4 are marked with bold fonts.

index VIX and Treasury Yield (TY) affecting the transition probabilities for all the cryptocurrency return series.

4.3.3 Hidden States classification and interpretation

Table 4.6 provides information on the hidden states for each coin; that is, the states' occupancies as the average time spent at each state i , $i = 1, \dots, 4$, the average returns and the corresponding standard deviation. At a first glance, the hidden process identifies periods with different underlying volatilities for every coin, i.e., periods with high and low volatilities. In more detail, for the BTC series, it identifies periods with negative average returns and high volatilities (state 1), periods with positive returns and low volatility (states 2 and 3) and calm periods with average returns close to zero and very low volatility. This segmentation in the return series resembles the bear/turbulent (state 1) and bull (states 2 and 3) markets, while state 4 corresponds to a stable/calm regime. Furthermore, we observe that the states 2 and 3 have similar (almost equal) average returns. The similar average returns and different volatility indicate that the hidden process segments the return series into two subseries with the same skewness but very different kurtosis, see Timmermann (2000).

Concerning the ETH series, we observe that the highest mean returns occur in the state with the highest volatility. The hidden process alternates between a high volatility and a low volatility regime with almost zero average returns — states 1 and 2 respectively — for the 95% of the overall time, while states 3 and 4 serve as auxiliary states with almost equal average returns but different volatilities. Finally, the hidden process in the XRP series spends most of the time (80%) in the high volatility regime 2 and the low volatility regime 3. We also observe that state 1 has extremely high average returns compared to the returns of states 2 and 3 but is associated with very high risk (high volatility) as a trade-off. Lastly, hidden state 4 serves as an auxiliary state with low occupancy, capturing the extreme values (outliers) of the returns series. The information in Table 4.6 can be visualized in Figures 4.4 and 4.5 which depict the state-switching dynamics of the three cryptocurrencies according to their hidden state classification. Figure 4.4 illustrates the Bitcoin (upper plot), Ether (middle plot) and Ripple (lower plot) returns conditionally on the a realization of the hidden process using the 4-state NHHM model. Gray lines correspond to the percentage log returns, while red, yellow, purple and green dots correspond to the times that the hidden state was in states 1,2,3 and 4, respectively. These graphical representations serve as an easy way to visualize the evolution of the hidden process in

Coin	State	Occupancies	Average	Std
BTC	1	0.35	-0.46	5.91
	2	0.30	0.49	2.49
	3	0.24	0.47	1.69
	4	0.11	0.02	0.53
ETH	1	0.41	0.88	8.90
	2	0.54	-0.06	2.61
	3	0.02	0.16	1.13
	4	0.02	0.18	0.76
XRP	1	0.17	3.59	14.98
	2	0.35	0.76	4.46
	3	0.45	0.41	1.77
	4	0.03	6.82	0.86

TABLE 4.6: Information on the states as derived from the experiments on the BTC, ETH, XRP return series. First column reports the cryptocurrencies and the second column the different regimes. The third column reports the states' occupancies, i.e., the average time spend at each regime. The fourth column reports the average returns at each state and finally, the fifth column reports the state's estimated standard deviation.

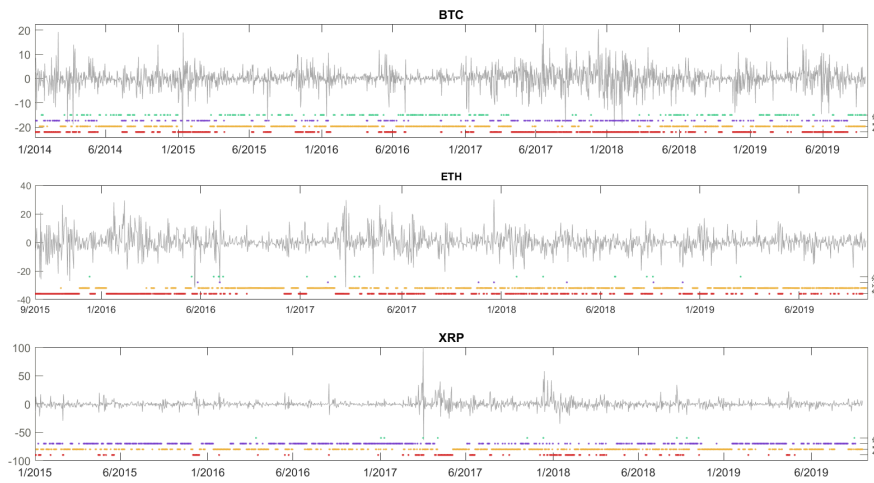


FIGURE 4.4: Bitcoin (upper plot), Ether (middle plot) and Ripple (lower plot) percentage return series conditionally on a realization of the hidden process. The hidden process is estimated using the 4-state NHHM on the aforementioned cryptocurrencies' percentage return series. Red, yellow, purple and green dots indicate the time periods that the hidden process is at states 1, 2, 3 and 4 respectively.

reference with the returns for each coin. While frequent alternations between the hidden states are prevalent in all three time-series, the transitional patterns are markedly different. For instance, in the BTC return series, there exist frequent alternations between states 1 and 3 and between 2 and 4, while in the ETH and XRP series they are between states 1 and 2 and between 2 and 3, respectively.

Figure 4.5 shows the estimated ex ante smoothed probabilities of each state for each time period, i.e., $P(Z_t = m | y_1, \dots, y_T)$, $m = 1, \dots, 4$, over all coins. We ob-

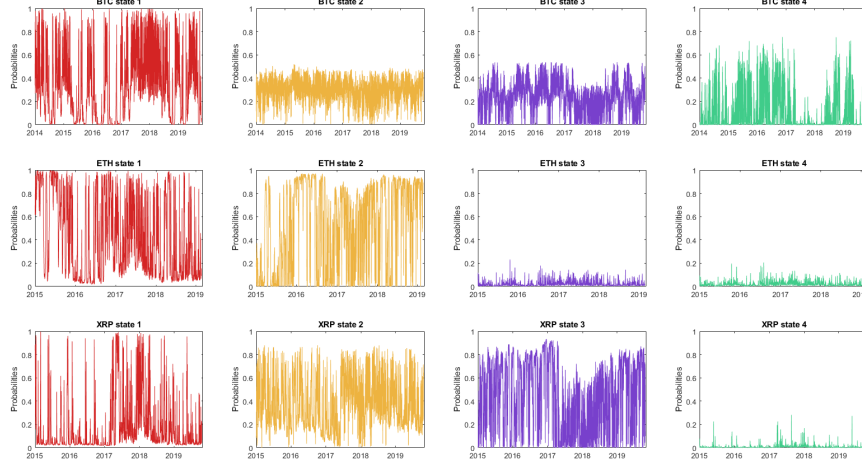


FIGURE 4.5: Smoothed probabilities of being in state m , i.e., $P(Z_t = m | y_1, \dots, y_T)$, $m = 1, \dots, 4$ using the 4-state Non-Homogeneous Hidden Markov model. Columns 1 to 4 correspond to states 1 to 4 while rows 1 to 3 correspond to Bitcoin, Ether and Ripple, respectively.

serve that hidden state 1 for the BTC series, hidden states 1 and 2 for the ETH series and hidden states 1, 2 and 3 for the Ripple series occur with high probability. The identification of these particular hidden states is sound with low probabilities of misclassification. However, we get mixed insights on the occurrence of the other states which are neither high nor low. A measure of the persistence of the underlying state densities are the so-called *persistent probabilities*, i.e., the probabilities of remaining at the same state $p_{ii} = P(Z_t = i | Z_{t-1} = i)$, $i = 1, \dots, 4$, see for example Timmermann (2000) and Yuan (2011). Figure 4.6 shows the persistent probabilities for BTC (upper plot), ETH (middle plot) and XRP (lower plot). Columns 1 to 4 correspond to states 1 to 4. The higher the probabilities are the longer the process is expected to remain at state i and hence transitions to different states are less frequent.

4.4 Discussion

In this work, we modeled the return series of the three largest (in terms of market capitalization) cryptocurrencies, Bitcoin, Ether and Ripple using a Hidden Markov framework. We employed a multi-state Bayesian Hidden Markov methodology with a predefined set of financial and cryptocurrency specific predictors to capture the time-varying characteristics and heteroskedasticity of the cryptocurrencies' return

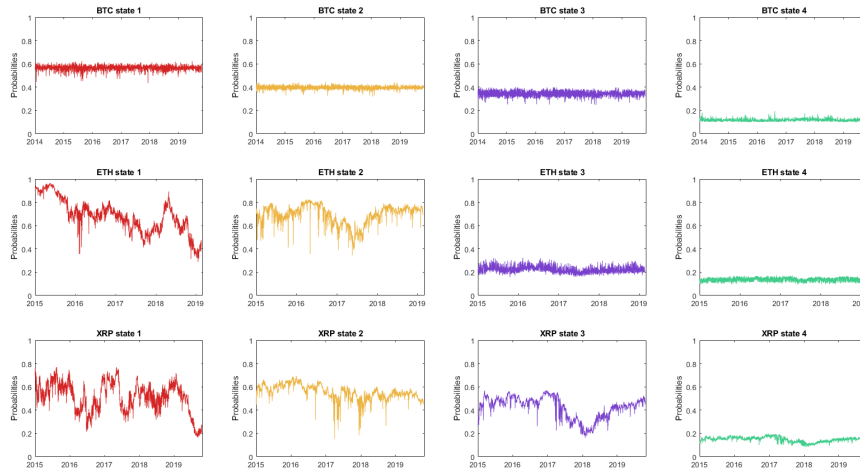


FIGURE 4.6: Persistence probabilities for each regime, based on the mean posterior transition probabilities $p_{ii} = P(Z_t = i | Z_{t-1} = i)$, $i = 1, \dots, 4$. The rows 1 to 3 corresponds to the Bitcoin, Ether and XRP return series respectively and columns 1 to 4 correspond to the states 1 to 4 respectively.

series. The employed methodology is motivated by the existing evidence of structural breaks and regime/states switches in the cryptocurrency series. In line with the literature, we chose the best model among 9 different Hidden Markov models — the standard Markov-Switching Random Walk (MS-RW) model, the Homogeneous Hidden Markov (HHM) and the Non-Homogeneous Hidden Markov (NHHM) models with up to five hidden states — and 3 single regime models — the Random Walk (RW) model, the linear AutoRegressive (AR) model and the Kitchen Sink (KS) model — based on their out-of-sample predictive ability.

The out-of-sample forecasting exercise indicated that the 4-states NHHM model has the best forecasting performance for all three series with significant improvements over the single regime models. In addition, the regime switches reveal time-varying connections of traditional financial, economic and cryptocurrency specific predictors with cryptocurrency returns. In particular, we identified predictors affecting the series linearly, i.e., through the observed process, and non-linearly, i.e., through the unobserved process. From the complete set of twelve predictors, Bitcoin series is affected linearly and non-linearly by five predictors, while the Ether and Ripple series are affected by seven predictors. In general, only the US Treasury Yield and the CBOE stock market volatility index, VIX, have predictive power on all three series. Turning to the in-sample analysis, the 4-states NHHM model segments the return series into four subseries with distinct state-switching dynamics and economic interpretation. For Bitcoin, we find that the most frequently occurring state (state 1)

corresponds to a bear regime (i.e., negative returns and high volatility), states 2 and 3 correspond to bull regimes (positive returns and low volatility) but with different kurtosis and state 4 corresponds to a calm regime (returns close to 0 and low volatility). Regarding the Ether return series, we observe frequent alternations between the high and low volatility states (states 1 and 2), while states 3 and 4 serve as auxiliary states with low occupancies. Finally, state 1 of the Ripple series corresponds to periods with extremely high average returns but, as a trade off, also with high risk. States 2 and 3 are the states with the highest occupancies, while state 4 serves as an auxiliary state.

Our empirical findings demonstrate that multi-state Non-Homogeneous Hidden Markov models offer improved forecasts on all considered cryptocurrency return series. In addition, they provide evidence to support the existence of predictors with state-dependent, time-varying predicting power on the cryptocurrency series. These insights are particularly useful to investors, portfolio-managers and policy-makers. Importantly, they refine our understanding of the dynamic relationship between traditional financial markets and cryptocurrency returns, which is characterized by regime switches and frequent alternations. Thus, when taken into account, these dynamics can be leveraged to optimize investors' decisions regarding portfolio allocation and risk-diversification between conventional and cryptocurrency assets. Moreover, along with the absence of predictors that systematically predict cryptocurrency returns in all possible states, the frequent alternations suggest that interested parties ought to be cautious when using forecasts to inform their decisions. Finally, the study of regimes with markedly different economic characteristics and transition dynamics is important for regulatory authorities and cryptocurrency entrepreneurs. Specifically, this information should be used to improve the implementation of policies that seek to mitigate risks associated with cryptocurrency exchange rates and promote their wider public adoption.

Chapter 5

The X-ray luminosity function of Active Galactic Nuclei in the redshift interval $z = 3 - 5$

5.1 Introduction

In recent years observations have established that supermassive black holes (SMBHs) are nearly ubiquitous in local spheroids Magorrian et al. (1998) and Kormendy and Ho (2013). These relic black holes are believed to have grown their masses at earlier times mostly via accretion of material from larger scales (e.g. Soltan, 1982; Marconi et al., 2004). Questions that remain open are when during the lifetime of the Universe these events occurred and under what physical conditions black holes grow their masses. Moreover, observations show that in the local Universe correlations exist between the mass of SMBHs and the properties of the stellar component of the bulges in which they reside, such as velocity dispersion (Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Gültekin et al., 2009; Graham et al., 2011), luminosity (McLure and Dunlop, 2002; Marconi and Hunt, 2003; Gültekin et al., 2009), dynamical mass (Magorrian et al., 1998; Marconi and Hunt, 2003; Häring and Rix, 2004; Graham, 2012) and central light concentration (Graham et al., 2001; Savorgnan et al., 2013). Such correlations suggest a link between the growth of black holes and the formation of their host galaxies, although the exact nature of such an interplay is still not well understood. Processes that can establish such correlations include a common gas reservoir that both feeds the central black hole and forms stars on larger scales, outflows related to the energy output from the AGN itself that affect the Inter-Stellar Medium and regulate the formation of stars (Silk and Rees, 1998; Fabian, 1999; King, 2003; King, 2005; Di Matteo, Springel, and Hernquist, 2007; Croton et al., 2006), or the merging history of galaxies and heir SMBHs (e.g. Jahnke

and Macciò, 2011).

One approach for improving our understanding of the formation of SMBHs as a function of cosmic time and their relation to their host galaxies is population studies of Active Galactic Nuclei (AGN), which signpost accretion events onto SMBHs. This requires a census of the AGN population across redshift to constrain for example, the accretion history of the Universe or study the incidence of active black holes among galaxies. In that respect, the AGN luminosity function, i.e. their comoving space density as a function of redshift and accretion luminosity, is one of the fundamental quantities that characterise the demographics of active black holes. The cosmic evolution of AGN leaves imprints on the shape and overall normalisation of the luminosity function. The total mass density locked into black holes at different epochs can be inferred by direct integration of the AGN luminosity function, under assumptions about the radiative efficiency of the accretion process and after applying appropriate bolometric luminosity corrections (Marconi et al., 2004; Aird et al., 2010; Ueda et al., 2014). The space density of AGN split by host galaxy properties, such as stellar mass, morphology or level of star-formation, provides clues on the interplay between black hole accretion and galaxy evolution (Georgakakis et al., 2009; Georgakakis et al., 2011; Aird et al., 2012; Bongiorno et al., 2012; Georgakakis et al., 2014).

Selection at ultraviolet (UV)/optical wavelengths (e.g. Richards et al., 2008; Ross et al., 2012) currently provides the largest spectroscopic AGN samples for luminosity function calculations (e.g. Ross et al., 2013). The downside is that the UV/optical continuum of AGN is sensitive to dust extinction along the line-of-sight and dilution by the host galaxy at faint accretion luminosities. Observations at X-ray wavelengths can mitigate these issues (e.g. Brandt and Alexander, 2015). X-ray photons, particularly at rest-frame energies $> 2 \text{ keV}$, can penetrate nearly unaffected large columns of intervening gas and dust clouds ($N_H \gtrsim 10^{22} \text{ cm}^{-2}$), thereby providing samples least affected by obscuration biases. Moreover, the X-ray emission associated with stellar processes is typically 2 orders of magnitude fainter than the AGN radiative output and therefore contamination or dilution effects by the host galaxy are negligible at X-rays over a wide baseline of accretion luminosities. X-ray surveys also benefit from a well defined selection function that is relatively easy to quantify and account for in the analysis. The disadvantage of X-ray selection is that the detected AGN are often optically faint and therefore spectroscopic follow-up observations are often expensive. Nevertheless, intensive multiwavelength campaigns in recent years

substantially increased the number of X-ray survey fields with sufficient quality ancillary data for reliable X-ray source identification and redshift measurements using either spectroscopy or photometric methods. Early results by Cowie et al. (2003) on the redshift evolution of the X-ray AGN space density and Ueda et al. (2003) on the hard band (2-10 keV) X-ray luminosity function and obscuration distribution of AGN have been expanded recently both in terms of data and analysis methodology (Yencho et al., 2009; Ebrero, J. et al., 2009; Aird et al., 2010; Burlon et al., 2011; Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015; Miyaji et al., 2015). Although important details on the shape of the X-ray luminosity function and the obscuration distribution of AGN are still debated (Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015; Miyaji et al., 2015), the overall evolution of the X-ray luminosity function is reasonably well constrained at least to $z \approx 3$. The initial increase of the AGN X-ray luminosity density from $z = 0$ to $z \approx 1.5$ is followed by a broad plateau up to $z \approx 2.5-3$ and a decline at higher redshift. However, the amplitude of the X-ray AGN evolution at $z \gtrsim 3$ is still not well constrained. Early studies suggested a moderate decline of the AGN space density at $z > 3$ (Yencho et al., 2009; Ebrero, J. et al., 2009; Aird et al., 2010), contrary to claims for a rapid drop (Brusa, M. et al., 2009; Civano et al., 2011; Vito et al., 2013; Kalfountzou et al., 2014) that can be parametrised by an exponential law in redshift (Gilli, R., Comastri, A., and Hasinger, G., 2007) similar to the optical QSO space density evolution (Schmidt, Schneider, and Gunn, 1995; Richards et al., 2006). Central to this debate is the typically small X-ray AGN sample sizes at $z > 3$. For example, there are 209 and 141 X-ray AGN with spectroscopic or photometric redshifts above $z = 3$ in the most recent compilations of Kalfountzou et al. (2014) and Vito et al. (2014) respectively. These numbers should be compared with sample sizes of few thousands AGN at $z < 3$ for the most recent X-ray luminosity function studies (e.g. Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015; Miyaji et al., 2015). Better constraints on the form and amplitude of the evolution of X-ray AGN at $z > 3$ will have implications for the contribution of this population to the UV photon field density that is needed to keep the Universe ionised at high redshift. Haardt and Madau (2012) used the Ueda et al. (2003) X-ray luminosity function to predict a moderate contribution of AGN to the hydrogen ionising radiation field at $z > 3$, in broad agreement with constraints derived from UV/optical selected QSO luminosity functions (e.g. Fontanot, F. et al., 2007; Masters et al., 2012; McGreer et al., 2013) and other X-ray AGN studies (Barger et al., 2003b; Grissom, R. L., Ballantyne, D. R., and Wise, J. H., 2014). There are also claims however, that AGN provide an important contribution to the photoionisation

rate at high redshift (Fiore, F. et al., 2012; Glikman et al., 2011; Giallongo, E. et al., 2015). This discrepancy emphasises the need for further work to improve measurements of the AGN space density at high redshift and to better understand their role in the re-ionisation of the Universe.

In this chapter, we combine deep *Chandra* and wide-area/shallow *XMM-Newton* survey fields to compile one of the largest samples of X-ray selected AGN at $z = 3 - 5$ to date. A Bayesian methodology is developed to correctly account for photometric redshift uncertainties and to determine in a non-parametric way the AGN comoving space density in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$ and over 3 decades in X-ray luminosity [$\log L_X(2 - 10 \text{ keV}) \approx 43 - 46$ in erg s^{-1}]. Although parametric models are also fit to the data, we emphasise the importance of non-parametric estimates to determine in a model-independent way the shape and overall evolution of the X-ray luminosity function. Throughout this paper we adopt $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$.

5.2 Data

For the determination of the X-ray luminosity function in the redshift interval $z = 3 - 5$ we combine *Chandra* and *XMM-Newton* X-ray surveys with different characteristics in terms of area coverage and X-ray depth. These are the 4 Ms *Chandra* Deep Field South (CDFS; Xue et al., 2011; Rangel et al., 2013), the 2 Ms *Chandra* Deep Field North (CDFN; Alexander et al., 2003; Rangel et al., 2013), the Extended Groth Strip International Survey field (AEGIS, Davis et al., 2007; Laird et al., 2009; Nandra et al., 2015), the *Extended Chandra Deep Field- South* (ECDFS; Lehmer et al., 2005), the *Chandra* Cosmological Evolution Survey (C-COSMOS) field (C-COSMOS, Elvis et al., 2009) and the equatorial field of the *XMM-XXL* survey.

5.2.1 Chandra survey fields

The *Chandra* observations of the CDFS, CDFN, AEGIS, ECDFS and C-COSMOS were analysed in a homogeneous way by applying the reduction and source detection methodology described by Laird et al. (2009). Specific details on the analysis of the 4 Ms CDFS and the 2 Ms CDFN fields are presented by Rangel et al. (2013). The *Chandra* survey of the AEGIS field has two tiles. The wide and shallow one (AEGIS-W) consists of 8 *Chandra* pointings of 200 ks each. These data are described by Laird et al. (2009). The deep survey of the AEGIS field (AEGIS-XD) increased to a total of 800 ks the exposure time of the central regions of the AEGIS-W. The additional data

TABLE 5.1: Number of X-ray sources in the full-band selected sample

field	solid angle (arcmin ²)	Number of X-ray sources	Number of optical/infrared IDs	photometric redshift full sample	photometric redshift 3 < z < 4	photometric redshift 4 < z < 5	full sample	spectroscopic redshift 3 < z < 4	spectroscopic redshift 4 < z < 5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
4Ms CDFS	271.4	422	418	151	16	3	268	9	1
CDFN	412.9	453	436	171	11	5	265	9	2
ECDFS	643.0	407	399	267	30	11	134	3	1
AEGIS-XD	934.6	818	804	478	37	13	326	16	0
AEGIS-W	724.3	415	413	267	26	8	146	2	1
COSMOS	3258.0	1435	1393	635	38	14	758	21	4
XMM-XXL	64728.0	7493	3798	—	—	—	2553	54	5
Total	70972.2	—	—	—	158	54	—	114	14

TABLE 5.2: (1) Name of the X-ray survey fields used in this work. The last row lists the total number of sources for the combined fields. For that last row only the columns that correspond to the total number of sources with photometric or spectroscopic redshift in the intervals $3 < z < 4$ and $4 < z < 5$ are listed. This is because these are the relevant numbers to the analysis presented in this paper. (2) Solid angle of each sample in square arcminutes after excluding regions of poor/no photometry. (3) Total number of X-ray sources detected in the 0.5-7 keV energy band in the case of the *Chandra*, or the 0.5-8 keV band for the XMM-XXL sample. (4) Number of full-band detected X-ray sources with optical/infrared associations. (5) Number of sources with photometric redshift estimates in each survey field. Photometric redshifts for the XMM-XXL are not available. (6) Number of sources with photometric redshifts in the interval $z = 3 - 4$. This the sum of the photometric redshift Probability Distribution Functions rounded to the nearest integer. No photometric redshift estimates are available for the XMM-XXL sample. (7) Same as column (6) but for the redshift interval $z = 4 - 5$. (8) Number of sources with spectroscopic redshift estimates in each survey field. (9) Number of sources with spectroscopic redshifts in the interval $z = 3 - 4$. (10) Number of sources with spectroscopic redshifts in the interval $z = 4 - 5$.

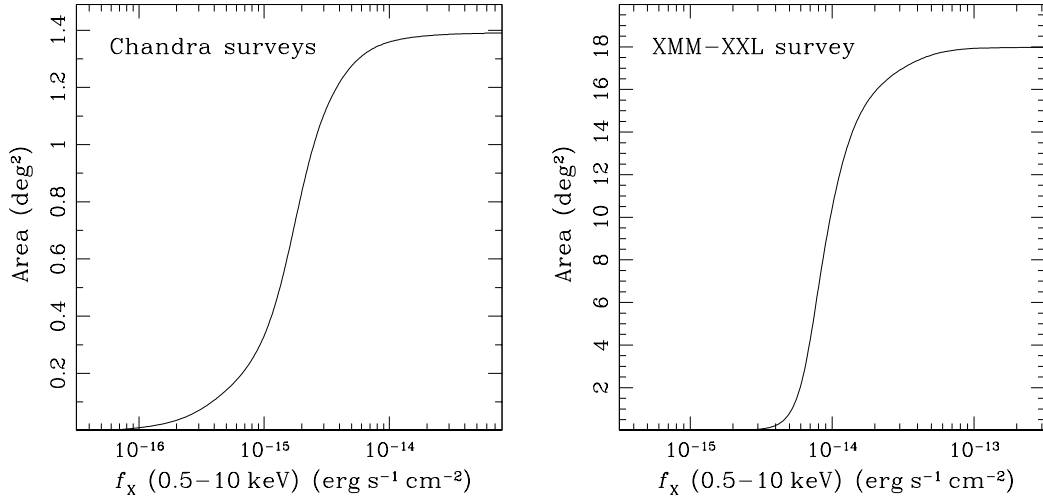


FIGURE 5.1: X-ray Sensitivity curves for the combined Chandra surveys used in the analysis (left panel) and the XMM-XXL survey (right panel).

overlap with the central 3 of the original 8 Chandra pointings observed as part of the AEGIS-W. The AEGIS-XD survey data reduction and source catalogue generation are described by Nandra et al. (2015).

The X-ray sources used in this paper are detected in the 0.5–7 keV (full) spectral band with Poisson false detection threshold $< 4 \times 10^{-6}$. The count rates in the 0.5–7 keV band are converted to fluxes in the 0.5–10 keV band assuming a power-law spectral index with $\Gamma = 1.9$. This is steeper than the $\Gamma = 1.4$ adopted for the X-ray flux estimation in the published catalogues of the CDFS, CDFN and AEGIS fields (Laird et al., 2009; Rangel et al., 2013; Nandra et al., 2015). The choice of the $\Gamma = 1.9$ is motivated by the fact that at high redshift, $z \gtrsim 3$, the observer-frame 0.5–7 keV band corresponds to harder rest-frame energies, which are least affected by obscuration. A diagnostic of the spectral shape of X-ray sources is their hardness ratio defined as $HR = (H - S) / (H + S)$, where S , H are the observed count rates in the 0.5–2 and 2–7 keV spectral bands, respectively. For sources in the range $z = 3 - 5$ we find a median hardness ratio $HR \approx -0.2$. This is consistent with a power-law X-ray spectrum with $\Gamma \approx 1.8$, i.e. similar to the mean spectral index of the intrinsic AGN spectra ($\Gamma \approx 1.9$, Nandra and Pounds, 1994). We therefore choose to fix $\Gamma = 1.9$ for the determination of fluxes. We note however, that the choice of Γ (1.4 vs 1.9) has a small impact on the results presented in this paper.

Sensitivity curves, which measure the total survey area that is sensitive to sources of a particular flux are calculated following methods described in Georgakakis et al. (2008). The overlap between the ECDFS and the 4 Ms CDFS or the AEGIS-W and the AEGIS-XD is accounted for by defining independent spatial regions for each

survey. Spatial masks that describe both the boundaries of the optical/infrared imaging of each field and regions of poor photometry because of bright stars (Aird et al., 2015) are also taken into account in the X-ray sensitivity calculations. The sensitivity curves in the 0.5-10 keV band are presented in Figure 5.1. Table 5.1 presents the number of X-ray sources in each field. The same spatial masks used for the construction of sensitivity maps are used to filter the X-ray source catalogue.

The optical identification of the X-ray sources in the CDFN, AEGIS-XD, AEGIS-W, ECDFS and C-COSMOS fields are based on the Likelihood Ratio method (Sutherland and Saunders, 1992) as implemented in Aird et al. (2015). The multiwavelength associations of the 4 Ms CDFS X-ray sources are presented by Hsu et al. (2014). They apply a Bayesian methodology, based on the work of Budavári and Szalay (2008), to different catalogues available in that field including the CANDELS/*H*-band selected photometry presented by Guo et al. (2013), the Taiwan ECDFS Near-Infrared Survey (TENIS; Hsieh et al., 2012) and the MUSYC/*BVR*-selected catalog of Cardamone et al. (2010). The number of X-ray sources with secure optical or infrared counterparts in each field are presented in Table 5.1.

Extensive spectroscopic campaigns have been carried out in the fields of choice. For the CDFN, ECDFS and AEGIS-W we use the compilation of spectroscopic redshifts presented by Aird et al. (2015). In the 4 Ms CDFS we use the spectroscopic redshifts compiled by Hsu et al. (2014). In the case of AEGIS-XD we use the spectroscopic redshift catalogue presented by Nandra et al. (2015). Redshifts in the C-COSMOS are from the public releases of the VIMOS/zCOSMOS bright project Lilly et al. (2009) and the Magellan/IMACS observation campaigns (Trump et al., 2009), as well as the compilation of redshifts for X-ray sources presented by Civano et al. (2012). The spectroscopic redshifts used in this paper have quality flags in the published catalogues from which they were retrieved that indicate a probability better than $\approx 95\%$ of being correct.

For X-ray sources without spectroscopy, photometric redshifts are estimated using the multiwavelength photometric catalogues available for each survey field. The photometric redshifts of the X-ray sources in the 4 Ms CDFS, the AEGIS-XD and the COSMOS fields are determined following the methodology described by Salvato et al. (2009) and Salvato et al. (2011). Specific details can be found in Hsu et al. (2014); 4 Ms CDFS, Nandra et al. (2015); AEGIS-XD and Salvato et al. (2011); COSMOS field. The estimated rms scatter of the X-ray AGN photometric redshifts is $\sigma_{\Delta z/(1+z)} = 0.016$, 0.014 and 0.04 for the C-COSMOS, 4 Ms CDFS and AEGIS-XD samples, respectively. The corresponding outlier fraction, defined as

$\Delta z/(1+z) > 0.15$, is about 5-6% in all three fields. In the case of ECDFS, AEGIS-W and CDFN we use the photometric redshifts estimated by Aird et al. (2015). For these fields $\sigma_{\Delta z/(1+z)} = 0.06$ and the outlier fraction is about 15%. The latter value is larger than in the C-COSMOS, 4 Ms CDFS and AEGIS-XD fields. This is related to differences in the methodology of estimating photometric redshifts and ultimately to the choice of template SEDs used in the calculation. Nevertheless, the photometric redshifts estimated by Aird et al. (2015) are assigned appropriately larger uncertainties, approximated by the corresponding Probability Distribution Functions (PDZ), that reflect the higher outlier fraction.

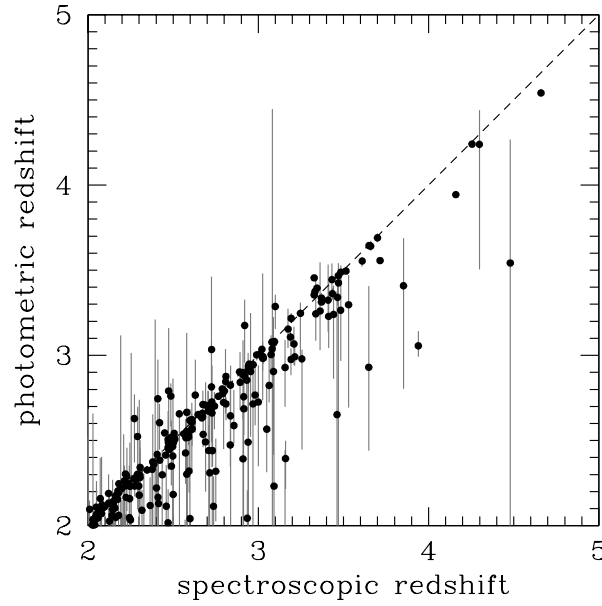


FIGURE 5.2: Spectroscopic vs photometric redshift measurements for the sample of X-ray selected sources used in this paper. The data points correspond to the median value of the PDZ. The errorbars correspond to the 90% confidence interval around the median.

Figure 5.2 plots spectroscopic vs photometric redshifts for the sample used in this paper and illustrates the overall quality of the photometric redshifts estimates. In the X-ray luminosity function calculations we use the full photometric redshift PDZ. These are typically unimodal but at increasing redshift they broaden and secondary peaks may also appear. Also the Aird et al. (2015) PDZs are typically broader than those in the C-COSMOS, 4 Ms CDFS and AEGIS-XD fields. Examples of PDZs used in this paper are shown in Figure 5.3. Table 5.1 presents the number of photometric and spectroscopic redshifts in each field, both total and in the intervals $z = 3 - 4$ and $4 - 5$. Source without optical identifications and hence, without redshift estimates, are a minority in the Chandra surveys sample, 107 in total. These sources can be either moderate redshift ($z \approx 1 - 3$) AGN with red SEDs because of

e.g. obscuration and/or old stellar populations (Koekemoer et al., 2004; Schaerer, D. et al., 2007; Rodighiero, G. et al., 2007; Del Moro, A. et al., 2009), or high redshift systems ($z \gtrsim 3$). Since the redshift distribution of these sources is not known they are assigned a flat PDZ in the redshift interval $z = 1 - 6$ and zero at other redshifts. Fixing the lower limit of the redshift range above to a value between $z = 0$ and $z = 2$ does not change the results and conclusions.

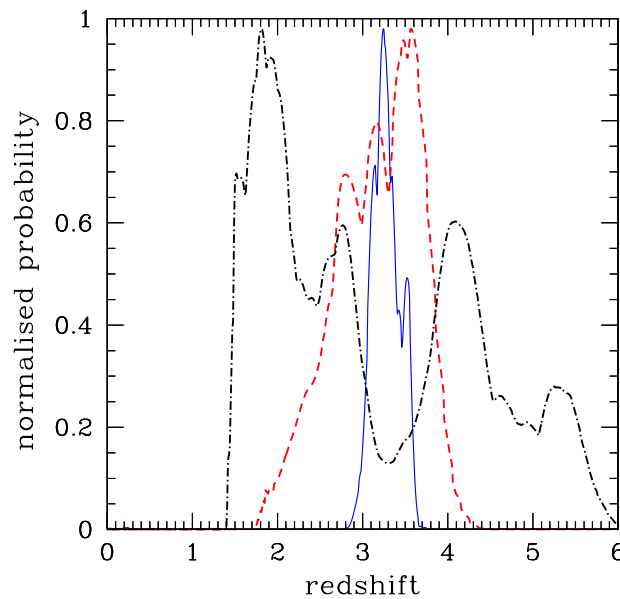


FIGURE 5.3: Examples of photometric redshift Probability Distribution Functions for X-ray selected AGN in the AEGIS field. The different colours correspond to different sources. There is a variety of PDZ shapes in the sample, including uni-modal and relatively narrow (blue solid), unimodal but broad (red dashed) and multi-modal (black dot-dashed).

Additionally we have tested that the differences in the accuracy and outlier fraction of the photometric redshifts in the different fields used in this work do not affect the final results. For the COSMOS, AEGIS-XD and 4 Ms CDFS we can substitute the photometric redshift PDZs adopted in this paper (based on the methods of Salvato et al. 2009, 2011) with those estimated following the methodology of Aird et al. (2015). For these fields we can therefore estimate the X-ray luminosity function at $z = 3 - 5$ (see next sections) using two different sets of photometric redshifts, those of Aird et al. (2015) and those determined following Salvato et al. (2009) and Salvato et al. (2011). We find no systematic differences between the X-ray luminosity functions at $z = 3 - 5$ estimated using the two independent photometric redshift catalogues.

5.2.2 XMM-XXL survey data

The Chandra survey fields are complemented by the wide-area and shallow XMM-XXL survey (PI: Pierre). The XMM-Newton observations of this programme cover a total of about 50 deg^2 of sky split into two nearly equal area fields. In this paper we use the equatorial region of the XMM-XXL, which overlaps with the Canada-France-Hawaii Legacy Survey (CFHTLS) W1 field and covers an area of about 25 deg^2 . The approximate centre of this field lies at right ascension $\alpha = 10 : 22 : 42.20$ and declination $\delta = -04 : 52 : 51.03$.

The reduction of the *XMM* data, the construction of the X-ray catalogue and the association of the X-ray sources with optical counterparts are described by Liu et al. (2016) based on the methods presented in Georgakakis and Nandra (2011). In brief, the X-ray data reduction is carried out using the XMM Science Analysis System (SAS) version 12. We analyse XMM-Newton observations related to the *XMM-XXL* programme that were made public prior to 23 January 2012. XMM-XXL data observed after that date are not included in the analysis. As a result our final catalogue of the equatorial XMM-XXL field misses about 5 deg^2 worth of X-ray coverage. The `EPCHAIN` and `EMCHAIN` tasks of `SAS` are employed to produce event files for the EPIC (European Photon Imaging Camera; Strüder, L. et al., 2001; Turner, M. J. L. et al., 2001) PN and MOS detectors respectively. Flaring periods resulting in elevated EPIC background are identified and excluded using a methodology similar to that described by Nandra et al. (2007). We use X-ray sources detected in the $0.5 - 8 \text{ keV}$ spectral band with Poisson false detection probability of $< 4 \times 10^{-6}$. The final sample consists of 7493 unique sources detected in $0.5-8 \text{ keV}$ spectral band. The fluxes listed in the final source catalogue are in the $0.5-10 \text{ keV}$ band assuming a power-law spectral energy distribution with $\Gamma = 1.4$. These X-ray sources are matched to the SDSS-DR8 photometric catalogue (Aihara et al., 2011) using the Maximum-Likelihood method (Sutherland & Saunders 1992). We assign counterparts to 3798 sources with Likelihood Ratio $LR > 1.5$. At that cut the spurious identification rate is about 6% and the total number of $0.5-8 \text{ keV}$ detected sources with optical counterparts is 3798 (see Table 5.1).

Redshifts for the *XMM-XXL* X-ray sources are from several follow-up spectroscopic campaigns. The *XMM-XXL* field overlaps with the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al., 2013) programme, which provides spectroscopy for UV/optically selected broad-line QSOs and luminous red galaxies. Stalin et al. (2010) presented spectroscopy for X-ray sources selected

in the original XMM-LSS survey (Clerc et al., 2014), which is part of the equatorial XMM-XXL survey field. Most of the redshifts however, are from a total of five special SDSS plates dedicated to follow-up spectroscopy of X-ray sources as part of the Ancillary Programs of SDSS-III. The overlap between those plates and the XMM-XXL survey region is 17.98 deg^{-2} . Targets were selected to have $f_X(0.5 - 10 \text{ keV}, \Gamma = 1.4) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ and $15 < r < 22.5$, where r is either the SDSS PSF magnitude in the case of optically unresolved sources (SDSS type=6) or the SDSS model magnitude for resolved sources. Specific details on these spectroscopic observations, including spectral classification, visual inspection and redshift quality assessment are presented by Menzel et al. (2016). The total number of XMM-XXL sources with secure spectroscopic measurements are presented in Table 5.1. Also shown in this table are the number of sources with spectroscopic redshifts in the interval $z = 3 - 5$.

Although when constructing the X-ray source catalogue of the XMM-XXL field fluxes are estimated for a power-law X-ray spectrum with spectral index $\Gamma = 1.4$, in the rest of the analysis we adopt $\Gamma = 1.9$ for the calculation of fluxes, luminosities and sensitivity maps. This is because at the depth of the XMM-XXL survey AGN at $z > 3$ are powerful QSOs with $L_X(2 - 10 \text{ keV}) \gtrsim 10^{44} \text{ erg s}^{-1}$. The fraction of obscured AGN among such luminous sources is a decreasing function of luminosity (Ueda et al., 2003; Akylas, A. et al., 2006; Merloni et al., 2014; Ueda et al., 2014; Buchner et al., 2015) and therefore a spectral index of $\Gamma = 1.9$, which represents the intrinsic unobscured power-law X-ray spectrum of both local Seyferts e.g Nandra and Pounds, 1994 and luminous high-redshift QSOs (e.g. Vignali et al., 2005; Shemmer et al., 2005; Just et al., 2007) is appropriate for this population. The 0.5-10 keV fluxes estimated for $\Gamma = 1.9$ are about 35% fainter than those for $\Gamma = 1.4$. The XMM-XXL sensitivity curve in the 0.5-10 keV band is shown in Figure 5.1 and is estimated following methods described in Georgakakis and Nandra (2011). In the calculation of the sensitivity curve we only consider the overlapping area between the SDSS-III Ancillary Programs spectroscopic plates used to target X-ray sources and the XMM-XXL survey region. We also take into account the flux limit for follow-up spectroscopy $f_X(0.5 - 10 \text{ keV}, \Gamma = 1.4) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$. This limit appears as a smooth drop in area rather than a sharp cut in Figure 5.1 because the Poisson probability of measuring a flux above this limit is used to determine the sensitivity curve.

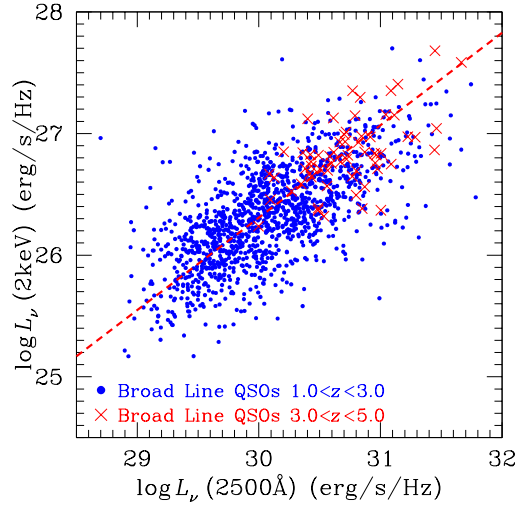


FIGURE 5.4: Monochromatic 2 keV X-ray luminosity, $L_\nu(2 \text{ keV})$, plotted as a function of monochromatic 2500 Å UV luminosity, $L_\nu(2500 \text{ Å})$. The data points (cross or dots) are *XMM-XXL* X-ray selected broad line QSOs with secure spectroscopic redshifts in the interval $z = 1 - 5$. Sources at $z > 3$ are highlighted with different symbols (crosses). The red dashed line is the bisector best-fit $L_\nu(2 \text{ keV}) - L_\nu(2500 \text{ Å})$ relation determined by Lusso, E. et al. (2010); $\log L_{2\text{keV}} = 0.760 \log L_\nu(2500 \text{ Å}) + 3.508$. For the *XMM-XXL* QSOs the X-ray luminosity density at 2 keV is estimated from the 0.5-2 keV flux assuming a power-law X-ray spectrum with index $\Gamma = 1.9$. The 2500 Å monochromatic luminosity is determined from the SDSS photometry. For the k-corrections we adopt the simulated QSO templates of McGreer et al. (2013). For a QSO with redshift z the SDSS photometric bands with effective wavelengths that bracket the wavelength $2500 \times (1 + z) \text{ Å}$ are identified. The mean model QSO SED at that redshift is then scaled to the observed SDSS optical magnitudes in those two bands. The monochromatic luminosity at 2500 Å is then estimated from the scaled model SED. At redshifts $z \gtrsim 2.7$ the rest-frame 2500 Å lies beyond the effective wavelength of the SDSS z -band (9134 Å). For these sources the flux density at $2500 \times (1 + z) \text{ Å}$ is an extrapolation using the model SED. The results do not change if we simply linearly interpolate between the observed flux densities of the SDSS bands that bracket the rest-frame 2500 Å. This model-independent approach however, does not allow extrapolation beyond $z \gtrsim 2.7$.

5.3 Methodology

5.3.1 Statistical background

We intend to estimate the number density per comoving volume ($\text{Mpc}^3, \frac{dV}{dz}$) of AGNs, as a function of various properties, specifically X-ray luminosity, redshift, and obscuring column density. This extended likelihood function will describe the evolution of the X-ray population and all sub-populations (e.g., Compton-thick AGNs). The difficulty is that each of these properties influences our ability to detect objects with such properties. Therefore, our aim is to estimate the parameters of the luminosity function of AGN as a function not only of luminosity (L_x) and redshift (z), but also as a function of the absorption (N_H), i.e., a fraction of the observed AGNs to the total

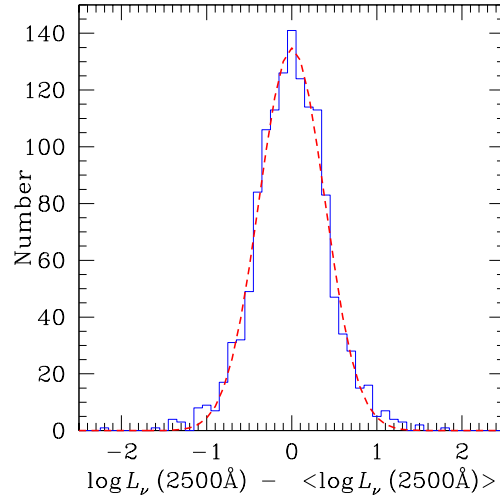


FIGURE 5.5: Histogram of the difference $\Delta \log L_\nu(2500 \text{ \AA})$ between the monochromatic 2500 \AA luminosity of broad-line QSOs in the *XMM-XXL* (blue points of Figure 5.4) and the mean $\langle \log L_\nu(2500 \text{ \AA}) \rangle$ value predicted by the bisector best-fit relation of Lusso et al. (2010; red dashed line of Figure 5.4) for a given $\log L_\nu(2 \text{ keV})$. The red dashed line shows the best-fit Gaussian distribution. The mean value of that distribution is consistent with zero and the standard deviation is $\sigma = 0.4$.

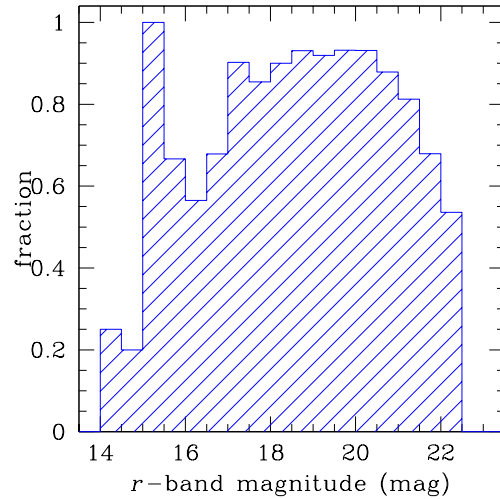


FIGURE 5.6: Fraction of *XMM-XXL* X-ray sources with successful redshift measurement as a function of the *r*-band magnitude of the optical counterpart. The fraction is defined as the ratio between the number of potential targets [$f_X(0.5 - 10 \text{ keV}, \Gamma = 1.4) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ and secure optical counterparts with $r < 22.5 \text{ mag}$] and X-ray sources with successful redshift measurements and fluxes/magnitudes within the above cuts.

AGNs.¹

The basic idea is to generate a source population, with each member having a set

¹To account for the absorption, we use the CR-HR (Count Rate-Hardness Ratio) method which is based on the work of Clerc et al. (2014). Using conversion factor tables, we convert the L_x, z, Γ to count rates for 3 energy bands. The indicator of absorption will be the hardness ratio.

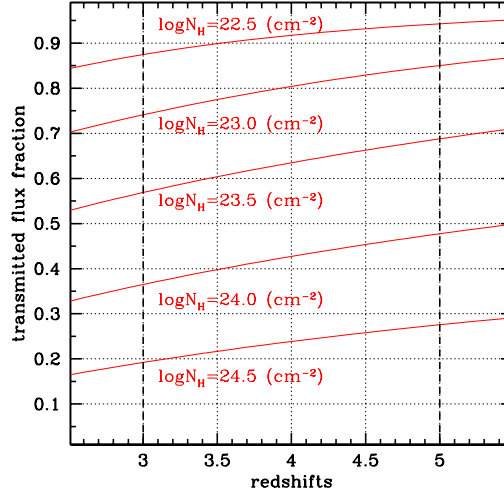


FIGURE 5.7: Impact of different levels of obscuration on the 0.5-10 keV flux of a source as a function of redshift. The vertical axis is the ratio of the flux of an AGN obscured by column density N_H relative to the flux of the same source in the case of zero obscuration. The red curves correspond to different levels of obscuration in the range $\log N_H = 22.5 - 24.5$ (cm^{-2} units). The vertical dashed lines mark the redshift interval of interest, $z = 3 - 5$. For fixed obscuration and intrinsic luminosity the flux of higher redshift AGN is less affected. The differential flux suppression between $z = 3$ and $z = 5$ is small, $\lesssim 10\%$. For the calculation of X-ray k-corrections we adopt the model X-ray spectral of Brightman and Nandra (2011b). These are based on Monte Carlo simulations of an illuminating source at the centre of a sphere with constant density and a conical region (apex at the centre of the sphere) cut-off to approximate a toroidal geometry. These simulations take into account both Compton scattering and photoelectric absorption of the X-ray photons by the obscuring medium. We adopt $\Gamma = 1.9$ for the intrinsic AGN spectrum, an opening angle of the conical region of 60 deg and a viewing angle of 45 deg , i.e. a line-of-sight intersecting the obscuring material.

of L_x, z, N_H, Γ parameters. The luminosity function which relates luminosities and redshifts, is found in Aird et al. (2010) and the function for the N_H that we use, can be found in Ueda et al. (2003). Using the appropriate conversion factors, we take the CR, HR for each set of L_x, z, N_H, Γ at the energy bands we have chosen to work with.

Considering that the sources can be described by a non-homogeneous Poisson process, we build a model for the L_x, z, N_H, Γ . Given a set of parameters, we obtain the likelihood of the two-dimensional variable $CR - HR$ which is again a non-homogeneous Poisson model. Assuming priors for the parameters we want to estimate, we calculate the posterior distribution of the parameters. Due to the high dimensionality of the parameter space, the problem requires multidimensional integrals which can not be done analytically. Hence the maximum likelihood estimation method can not be applied directly.

An interesting approach to the problem is the Bayesian methodology. Using

Markov Chain Monte Carlo (MCMC) methods, we can derive the posterior distributions of the parameters. Therefore, the inference for the parameters is possible even for very difficult problems. So, let us look at the problem and the methodology from its basis.

Let L_x, z, N_H be the luminosity, redshift and absorption column, respectively. Let us assume we can describe the probability to detect a source having a particular value of luminosity L_x , redshift z and absorption N_H . After we analyzing each object in detail, we can bin our sample in such small bins that only one item, denoted with d_i hereinafter, at most can be in each bin with these values. Even though in a real survey the data are divided into two parts, i.e., the data from the detected sources d_i and the data from the sources that exist in the universe but are not detectable, we assume that the space can be divided in small boxes that have one or zero sources. Note that, finding zero sources does not mean that there were zero sources in the box but rather that we could not detect them. Hence the likelihood of the observed data is given by the product of the probability of finding no source when we expected λ items times the product of the probabilities of the observed sources. Since observing one set of (L_{x_i}, z_i, N_{H_i}) is independent from observing another vector (L_{x_j}, z_j, N_{H_j}) , we can assume that the sources in the universe are non-homogeneous Poisson process with parameter $\lambda(L_{x_i}, z_i, N_{H_i})$. The expected number of sources is given by λ which is defined as

$$\lambda = \int \int \int A(L_x, z, N_H) \phi(L_x, z | \theta) f(L_x, z, N_H | \theta) \frac{dV}{dz} dz (\log(L_x)) d \log N_H$$

where,

- $A(L_x, z, N_H)$ is the area of the survey sensitive to a source of a particular luminosity L_x , redshift z and absorption N_H . Then $A(L_x, z, N_H)$ describes the probability of a source being detected.
- $\phi(L_x, z | \theta)$ is the luminosity function and it is the differential number density of sources per unit co-moving volume as a function of L_x and z , given an assumed model described by the set of parameters θ .
- $f(L_x, z, N_H | \theta)$ the N_H function for different luminosities and redshifts.

Let also

- $\{d_i\}$ be the set of the observed discrete data from all sources ($i = 1, \dots, M$) with specific properties, as indicated with the values of (L_x, z, N_H) .

- $p(d_i|L_x, z, N_H)$ be the probability of the observed data from an individual detected source i given their magnitudes.

In the ideal case that we can detect all the sources in the space, the likelihood is the product of the probabilities of observing exactly one source in $dzdL_xdN_H$ at each (L_{x_i}, z_i, N_{H_i}) for the M sources in the sample times the probabilities of observing zero sources in the accessible regions.

$$\mathcal{L}(L_x, z, N_H|\theta) = \prod_{i \in A_{obs}} \frac{(\lambda(z_i, L_{x_i}, N_{H_i})dzdLdN_H)^1}{1!} e^{-\lambda(z_i, L_{x_i}, N_{H_i})dzdL_xdN_H} \prod_{j \in A_{mis}} \frac{e^{-\lambda(z_j, L_{x_j}, N_{H_j})dzdL_xdN_H} (\lambda(z_j, L_{x_j}, N_{H_j})dzdL_xdN_H)^0}{0!}.$$

Substituting,

$$\mathcal{L}(L_x, z, N_H|\theta) = e^{-\sum_{i \in A_{mis}} \lambda_i} \prod_{i=1}^M e^{\lambda_i} \int \int \int \phi(L_x, z, N_H|\theta) A(L_x, z, N_H) f(L_x, z, N_H|\theta) \frac{dV}{dz} dz d(\log(L_x)) dN_H.$$

The luminosity function that we consider is

$$\phi(L_x, z) = \frac{d\Phi(L_x, z)}{d\log(L_x, z)} = K_{norm} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1}$$

with

$$\log L_*(z) = \log L_0 - \log \left[\left(\frac{1+z_c}{1+z} \right)^{p_1} + \left(\frac{1+z_c}{1+z} \right)^{p_2} \right].$$

A simple form of the N_H function as a function of L_x, z is

$$f(L_x, z, N_H) = \begin{cases} 2 - \frac{5+2\epsilon}{1+\epsilon} \psi(L_x, z), & 20.0 \leq \log N_H < 20.5, \\ \frac{1}{1+\epsilon} \psi(L_x, z), & 20.5 \leq \log N_H < 23.0, \\ \frac{\epsilon}{1+\epsilon} \psi(L_x, z), & 23.0 \leq \log N_H < 24.0. \end{cases}$$

where

$$\psi(L_x, z) = \min [\psi_{max}, \max [\psi_{44} - \beta(\log L_x - 44), 0]],$$

and

$$\psi_{max} = \frac{1+\epsilon}{3+\epsilon}.$$

The next step is to convert (L_x, z, N_H) to (CR, HR) :

$$CR = \frac{L_x}{4\pi D_L^2} \times \text{conversion.factor.}$$

We chose to work in 3 energy bands. Therefore 3 count rates are calculated, CR_1, CR_2, CR_3 at bands $[0.5 - 2]$ $[2 - 7]$ $[0.5 - 7]$ keV and a hardness ratio which is defined by

$$HR = \frac{CR_2 - CR_1}{CR_2 + CR_1}$$

The likelihood function as a function of count rates and hardness ratio is now different and can not be written in closed form since it requires calculating the inverse of a 3-dimensional integral. To overcome this difficulty we use Poisson statistics and proceed as follows.

Given that we can not observe all the sources in the universe, we assume that the probability of the observed data from an individual source given the assumed model is given by marginalizing the $p(d_i|L_x, z)$ over the model distribution of L_x and z . The likelihood is

$$\mathcal{L}(\{d_i\}|\theta) = e^{-\lambda} \prod_{i=1}^M \int \int \int p(d_i|L_x, z) \phi(L_x, z|\theta) \frac{dV}{dz} dz d(\log(L_x)).$$

Let N be the counts (number of photons) at the detector. Assuming a simple model, where the luminosity function is independent from z the probability $p(d_i|L_x, z)$ is given by:

$$p(d_i|L_x, z) = \frac{(s+b)^N}{N!} e^{-(s+b)}.$$

This is a Poisson distribution with parameter $(s+b)$ where b is a constant that corresponds to the number of photons that come from the background (not from the sources) and $s = \text{flux} * \text{exposure.time} / \text{conversion.factor}$ which corresponds to the number of photons that come from the AGN. Flux can be given from the relation $\text{flux} = \frac{L_x}{4\pi D^2}$, D^2 is the luminosity distance. Assuming that $A(L_x, z) = 1$ and therefore $\lambda \propto 1$, the likelihood of the observed data is given by:

$$\mathcal{L}(\{d_i\}|\theta) = e^{-\lambda} \prod_{i=1}^M \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{\text{norm}} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)).$$

We assume non-informative prior distributions for the parameters $(\gamma_1, \gamma_2, L_*)$. Specifically we assume that the parameters are uniformly distributed in some bounded intervals. Based on the previous, the joint posterior distribution of the parameters is

given by Bayes theorem:

$$p(L_*, \gamma_1, \gamma_2) \propto \prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \times \\ \mathbb{I}[\alpha_1 < \gamma_1 < b_1] \mathbb{I}[\alpha_2 < \gamma_2 < b_2] \mathbb{I}[\alpha_3 < L_* < b_3],$$

where K_{norm} is a normalizing constant such as

$$\int \phi(L_x) d(\log(L_x)) = N.$$

Note that N is the number of observed sources. Therefore,

$$K_{norm} = N \left[\int \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right]^{-1}.$$

The marginal posterior distribution of the parameters, since it is not easy to find it in an analytical form, we use an MCMC algorithm (Metropolis-Hastings Random Walk). In particular, we propose values for

$$\gamma_1^{candidate} \sim \mathcal{N}(\gamma_1^{current}, v_1),$$

$$\gamma_2^{candidate} \sim \mathcal{N}(\gamma_2^{current}, v_2),$$

$$\log(L_*^{candidate}) \sim \mathcal{N}(\log(L_*^{current}), v_3).$$

We choose variances (v_1, v_2, v_3) for each case such that the acceptance probability of the proposed values will be 30% – 50%. The acceptance probabilities will be:

For γ_1 :

$$p_1 = \max \left\{ 1, \frac{f(\gamma_1^{candidate}|\cdot)}{f(\gamma_1^{current}|\cdot)} \cdot \frac{q(\gamma_1^{current})}{q(\gamma_1^{candidate})} \right\} = \max \{1, l_1\}. \text{ where } f(\cdot) \text{ is the full conditional}$$

probability distribution and q the proposal distribution

$$\begin{aligned}
 l_1 &= \frac{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1^{can}, \gamma_2, L_*) \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))}{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1, \gamma_2, L_*) \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\
 &\quad \times \frac{\mathbb{I}[\alpha_1 < \gamma_1^{can} < b_1]}{\mathbb{I}[\alpha_1 < \gamma_1 < b_1]} \\
 &= \frac{K_{norm}^N(\gamma_1^{can}, \gamma_2, L_*)}{K_{norm}^N(\gamma_1, \gamma_2, L_*)} \prod_{i=1}^N \frac{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))}{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\
 &\quad \times \frac{\mathbb{I}[\alpha_1 < \gamma_1^{can} < b_1]}{\mathbb{I}[\alpha_1 < \gamma_1 < b_1]}.
 \end{aligned}$$

For computational reasons when we run the algorithm at the M-H step if the candidate value for the parameter satisfies the $\mathbb{I}[\alpha_1 < \gamma_1 < b_1]$ then we generate a random number u_1 from the uniform distribution $\mathcal{U}[0, 1]$ and compare the $\log(u_1)$ with $\log(p_1)$ and then accept or not the candidate value. For γ_1 , the $\log(p_1)$ will be given by,

$$\begin{aligned}
 \log(p_1) &= N \log(K_{norm}(\gamma_1^{can}, \gamma_2, L_*)) - N \log(K_{norm}(\gamma_1, \gamma_2, L_*)) \\
 &\quad + \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right) \\
 &\quad + \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right).
 \end{aligned}$$

For γ_2 : $p_2 = \max \left\{ 1, \frac{f(\gamma_2^{candidate})}{f(\gamma_2^{current| \cdot})} \cdot \frac{q(\gamma_2^{current| \cdot})}{q(\gamma_2^{candidate})} \right\} = \max \{1, l_2\}.$

$$\begin{aligned}
 l_2 &= \frac{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1, \gamma_2^{can}, L_*) \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x))}{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\
 &\quad \times \frac{\mathbb{I}[\alpha_2 < \gamma_2^{can} < b_2]}{\mathbb{I}[\alpha_2 < \gamma_2 < b_2]} \\
 &= \frac{K_{norm}^N(\gamma_1, \gamma_2^{can}, L_*)}{K_{norm}^M(\gamma_1, \gamma_2, L_*)} \prod_{i=1}^N \frac{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x))}{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\
 &\quad \times \frac{\mathbb{I}[\alpha_2 < \gamma_2^{can} < b_2]}{\mathbb{I}[\alpha_2 < \gamma_2 < b_2]}.
 \end{aligned}$$

The logarithm of the probability is:

$$\begin{aligned} \log(p_2) = & M \log(K_{norm}(\gamma_1, \gamma_2^{can}, L_*)) - M \log(K_{norm}(\gamma_1, \gamma_2, L_*)) \\ & + \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x)) \right) \\ & + \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right), \end{aligned}$$

under the $\mathbb{I}[\alpha_2 < \gamma_2 < b_2]$

$$\text{For } L_*: p_3 = \left\{ 1, \frac{f(L_*^{candidate} | \cdot)}{f(L_*^{current} | \cdot)} \cdot \frac{q(L_*^{current})}{q(L_*^{candidate})} \right\} = \max\{1, l_3\}.$$

$$\begin{aligned} l_3 = & \frac{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1, \gamma_2, L_*^{can}) \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))}{\prod_{i=1}^N \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1, \gamma_2, L_*) \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\ & \times \frac{\mathbb{I}[\alpha_3 < L_*^{can} < b_3]}{\mathbb{I}[\alpha_3 < L_* < b_3]} \\ = & \frac{K_{norm}^N(\gamma_1, \gamma_2, L_*^{can})}{K_{norm}^N(\gamma_1, \gamma_2, L_*)} \prod_{i=1}^N \frac{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))}{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\ & \times \frac{\mathbb{I}[\alpha_3 < L_*^{can} < b_3]}{\mathbb{I}[\alpha_3 < L_* < b_3]}. \end{aligned}$$

The logarithm of the probability is:

$$\begin{aligned} \log(p_3) = & N \log(K_{norm}(\gamma_1, \gamma_2, L_*^{can})) - N \log(K_{norm}(\gamma_1, \gamma_2, L_*)) \\ & + \sum_{i=1}^M \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right) \\ & + \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right). \end{aligned}$$

under the $\mathbb{I}[\alpha_3 < L_* < b_3]$.

The MCMC algorithm converges to the posterior probability distribution. Therefore, after a burn-in period the algorithm generates values from the posterior (in this case form the marginal posterior) distribution of the parameters. Considering each time a given loss function, the mode, the mean or another point of each distribution can be used as a point estimator of each parameter. In multi-parameter problems

we usually make the assumption of independent parameters and as we have already seen, the proposed values for the Metropolis Hastings MCMC algorithm are taken from the univariate Normal distribution for each parameter. After the simulation, a wise thing to do is to check if the parameters are correlated. Thus, we calculate the covariance matrix (Σ) of the parameters. Following the previously described approach, we saw that the parameters are correlated and hence, we performed another MCMC algorithm under the assumption of correlated parameters.

The proposed values were taken from the 3-dimensional Normal distribution:

$$(\gamma_1^{can}, \gamma_2^{can}, L_*^{can}) \sim \mathcal{MN}((\gamma_1^{current}, \gamma_2^{current}, L_*^{current}), c\Sigma),$$

where Σ the 3×3 covariance matrix from the previous algorithm and c is a tuning constant, such as the acceptance probability of the proposed values will be between 30 – 50%. The acceptance probabilities for the 3-dimensional vector will be:

$$\begin{aligned} p &= \frac{\prod_{i=1}^M \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1^{can}, \gamma_2^{can}, L_*^{can}) \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x))}{\prod_{i=1}^M \int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} K_{norm}(\gamma_1, \gamma_2, L_*) \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\ &\times \frac{\mathbb{I}[(\alpha_1, \alpha_2, \alpha_3) < (\gamma_1^{can}, \gamma_2^{can}, L_*^{can}) < (b_1, b_1, b_3)]}{\mathbb{I}[(\alpha_1, \alpha_2, \alpha_3) < (\gamma_1, \gamma_2, L_*) < (b_1, b_2, b_3)]} \\ &= \frac{K_{norm}^M(\gamma_1^{can}, \gamma_2^{can}, L_*^{can})}{K_{norm}^M(\gamma_1, \gamma_2, L_*)} \prod_{i=1}^M \frac{\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x))}{\int \frac{(s+b)^N}{N!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x))} \\ &\times \frac{\mathbb{I}[(\alpha_1, \alpha_2, \alpha_3) < (\gamma_1^{can}, \gamma_2^{can}, L_*^{can}) < (b_1, b_1, b_3)]}{\mathbb{I}[(\alpha_1, \alpha_2, \alpha_3) < (\gamma_1, \gamma_2, L_*) < (b_1, b_2, b_3)]}. \end{aligned}$$

The logarithm of the probability is:

$$\begin{aligned} \log(p) &= M \log(K_{norm}(\gamma_1^{can}, \gamma_2^{can}, L_*^{can})) - M \log(K_{norm}(\gamma_1, \gamma_2, L_*)) \\ &+ \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*^{can}} \right)^{\gamma_1^{can}} + \left(\frac{L_x}{L_*^{can}} \right)^{\gamma_2^{can}} \right]^{-1} d(\log(L_x)) \right) \\ &+ \sum_{i=1}^N \log \left(\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)) \right). \end{aligned}$$

The algorithm with the assumption that the parameters are correlated has better mixing properties.

Finally, the integral for calculating K_{norm} as well as the integral

$$\int \frac{(s+b)^{N_i}}{N_i!} e^{-(s+b)} \left[\left(\frac{L_x}{L_*} \right)^{\gamma_1} + \left(\frac{L_x}{L_*} \right)^{\gamma_2} \right]^{-1} d(\log(L_x)),$$

can be calculated:

- via Monte Carlo integration by utilizing a special case called importance sampling. Let us say that we want to calculate the integral-mean value:

$$\theta = E_f[h(X)] = \int h(x)f(x)dx$$

If it is difficult to simulate from $f(x)$, then we simulate from a function $g(x)$ called envelope function and then:

$$\theta = \int \frac{h(X)f(X)}{g(X)} g(x)dx = E_g \left[\frac{h(X)f(X)}{g(X)} \right]$$

In our problem, the envelope function is the truncated exponential distribution

$$g(L_x) = \begin{cases} \lambda \frac{e^{-\lambda(L_x-\alpha)}}{1-e^{-\lambda(\beta-\alpha)}} & \alpha \leq L_x \leq \beta, \\ 0, & \text{elsewhere.} \end{cases}$$

- numerically by using the composite trapezoidal rule

$$\int_a^b f(x)dx \approx \left[\frac{b-a}{2} \right] \left(\frac{f(a)}{2} + \sum_{k=1}^{n-1} \left(f(a + k \frac{b-a}{n}) \right) + \frac{f(b)}{2} \right).$$

The following sections will describe the X-ray Luminosity function estimation methodology for the various survey fields. The reader who is not familiar with the various astrophysical survey fields or is only interested in the mathematical/statistical interpretation should omit the remaining parts of this section.

5.3.2 X-ray Luminosity function estimation: Chandra survey fields

A Bayesian approach is adopted for the determination of the X-ray luminosity function. The X-ray sources detected in a survey are essentially Poisson realisations of a parent sample and therefore the likelihood can be written as the product of the Poisson probabilities of individual sources. Following the works of Marshall et al. (1983), Polpo et al. (2002), Aird et al. (2010) and Buchner et al. (2015) the likelihood can be written as

$$\mathcal{L}(d_i | \theta) = e^{-\lambda} \times \prod_{i=1}^N \int d \log L_X \frac{dV}{dz} p(d_i | L_X, z) \phi(L_X, z | \theta), \quad (5.1)$$

where dV/dz is the comoving volume per solid angle at redshift z , d_i signifies the dataset and θ represents the parameters of the luminosity function model, $\phi(L_X, z | \theta)$, that are to be estimated. The multiplication is over all sources, N , and the integration is over redshift and X-ray luminosity. The quantity $p(d_i | L_X, z)$ is the probability of a particular source having redshift z and X-ray luminosity L_X . This captures uncertainties in the determination of both redshifts (e.g. photometric redshifts measurements) and X-ray fluxes because of Poisson statistics and the Eddington bias. In equation 5.1, λ is the expected number of detected sources in a survey for a particular set of model parameters θ

$$\lambda = \int d \log L_X \frac{dV}{dz} A(L_X, z) \phi(L_X, z | \theta). \quad (5.2)$$

where, $A(L_X, z)$ is the sensitivity curve that quantifies the survey area over which a source with X-ray luminosity L_X and redshift z (and hence flux f_X) can be detected. Note that the selection function term, $A(L_X, z)$, is not included within the integral of equation 1. The reader is referred to an extensive discussion in Loredo (2004) on that point.

The goal of this paper is the estimation of the X-ray luminosity function in the redshift interval $z = 3 - 5$. However, defining a sample of X-ray AGN in a relatively narrow redshift range is not straightforward. This is because the redshifts of many sources are determined by photometric methods and therefore have uncertainties, which are not negligible compared to the size of the redshift interval. It may happen for example, that the errors of the photometric redshift of a particular source straddle one (or both) boundaries of the redshift range of interest (see Fig. 5.3).

We deal with this difficulty by simply using all sources in the X-ray sample and splitting the luminosity function model into two terms

$$\phi(L_X, z | \theta) = \phi_1(L_X, z \in [z_1, z_2] | \theta_1) + \phi_2(L_X, z \notin [z_1, z_2] | \theta_2). \quad (5.3)$$

The first term ϕ_1 refers to the luminosity function within the redshift interval of interest, $z = z_1 - z_2$, and has its own set of parameters, θ_1 . The second term, ϕ_2 , corresponds to the X-ray AGN space density outside that redshift range and has a

different set of parameters, θ_2 , which are treated as nuisance parameters. The advantage of this approach is that it allows an estimate of the AGN space density at $z = 3 - 5$ nearly independent of the shape and form of the evolution of the X-ray luminosity function at lower redshift. In any X-ray selected sample the bulk of the AGN population lies at low redshift $z < 3$. This may introduce systematics in the determination of the AGN space density at $z \gtrsim 3$, if the same evolutionary law is fit to the data across all redshifts. In this case it is possible that the determination of the model parameters is dominated by the regions of the redshift/luminosity parameter space with the most data. Our approach minimises the impact of this potential source of bias. In our analysis the ϕ_2 term is modelled as a step function, i.e. the sum of constants that correspond to the AGN space densities at different luminosity and redshift bins. The values of these constants are determined by the data via equation 5.1. For the calculation of equation 5.1 assumptions need to be made on the redshift errors. For sources with photometric redshift determinations we adopt the corresponding redshift Probability Distribution Function (PDZ) as a measure of the uncertainty. Spectroscopic redshifts in the sample have reliabilities $\gtrsim 95\%$ and therefore their corresponding PDZs are assumed to be delta functions at the spectroscopic redshift of the source. Sources without optical counterparts are assigned a flat PDZ in the redshift range $z = 1 - 6$ (see Section 5.2.1).

Poisson statistics are used to determine the flux distribution that is consistent with the extracted source and background counts. A power-law X-ray spectrum with $\Gamma = 1.9$ is adopted in this calculation. The flux distribution in the 0.5-10 keV band is then convolved with the PDZ to estimate the luminosity distribution of each source at rest-frame energies 2-10 keV. The relevant k-corrections also assume a power-law X-ray spectrum with $\Gamma = 1.9$. The two-dimensional probability distribution in L_X and z is the term $p(d_i | L_X, z)$ of equation 5.1. In practice we use importance sampling (Press et al., 1992) to evaluate the integral of equation 5.1. For each source we draw L_X and z samples based on the PDZ and Poisson X-ray counts distribution of that source. The luminosity function is then evaluated for each sample point, L_X, z . The integral of equation 5.1 is simply the average luminosity function of the sample.

5.3.3 X-ray Luminosity function estimation: XMM-XXL field

In the case of the XMM-XXL field, only spectroscopically confirmed sources in the redshift interval $z = 3 - 5$ are used. For that sample there is no need to apply equation 5.3 to determine the corresponding AGN space density. The limitation of that sample however, is that it is both X-ray and optical flux limited because of

the magnitude limit of the spectroscopic follow-up observations. Both cuts need to be accounted for to infer the X-ray luminosity function via equation 5.1. We do that by exploiting the fact that the $z > 3$ XMM-XXL sample consists of powerful [$L_X(2 - 10 \text{ keV}) \gtrsim 10^{44} \text{ erg s}^{-1}$] broad-line QSOs. We use the well-established correlation between monochromatic X-ray [$L_\nu(2 \text{ keV})$] and UV [$L_\nu(2500 \text{ \AA})$] luminosities of broad-line QSOs (e.g. Steffen et al., 2006; Just et al., 2007; Lusso, E. et al., 2010) to link the observed optical magnitudes and X-ray fluxes of the sample and account for the selection effects. Figure 5.4 demonstrates the correlation between $L_\nu(2 \text{ keV})$ vs $L_\nu(2500 \text{ \AA})$ using X-ray selected broad-line QSOs from the XMM-XXL field in the redshift interval $z = 1 - 5$. The low redshift cut is to avoid X-ray AGN with relatively low luminosities, for which the UV/optical continuum shows non-negligible contribution from the host galaxy. We adopt the Lusso, E. et al. (2010) bisector best-fit $\log L_\nu(2 \text{ keV}) = 0.760 \log L_\nu(2500 \text{ \AA}) + 3.508$. At a given monochromatic optical luminosity we assume that the data points scatter around the above relation following a Gaussian distribution with standard deviation σ . Figure 5.5 shows that this is a reasonable assumption. From that figure we estimate $\sigma = 0.4$. Broad Absorption Line (BAL) QSOs represent up to about 26% of optically selected samples (Hewett and Foltz, 2003; Reichard et al., 2003; Gibson et al., 2009), and are known to be X-ray faint either because of absorption or intrinsic X-ray weakness (e.g. Gallagher et al., 2006; Luo et al., 2014). Such sources do not appear to skew the distributions plotted in Figures 5.4 and 5.5. This is likely related to the bright X-ray flux limit of the XMM-XXL survey, which selects against X-ray faint. Under assumptions on the optical Spectral Energy Distribution of QSOs, Figure 5.5 can be used to estimate the SDSS r -band optical magnitude distribution of X-ray sources of given L_X and z and then determine the fraction of this distribution that is brighter than the spectroscopic magnitude limit of the survey, $r_{\text{cut}} = 22.5$. In this case the expected number of detected sources with the surveyed area, λ , in equation 5.1 can be rewritten

$$\lambda = \int d\log L_X \frac{dV}{dz} dz A(L_X, z) B(L_X, z, | r) \times \phi(L_X, z | \theta) \eta(r), \quad (5.4)$$

where $B(L_X, z | r)$ is the SDSS r -band magnitude distribution of a source with L_X and z . The efficiency factor $\eta(r)$ is the success rate of measuring secure redshifts for X-ray sources as a function of the SDSS r -band magnitude. It accounts for the fact that not all X-ray sources with secure counterparts have successful spectroscopic

redshift measurements. Collisions between SDSS fibers or the finite number of science fibers on the SDSS spectroscopic plates mean that not all candidate sources for follow-up spectroscopy can be assigned a fiber. Moreover, for the sources that are observed the rate of secure redshift measurement depends on their optical brightness. At fainter magnitudes the signal-to-noise ratio of the optical spectra decreases and therefore the ability to estimate redshifts is affected. The probability of a source being assigned a fiber is random, while the redshift success rate depends, at least to the first approximation, on optical magnitude. The factor $\eta(r)$ is the number of spectroscopically confirmed X-ray sources in the magnitude interval $r \pm \Delta r$ divided by the total number of X-ray sources that are potential targets for follow-up spectroscopy in the same magnitude range. The parent X-ray sample of potential targets is selected to have $f_X(0.5 - 10 \text{ keV}, \Gamma = 1.4) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ and $15 < r < 22.5 \text{ mag}$ (see Section 5.2.2). The r -band magnitude dependence of $\eta(r)$ is shown in Figure 5.6. For magnitudes in the range $r = 17.0 - 21.5 \text{ mag}$ the efficiency factor $\eta(r)$ is nearly constant and larger than 80%. This fraction drops to about 50% at $r = 22.5$, the limiting magnitude for follow-up spectroscopy with the Sloantelescope, and is zero for $r > 22.5 \text{ mag}$.

In equation 5.4 for the calculation of the optical k-corrections we use the simulated QSO SEDs of McGreer et al. (2013). They are generated assuming a double power-law continuum in the rest-frame UV/optical part of the SED with a break-point at 1100 \AA . The short- and long-wavelength slopes are drawn from normal distributions with means of -1.7 and -0.5 , respectively. For both slopes the scatter is fixed to 0.3 . Emission lines with luminosity dependent equivalent widths as well as Ly α forest absorption are also added to the simulated SEDs as described in McGreer et al. (2013). A total of 60 000 model SEDs are generated in the redshift interval $z = 3 - 5$ and for AGN luminosities $\log L_\nu(2500 \text{ \AA}) \approx 29 - 32 \text{ erg s}^{-1}$. These are used to calculate the expected distribution of the observed r -band optical magnitudes for a given redshift and AGN luminosity, and determine the term $B(L_X, z, | r)$ in equation 5.4. For the calculation of X-ray k-corrections we assume a power-law X-ray spectrum with $\Gamma = 1.9$.

5.3.4 X-ray luminosity function models

The Bayesian framework outlined above explicitly requires a model with a set of free parameters that are constrained by the observations. We consider both non-parametric and parametric models for the X-ray luminosity function in the redshift range $z = 3 - 5$. The non-parametric model simply assumes that the space density

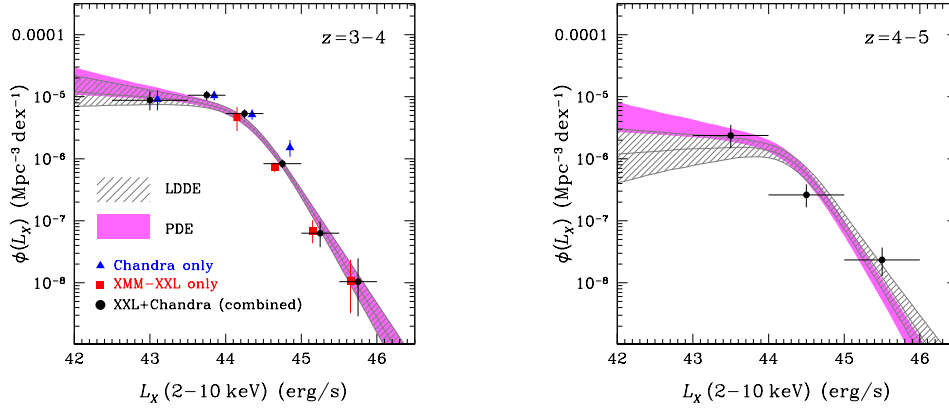


FIGURE 5.8: AGN space density as a function of 2-10 keV X-ray luminosity in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. In both panels the datapoints are the non-parametric binned X-ray luminosity function. The errors correspond to the 68th percentiles of the Probability Distribution Function. The black filled circles are the estimates for the combined *XMM-XXL* and *Chandra* deep field surveys. In the $z = 3 - 4$ panel (left) we also plot the constraints obtained separately from the *XMM-XXL* (red squares) and the *Chandra* deep survey fields (blue triangles). This is to demonstrate that in overlapping luminosity bins the *XMM-XXL* and *Chandra* data yield consistent results. The shaded regions are the 68% confidence intervals of the LDDE and PDE parametric models described in Section 5.3.4. The gray hatched region is for the LDDE and the pink region is for the PDE. The parametric models are estimated at the middle redshift of the redshift intervals, i.e. $z = 3.5$ and $z = 4.5$.

of AGN is constant within a given luminosity and redshift interval. In this case the free model parameters to be estimated are the AGN space densities in each luminosity and redshift bin. This is equivalent to the widely used $1/V_{\text{max}}$ ² approach for the determination of binned luminosity functions. The advantage of the non-parametric approach is that it allows investigation of the form and amplitude of the X-ray luminosity function evolution in a model-independent way.

We also consider four parametric models for the X-ray luminosity function and its redshift evolution, which have been extensively used in the literature. These are the Pure Luminosity Evolution (PLE), Pure Density Evolution (PDE), Luminosity Dependent Density Evolution (LDDE) and Luminosity And Density Evolution (LADE) models. The parametrisation of each model follows Vito et al. (2014). The X-ray luminosity function in the redshift range $z = 3 - 5$ is defined as the space density of AGN per logarithmic luminosity bin and is described by a double power-law of the form

$$\phi(L_X, z) = \frac{K}{\left(\frac{L_X}{L_\star}\right)^{\gamma_1} + \left(\frac{L_X}{L_\star}\right)^{\gamma_2}}, \quad (5.5)$$

²In the case of a step function (i.e. non-parametric) model for $\phi(L_X, z)$ and vanishing uncertainties for the redshifts and luminosities of individual sources, it is straightforward to show that the maximum likelihood value of $\phi(L_X, z)$ in equation 1 reduces to the Page and Carrera (2000) binned luminosity function estimator.

where K is the normalization, γ_1 and γ_2 are the faint and bright-end power-law slopes, respectively, and L_\star is the break luminosity. The PLE model assumes that L_\star is a function of redshift and evolves according to the relation

$$L_\star(z) = L_\star(z_0) \times \left(\frac{1+z}{1+z_0} \right)^p, \quad (5.6)$$

where we fix $z_0 = 3$ and the exponent p parametrises how fast the break luminosity evolves with redshift. In the case of PDE it is assumed that only the normalisation of the luminosity function evolves with the redshift according to the relation

$$K(z) = K(z_0) \times \left(\frac{1+z}{1+z_0} \right)^q, \quad (5.7)$$

where $z_0 = 3$ and the exponent q parametrises the speed of the normalisation factor evolution. The LADE model adopted here is similar to the Independent Luminosity and Density Evolution (ILDE) model described by Yencho et al. (2009). We use equation 5.6 to parametrise the redshift evolution of L_\star and also add a normalisation evolution term of the form

$$K(z) = K(z_0) \times 10^{q(z-z_0)}. \quad (5.8)$$

Finally we consider the LDDE parametrisation of Hasinger, G., Miyaji, T., and Schmidt, M. (2005), where the normalisation factor of equation 5.5 changes with redshift as

$$K(z) = K(z_0) \times \left(\frac{1+z}{1+z_0} \right)^{q+\beta(\log L_X - 44)}, \quad (5.9)$$

where as previously $z_0 = 3$ and the rate of the density evolution also depends on the X-ray luminosity via the parameter β .

Recall that in addition to the model parameters that describe the X-ray luminosity function in the redshift range $z = 3 - 5$ the likelihood function in equation 5.1 also includes terms that correspond to the total AGN space density outside the redshift range of interest, $z < 3$ or $z > 5$ (see equation 5.3). A non-parametric model is adopted to describe that term. This to minimise the impact of parametric model assumptions to the results and conclusions. We use 3 redshift bins in the range $z = 0 - 3$ with size $\Delta z = 1$. Each redshift bin is split into 5 logarithmic luminosity bins in the range $\log L_X(2 - 10 \text{ keV}) = 41 - 46$ (units of erg s^{-1}) with width $\Delta \log L_X = 1 \text{ dex}$. One additional term is used to model the luminosity function in the range $z = 5 - 6$. A constant AGN space density is assumed within the above luminosity and redshift

intervals. We therefore use a total of 16 nuisance parameters to model the AGN X-ray luminosity function at $z < 3$ or $z > 5$.

The MultiNest multimodal nested sampling algorithm (Feroz and Hobson, 2008; Feroz, Hobson, and Bridges, 2009) is used for both Bayesian parameter estimation and the calculation of the Bayesian evidence, \mathcal{Z} , of each model, i.e. the integral of the model likelihood over the parameter space allowed by the priors. The Bayesian evidence is used for model comparison, i.e. to select among different models the one that better describes the data.

5.3.5 Potential obscuration effects

The analysis presented in this paper uses the full-band selected sample to determine the X-ray luminosity function of AGN at $z > 3$. This is because of the higher sensitivity of the 0.5-10 keV band compared to e.g. the 2-10 keV band, which translates to a larger number of sources. A potential issue however, is that the analysis described above ignores the impact of obscuration on the observed AGN flux. Figure 5.7 shows that for AGN at $z > 3$ moderate line-of-sight columns, $\log N_H = 23$ (cm^{-2}), suppress the observed 0.5-10 keV flux of AGN by less than about 25%. Higher column densities however, have a larger impact on the 0.5-10 keV flux and therefore result in incompleteness in the sample that is not accounted for in our analysis. Nevertheless, we are primarily interested in the differential evolution of AGN in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. Figure 5.7 shows that for fixed obscuration and intrinsic AGN luminosity there is little difference in the level of flux suppression between redshifts $z = 3$ and $z = 5$. Incompleteness related to obscuration is therefore expected to be similar across the redshift range $z = 3 - 5$. This allows direct comparison of the inferred AGN space densities in the redshift bins $z = 3 - 4$ and $z = 4 - 5$, under the assumption that the distribution of AGN in obscuration does not change dramatically between these redshift intervals. Recent studies on the obscuration distribution of AGN (Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015) show that the obscured AGN fraction increases with redshift, at least to $z \approx 3$. In Buchner et al., 2015 for example, the obscured fraction in Compton thin AGN is about 50% at $z < 1$ and increases to 70% at $z \approx 3$. At higher redshifts however, current constraints on the evolution of the obscured AGN fraction are still limited by small number statistics. Nevertheless, there are suggestions that the obscured AGN fraction remains roughly constant with redshift at $z \gtrsim 3$ (Buchner et al., 2015).

Obscuration related effects potentially have a larger impact on the analysis of the *XMM-XXL* data. For these sources it is explicitly assumed that their optical/UV

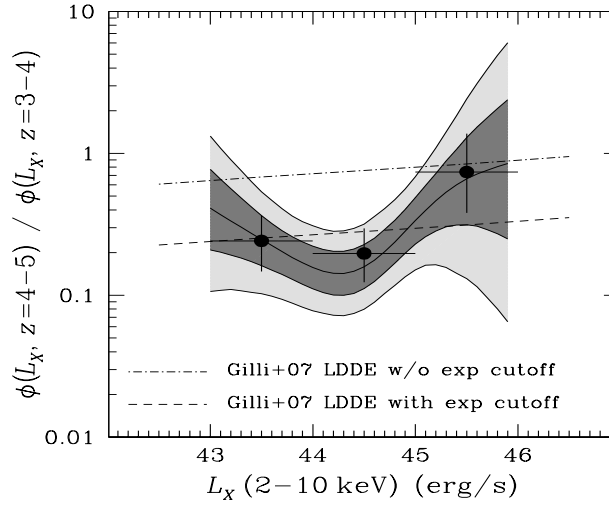


FIGURE 5.9: X-ray luminosity dependence of the ratio of the AGN space densities in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. The data points are for the non-parametric binned X-ray luminosity function. A double power-law is also fit independently to the data in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. The ratio of these parametric fits is shown by the light/dark grey shaded regions. The widths of the dark-grey and light-grey shaded regions at a given luminosity corresponds to the 68% and 95% confidence intervals. The predictions of the Gilli et al. (2007) LDDE parametrisation of the luminosity function with (dashed line) and without (dot-dashed line) an exponential cutoff at high redshift is also plotted for comparison.

continua are not obscured or reddened. This assumption allows us to place them on the $L_{\nu}(2 \text{ keV}) - L_{\nu}(2500 \text{ \AA})$ correlation for type-I AGN and correct for the spectroscopic magnitude cut as explained above. This poses a problem when comparing the Chandra with the XMM-XXL constraints on the X-ray luminosity function. Nevertheless, the shallow XMM-XXL survey is only sensitive to powerful QSOs [$L_X(2 - 10 \text{ keV}) \gtrsim 10^{44} \text{ erg s}^{-1}$] at $z > 3$. At lower redshift at least, it is well established that the obscured AGN fraction is rapidly decreasing with increasing accretion luminosity (e.g. Ueda et al., 2003; Akylas, A. et al., 2006; Ueda et al., 2014; Buchner, J. et al., 2014; Merloni et al., 2014). Obscuration incompleteness corrections are therefore likely to be small for luminosities $L_X(2 - 10 \text{ keV}) \gtrsim 10^{44} \text{ erg s}^{-1}$. We explore this assumption further in the next section by directly comparing the Chandra and XMM-XXL constraints in overlapping luminosity bins.

TABLE 5.3: Power-law fits to the data in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$

redshift interval	$\log K$ (Mpc^{-3})	$\log L_{\star}$ (erg/s)	γ_1	γ_2
(1)	(2)	(3)	(4)	(5)
$z = 3 - 4$	$-5.15^{+0.12}_{-0.17}$	$44.41^{+0.13}_{-0.12}$	$0.19^{+0.15}_{-0.13}$	$2.25^{+0.29}_{-0.24}$
$z = 4 - 5$	$-6.62^{+0.74}_{-0.82}$	$45.00^{+0.57}_{-0.58}$	$0.68^{+0.24}_{-0.38}$	$2.16^{+0.98}_{-0.47}$

Listed are the best-fit parameters for the simple power-law fits to the data in the redshift intervals $z = 3 - 4$, $z = 4 - 5$. The listed values are the median of the probability distribution function of each parameter. The errors correspond to the 16th and 84th percentiles around the median. The columns are: (1) redshift interval (2) X-ray luminosity function normalisation (see equation 5.5), (3) break luminosity of the X-ray luminosity function (see equation 5.5), (4) faint-end slope, (5) bright-end slope.

5.4 Results

5.4.1 The non-parametric X-ray luminosity function determination

Figure 5.8 plots the non-parametric estimate of the X-ray luminosity function in two redshift intervals, $z = 3 - 4$ and $z = 4 - 5$. For the low redshift sub-sample, $z = 3 - 4$, we also plot separately the AGN space density determined independently from the *Chandra* deep fields and the *XMM-XXL* survey. As expected these datasets probe different luminosity ranges. The *Chandra* fields provide constraints to $\log L_X(2 - 10 \text{ keV}) = 42.5$ (erg s^{-1}), but do not probe sufficient volume to detect luminous and rare sources with luminosities $\gtrsim 10^{45} \text{ erg s}^{-1}$. The *XMM-XXL* survey can constrain the bright-end of the luminosity function but is not sensitive enough to detect AGN below $L_X(2 - 10 \text{ keV}) \lesssim 10^{44} \text{ erg s}^{-1}$. Nevertheless, in the interval $\log L_X(2 - 10 \text{ keV}) = 44 - 45$ (erg s^{-1}) both the *Chandra* and *XMM* data provide meaningful constraints to the AGN space density. These independent non-parametric estimates are in good agreement within the 68% errors plotted in Figure 5.8. This suggests that obscuration effects do not have a strong impact on the AGN space densities estimated from the *XMM-XXL* data. In Figure 5.8 the $z = 4 - 5$ panel does not plot separately the space density estimates from the *Chandra* and then *XMM-XXL* survey fields. This is because at this redshift interval there is virtually no overlap in the luminosities probed by the two datasets. The *XMM-XXL* constrains only the brightest luminosity bin, $\log L_X(2 - 10 \text{ keV}) = 45 - 46$ (units erg s^{-1}). The *Chandra* fields include only sources that are fainter than $L_X(2 - 10 \text{ keV}) = 10^{45} \text{ erg s}^{-1}$.

A striking result illustrated in Figure 5.8 is the strong evolution of the AGN population between redshifts $z = 3$ and $z = 5$. This is further demonstrated in Figure 5.9, which plots as a function of luminosity the ratio of the non-parametric AGN space densities in the redshift intervals $z = 3 - 4$ and $z = 4 - 5$. At luminosities $L_X(2 - 10 \text{ keV}) \lesssim 10^{45} \text{ erg s}^{-1}$ there is a factor of 5 decrease in the AGN space density from $z = 3 - 4$ to $z = 4 - 5$. There is also evidence for luminosity dependent evolution. AGN with $L_X(2 - 10 \text{ keV}) \lesssim 10^{45} \text{ erg s}^{-1}$ in Figure 5.9 appear to evolve faster than intrinsically brighter sources. We quantify the statistical significance of the evidence for luminosity dependent evolution using the quantity

$$\mathcal{R} = \log \frac{\phi(L_X = 10^{45} - 10^{46}, z = 4 - 5)}{\phi(L_X = 10^{45} - 10^{46}, z = 3 - 4)} - \log \frac{\phi(L_X = 10^{43} - 10^{45}, z = 4 - 5)}{\phi(L_X = 10^{43} - 10^{45}, z = 3 - 4)}, \quad (5.10)$$

where $\phi(L_X, z)$ is the binned (non-parametric) luminosity function. \mathcal{R} is the logarithmic difference between the high luminosity ($L_X > 10^{45} \text{ erg s}^{-1}$) data point of Figure 5.9 and the sum of the two moderate luminosity data points ($L_X = 10^{43} - 10^{45} \text{ erg s}^{-1}$) of the same plot. We bin together moderate luminosity AGN to simplify the problem and also improve the statistical reliability of the results. We use the full probability density distribution of \mathcal{R} and find that at the 90% probability $\mathcal{R} > 0$, i.e. there is differential evolution between moderate ($L_X = 10^{43} - 10^{45} \text{ erg s}^{-1}$) and powerful ($L_X > 10^{45} \text{ erg s}^{-1}$) X-ray AGN from $z = 3 - 4$ to $z = 4 - 5$. We further investigate this using a simple parametric approach. A double power-law function (equation 5.5) is fit independently to the data in the redshift bins $z = 3 - 4$ and $z = 4 - 5$. In this exercise evolutionary effects within each of the two redshift intervals are ignored. The best-fit parameters are presented in Table 5.3. The ratio between the two double power-laws in the redshift bins $z = 3 - 4$ and $z = 4 - 5$ is plotted with the shaded region in Figure 5.9. The apparent increase of this ratio toward faint luminosities is because the faint-end slope of the $z = 4 - 5$ sample is poorly constrained and on the average steeper than that of the $z = 3 - 4$ sample. Nevertheless, in the interval $L_X(2 - 10 \text{ keV}) = 10^{43} - 10^{46} \text{ erg s}^{-1}$, where constraints on the AGN space density are available for both the $z = 3 - 4$ and the $z = 4 - 5$ sub-samples, the ratio between the two power-laws is consistent with luminosity-dependent evolution. Larger samples are needed however, particularly in the redshift interval $z = 4 - 5$, to reduce the uncertainties in Figure 5.9 and further explore the evidence for luminosity dependent evolution.

The amplitude of the AGN X-ray luminosity function evolution between the redshift intervals $z = 3 - 4$ and $z = 4 - 5$ in Figure 5.9 is further compared with the predictions of luminosity function parametrisations that include an exponential decline at $z \gtrsim 3$ (e.g. Gilli, R., Comastri, A., and Hasinger, G., 2007). Such models are motivated by the rapid evolution of optical QSO space density at high redshift (e.g. Schmidt, Schneider, and Gunn, 1995; Richards et al., 2006). We use the Gilli, R., Comastri, A., and Hasinger, G. (2007)³ LDDE parametrisation of the X-ray luminosity function with and without an additional exponential cutoff to predict the space density of AGN at $z = 3.5$ and $z = 4.5$. The ratio between these predictions is plotted in Figure 5.9. The amplitude of the AGN evolution inferred in this paper for luminosities $L_X(2 - 10 \text{ keV}) \lesssim 10^{45} \text{ erg s}^{-1}$ is consistent with the Gilli et al. (2007) LDDE parametrisation that includes an exponential cutoff. AGN with luminosities in the range $L_X(2 - 10 \text{ keV}) = 10^{45} - 10^{46} \text{ erg s}^{-1}$ lie in between the Gilli et al. (2007) model predictions with and without an exponential cutoff. This is suggestive of milder evolution, albeit at the $\approx 90\%$ confidence level.

5.4.2 Parametric X-ray luminosity function determination

Next we use Bayesian model comparison to assess which of the evolutionary models outlined in Section 5.3.4 provides a better description of the observations in the redshift interval $z = 3 - 5$. The PLE, PDE, LADE and LDDE parametric models of Section 5.3.4 are fit to the combined *Chandra* and *XMM-XXL* dataset. Table 5.4 presents the best-fit parameters for each parametric model. The Bayes factor (ratio between evidences) of each model relative to the model with the highest evidence (PDE) is also shown in that table.

The model with the highest evidence in Table 5.4 is the PDE, with the second best being the LDDE. The Bayes factor of the two models is $\Delta \log_{10} Z = 0.25$. Based on the Jeffreys interpretation of the Bayes factor (Jeffreys, 1961) this difference suggests that both models describe equally well the evolution in the redshift interval $z = 3 - 5$ of the X-ray selected AGN sample presented in this paper. The bulk of the AGN in the present sample have X-ray luminosities $L_X(2 - 10 \text{ keV}) < 10^{45} \text{ erg s}^{-1}$. Figure 5.9 shows that such sources experience similar reduction in their space density between $z = 3 - 4$ and $z = 4 - 5$, i.e. consistent with pure density evolution. More luminous AGN [$L_X(2 - 10 \text{ keV}) > 10^{45} \text{ erg s}^{-1}$], which appear to experience milder evolution in Figure 5.9, represent only a small fraction of the present sample. This combined

³<http://www.bo.astro.it/~gilli/counts.html>

with the fact that the PDE is a simpler model compared to the LDDE (5 vs 6 free parameters) results in similar Bayesian evidences for these two parametric models.

Table 5.4 also shows that the LADE parametrisation adopted in this work performs worse than the LDDE. The Bayes factor of the two models is $\Delta \log_{10} Z = 1.08$. This difference is strong evidence in favor of the LDDE (Jeffreys, 1961). The one model that performs significantly worse than the rest is the PLE. The Bayes factor of that model relative to the one with the highest evidence (PDE) is $\Delta \log_{10} Z = 4.02$. Based on the Jeffreys (1961) interpretation of the Bayes factor this is decisive evidence against the PLE model.

Although the PDE model is favoured by our analysis for the evolution of AGN in the redshift interval $z = 3 - 5$ in the next sections we use the LDDE model to compare with previous studies and to determine the contribution of X-ray AGN to the ionisation of the Universe. This is because of the small difference in the evidences of the PDE vs the LDDE and the long literature on the LDDE evolutionary model. Nevertheless, the results and conclusions are not sensitive to the particular choice of AGN evolutionary model.

5.4.3 Comparison with previous studies

A number of studies on the X-ray luminosity function of AGN have appeared in the literature recently. These include works focusing on the space density of X-ray AGN at high redshifts $z \geq 3$ (Civano et al., 2010; Kalfountzou et al., 2014; Vito et al., 2014) and results on the global X-ray luminosity function evolution across redshift (Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015; Miyaji et al., 2015). In this section we compare our X-ray luminosity function with previous studies using X-ray samples selected in the 0.5-2 or 0.5-10 keV bands, i.e. similar to the one presented in this paper. Vito et al. (2014) compiled one of the largest samples to date in the redshift interval $z = 3 - 5$. They select sources in the 0.5-2 keV band and also present non-parametric ($1/V_{\max}$) estimates of the X-ray luminosity function. These aspects of the Vito et al. (2014) analysis methodology are similar to ours. We also use recent results of Aird et al. (2015) as a representation of parametric approaches to fit the X-ray luminosity function of AGN to the full redshift interval accessible to current X-ray selected samples, $z \approx 0 - 5$.

Vito et al. (2014) use a sample similar to ours in size to determine the X-ray luminosity function in the range $3 < z \lesssim 5$. They use both photometric and spectroscopic redshifts (total of 141) and apply corrections to account for sources without photometric redshift determinations (total of 65). We compare our best-fit LDDE model

with their results in Figure 5.10. The redshift intervals of each panel are the same as in Figure 7 of Vito et al. (2014). We plot their non-parametric $1/V_{\text{max}}$ estimates that include their redshift incompleteness corrections. Our LDDE parametric models in Figure 5.10 is estimated at the middle of each redshift interval. We find that our X-ray luminosity function determination is systematically lower than the Vito et al. (2014) data points. This discrepancy ($> 3\sigma$ significance in e.g. the redshift interval $z = 3.47 - 3.90$ of Figure 5.10) may be related to the fact that Vito et al. (2014) account for X-ray obscuration in the determination of X-ray luminosities and the calculation of the V_{max} of individual sources. A total of 36 of their 141 sources have column densities $N_H \gtrsim 10^{23} \text{ cm}^{-2}$, which translates to suppression of their fluxes compared to the unobscured ($N_H = 0 \text{ cm}^{-2}$) case by $\gtrsim 20\%$ (see Figure 5.7). Such corrections are ignored in the analysis presented here. Alternatively the discrepancy may be related to how sources with photometric redshifts or sources without any redshift information are treated. Vito et al. (2014) use only the best-fit photometric redshift solution without taking into account the corresponding uncertainties. It is further assumed that the 65 sources in their sample without photometric redshifts all lie in the redshift interval $z = 3 - 5$ and that they follow the redshift distribution of the X-ray sources with redshift determinations (photometric or spectroscopic) in the range $z = 3 - 5$. The amplitude of this correction is $\approx +0.5$ and $+0.25$ dex increase at luminosities $L_X(2 - 10 \text{ keV}) \approx 10^{43}$ and $10^{45} \text{ erg s}^{-1}$ respectively. In that respect Vito et al. (2014) determine maximal X-ray luminosity functions. It is likely that some of the X-ray sources without redshift information lie outside the redshift range $z = 3 - 5$. Heavily obscured AGN or AGN hosted by early type hosts at moderate redshifts, $z \approx 1 - 3$, may also have red SEDs, similar to those of high redshift sources (Koekemoer et al., 2004; Schaerer, D. et al., 2007; Rodighiero, G. et al., 2007; Del Moro, A. et al., 2009).

We further investigate this issue by adopting a methodology similar to that of Vito et al. (2014) to determine the X-ray luminosity function. We use the best-fit photometric redshifts only, i.e. ignoring the photometric redshift probability distribution functions. Sources without optical identifications in the sample are assigned random redshifts in the range $z = 3 - 5$ based on the redshift distribution of sources with spectroscopic or photometric redshift measurements. The results are presented in Figure 5.11. It shows that an approach that ignores photometric redshift errors results in an overestimation of the AGN space density. This is not only because of sources without optical identifications and therefore without photometric redshift (total of 107 in our sample) estimates being forced to lie in the range $z = 3 - 5$. The

overestimation is mainly the result of ignoring photometric redshift uncertainties. In the likelihood equation 5.1 the probability of a source having redshift z and luminosity L_X is weighted by the luminosity function. Sources with broad (see Fig. 5.3) redshift probability distribution functions are more likely to lie at low redshift and moderate luminosities simply because the space density of AGN is higher there. Ignoring this weighting in equation 5.1 by e.g. fixing photometric redshifts to a single (best-fit) value overestimates the luminosity function at high redshifts.

Also plotted in Figure 5.10 are the flexible double power-law parametric model of Aird et al. (2015) for their 0.5-2 selected sample before applying corrections for obscuration (i.e. directly comparable to our analysis). The binned X-ray luminosity function estimates of Aird et al. (2015) are also shown. These data points are estimated using the N_{obs}/N_{mdl} method developed by Miyaji, T., Hasinger, G., and Schmidt, M. (2001) and are therefore not independent of the underlying parametric model plotted in Figure 5.10. Nevertheless, these binned estimates provide a measure of the associated uncertainties and can be used to identify redshift and luminosity intervals where the parametric model provides a poorer fit to the data. The Aird et al. (2015) X-ray luminosity function estimated from their 0.5-2 keV selected sample is in fair agreement with our results. This may be not surprising given the overlap between the two datasets and the similar approaches. Figure 5.12 plots the X-ray luminosity density evolution of AGN using our LDDE model and the Aird et al. (2015) total X-ray luminosity function that includes both obscured (Compton thin and thick) and unobscured sources. The X-ray luminosity function estimated in this work accounts for about 40% of the total X-ray luminosity density determined by Aird et al. (2015).

Finally, we also compare in Figure 5.10 the X-ray luminosity function with high redshift determinations of the optical QSO luminosity function. The conversion from UV/optical to X-rays ultimately depends on the scatter in the relations between bolometric and X-ray or UV luminosities. There are suggestions that the bolometric-to-X-ray luminosity ratio has a larger scatter than the bolometric-to-UV luminosity ratio (Hopkins, Richards, and Hernquist, 2007). We therefore convert from UV/optical to X-rays by convolving the best-fit parametric model of UV/optical QSOs from the relevant publications with the $L_\nu(2 \text{ keV}) - L_\nu(2500 \text{ \AA})$ relation of Lusso, E. et al. (2010) assuming a scatter of 0.4 dex (e.g. Figure 5.5). This calculation also requires assumptions on the shape of the UV SED of QSOs at wavelengths $\lambda \gtrsim 1450 \text{ \AA}$. We assume a power-law of the form $L(\nu) \propto \nu^{-0.5}$ (Berk et al., 2001; Telfer et al., 2002). A steeper slope, $L(\nu) \propto \nu^{-0.61}$ (Lusso et al., 2015), translates to an X-ray luminosity

function that is offset by $\delta \log L_X \approx -0.03$ dex compared to $L(\nu) \propto \nu^{-0.5}$. At the $z = 3.1$ and $z = 4.1$ panels of Figure 5.10 we overplot the Masters et al., 2012 best-fit double power-law models at $z = 3.2$ and $z = 4$. At the $z = 3.35$ panel of Figure 5.10 we show the Ross et al. (2013) LEDE (Luminosity Evolution and Density Evolution) model fit to the BOSS Stripe82 QSO data (Palanque-Delabrouille, N. et al., 2011). We choose not to extrapolate this model past redshift $z = 3.5$, the limiting redshift of the BOSS Stripe82 QSO sample. At the redshift interval $4.3 < z < 5.1$ (mean redshift $z = 4.7$) we transform to X-rays the optical QSO luminosity function determined by McGreer et al. (2013). The optical luminosity functions are plotted by shaded regions in Figure 5.10. The shape and normalisation of the UV/optical QSO luminosity functions in Figure 5.10 at luminosities $\gtrsim 10^{45} \text{ erg s}^{-1}$ are sensitive to the scatter of the $L_\nu(2 \text{ keV}) - L_\nu(2500 \text{ \AA})$ relation used in the convolution.

A striking result from Figure 5.10 is the steep faint-end slope of the Masters et al. (2012) best-fit double power-law models at $z = 3.2$ and $z = 4$ compared to the faint-end slope of the X-ray luminosity function. The ratio between the UV/optical and the X-ray luminosity functions is a proxy of the type-I fraction among AGN, $\mathcal{F}_{\text{type-I}}$, and can be used to explore the luminosity dependence of this quantity at high redshift. In this section we define type-I AGN as sources with blue UV/optical continua, typical of broad-line QSOs. We use the Masters et al. (2012) luminosity function and the LDDE parametrisation for the X-ray luminosity function to determine $\mathcal{F}_{\text{type-I}}$ and plot the results in the case of $z = 3.2$ in Figure 5.13. The luminosity dependence of the $\mathcal{F}_{\text{type-I}}$ at $z = 4$ is very similar and is not shown. We find evidence that the type-I AGN fraction at $z > 3$ is a non-monotonic function of luminosity. There is a minimum at $L_X(2 - 10 \text{ keV}) \approx 10^{44} \text{ erg s}^{-1}$, followed by a steep increase toward fainter luminosities.

The behaviour of $\mathcal{F}_{\text{type-I}}$ in Figure 5.13 is opposite to studies that find a drop in the type-I or X-ray unobscured ($N_H < 10^{22} \text{ cm}^{-2}$) AGN fraction with decreasing luminosity or equivalently that type-II or obscured ($N_H > 10^{22} \text{ cm}^{-2}$) AGN dominate at faint luminosities (Ueda et al., 2003; Akylas, A. et al., 2006; Ueda et al., 2014; Merloni et al., 2014; Aird et al., 2015). Studies on the obscuration distribution of AGN in the local Universe support a picture where the obscured AGN fraction increases with decreasing luminosity. Nevertheless, they also find evidence for a turnover (drop) of the obscured AGN fraction at very faint X-ray luminosities, below about $10^{42} \text{ erg s}^{-1}$ (e.g Burlon et al., 2011; Brightman and Nandra, 2011a). Recently, Buchner et al. (2015) extended these results to higher redshift and found evidence that the obscured AGN fraction peaks at a redshift-dependent luminosity and then

drops at both brighter and fainter luminosities. Figure 5.13 overplots the obscured AGN fraction derived by Buchner et al. (2015) in the redshift interval $z = 2.7 - 4$. For this comparison we assume that $\mathcal{F}_{\text{type-I}} = 1 - \mathcal{F}_{\text{obscured}}$. This is a simplistic assumption because the definition of Type-I QSOs in the case of UV/optical samples and unobscured/obscured AGN in X-ray samples like in Buchner et al. (2015) is different. Nevertheless, to the first order the quantity $1 - \mathcal{F}_{\text{obscured}}$ should be at least loosely related to $\mathcal{F}_{\text{type-I}}$. In Figure 5.13 there is qualitative agreement between the Buchner et al. (2015) $1 - \mathcal{F}_{\text{obscured}}$ parameter and our definition of $\mathcal{F}_{\text{type-I}}$.

At bright luminosities the quantity $\mathcal{F}_{\text{type-I}}$ in Figure 5.13 increases with increasing L_X . This is in agreement with previous studies based on X-ray (Ueda et al., 2014, e.g.), optical (e.g. Simpson, 2005) or infrared (e.g. Assef et al., 2013) data. At the brightest luminosities probed by our data, $L_X = 10^{45} - 10^{46} \text{ erg s}^{-1}$, the type-I fraction is about $75 \pm 25\%$. This number is relevant to the population of powerful QSOs (bolometric $L_{\text{bol}} \approx 10^{47} \text{ erg s}^{-1}$) with reddened UV/optical continua [extinction $E(B - V) \approx 5$] identified in recent wide-area infrared surveys (Stern et al., 2014). These reddened/obscured sources correspond to X-ray luminosities $L_X(2 - 10 \text{ keV}) \approx 5 \times 10^{45} \text{ erg s}^{-1}$ (Marconi et al., 2004) and are suggested to represent up to 50% of the QSO population at these luminosities (Assef et al., 2015). We caution however, that there are uncertainties on the inferred bolometric luminosities of these sources and hence, the obscured AGN fraction, depending on the assumed geometry, physical scale and covering fraction of the obscuring material (Assef et al., 2015). Additionally, the X-ray properties of these infrared selected AGN are poorly known. There are suggestions that they represent heavily obscured, even possibly Compton thick, systems (Stern et al., 2014), in which case they are expected to be underrepresented in our sample.

5.4.4 Contribution of AGN to ionisation of the Universe

The X-ray luminosity function can be used to set limits on the contribution of the AGN to the radiation field that keeps the Universe ionised at redshift $z > 3$. The advantage of X-ray selection is that it provides a better handle on the faint-end of the AGN luminosity function compared to UV/optically selected samples. The downside is that assumptions on the escape fraction of UV photons have to be made to convert AGN space densities to ionising photon densities. At the very least however, X-rays surveys can set strict upper limits on the contribution of AGN to the ionising photon field, under the assumption that all photons emitted by the central source escape to

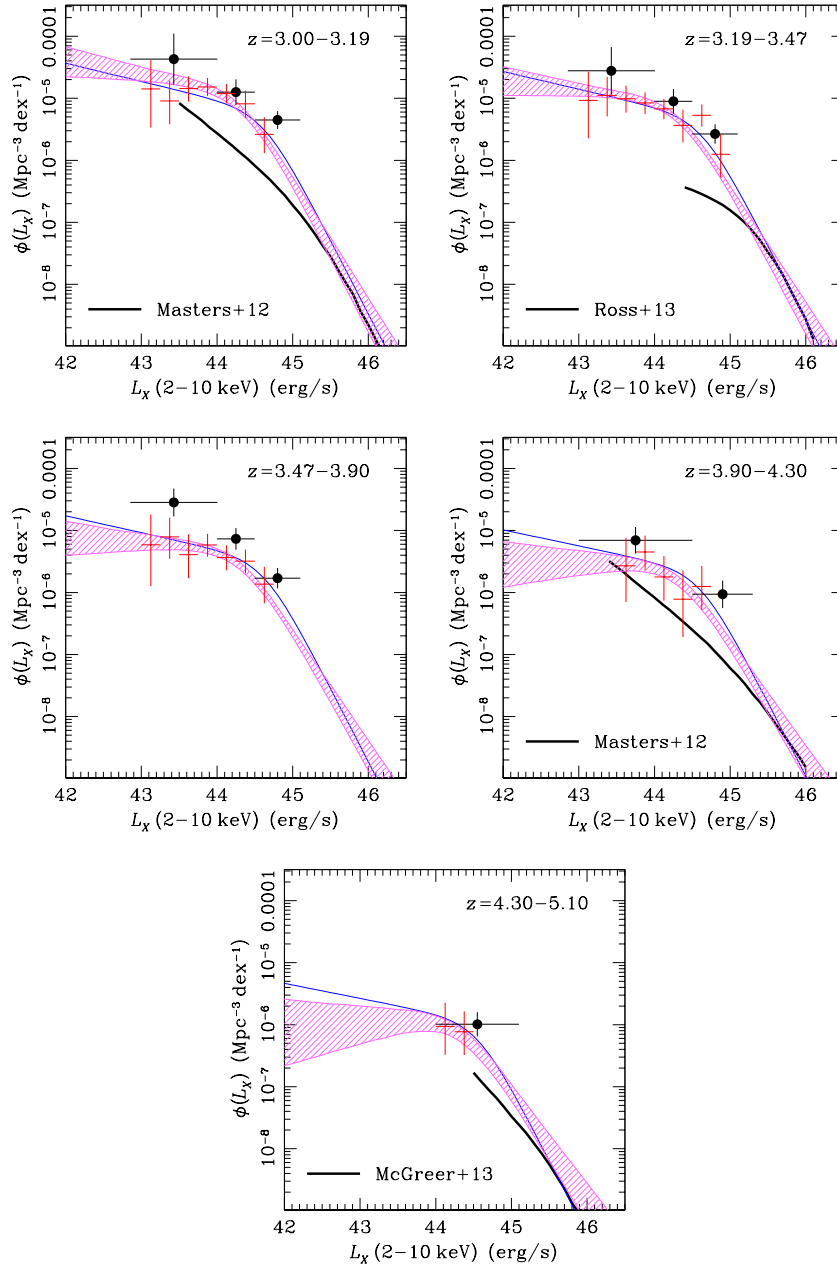


FIGURE 5.10: Comparison of our best-fit LDDE model (pink shaded region) with previous estimates of the X-ray luminosity function in the range $z = 3 - 5$. The panels correspond to the redshift intervals of Vito et al. (2014) to allow direct comparison with their results. The black filled circles are the $1/V_{\text{max}}$ binned luminosity function estimates of Vito et al. (2014). The thin blue curves are the flexible double power-law parametric model of Aird et al. (2015) for their 0.5-2 keV band selected sample without corrections for obscuration. For the Aird et al. (2015) 0.5-2 keV sample the binned X-ray luminosity function estimates are also shown by the red crosses. The UV/optical QSO luminosity functions of Masters et al. (2012), McGreer et al. (2013), and Ross et al. (2013) are also plotted by the thick black lines at the relevant redshift intervals.

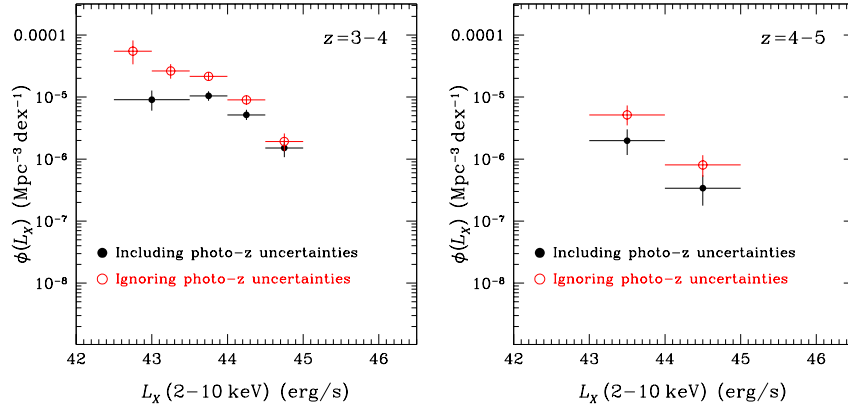


FIGURE 5.11: Impact of ignoring photometric redshift uncertainties on the inferred AGN space density. The red open circles correspond to the non-parametric binned X-ray luminosity function estimated by using the best-fit photometric redshift solution only and ignoring photometric redshift uncertainties (see text for details). The black data points correspond to the AGN space density estimated by using the full photometric redshift Probability Distribution Function to account for photometric redshift errors. For both set of data-points only the *Chandra* survey fields are used.

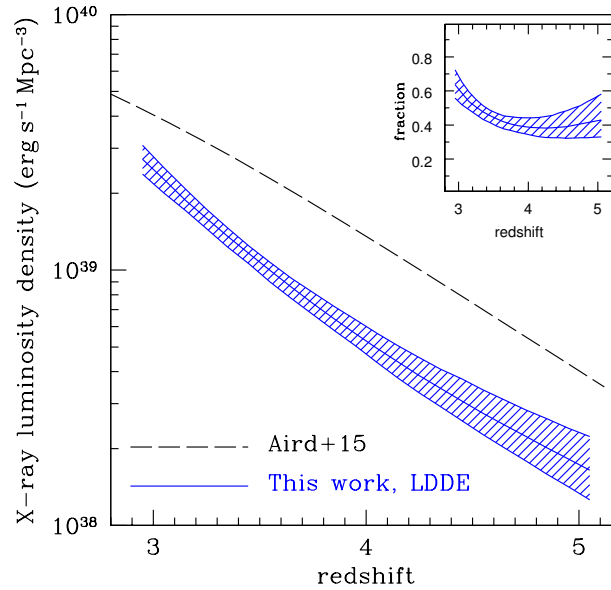


FIGURE 5.12: X-ray luminosity density as a function of redshift. The black dashed curve corresponds to the total X-ray luminosity density for both obscured (Compton thin and Compton thick) and unobscured AGN determined by Aird et al. (2015). The shaded blue region is the X-ray luminosity density estimated using our LDDE parametrisation for the evolution of the AGN population at $z > 3$. The inset plot shows the ratio between the blue (solid) and black (dashed) curves as a function of redshift. It represents the fraction of the luminosity density accounted by our LDDE model relative to the total luminosity density (corrected for obscuration effects) estimated by the Aird et al. (2015).

the Inter-Galactic Medium. The rate of hydrogen ionising photons is estimated by integrating the AGN Spectral Energy Distribution in the energy range 1-4 ryd

$$\dot{n} = \int_{1\text{ryd}}^{4\text{ryd}} \frac{L(\nu)}{h\nu} d\nu, \quad (5.11)$$

where $L(\nu)$ is the AGN monochromatic luminosity at frequency ν . The Lusso, E. et al. (2010) $L_\nu(2\text{keV}) - L_\nu(2500\text{\AA})$ relation is used to convert the 2-10 keV X-ray luminosity to UV monochromatic luminosity at 2500 Å. We then extrapolate to the wavelength range 227 Å (4 ryd) and 910 Å (1 ryd) assuming a double power law for the AGN SED of the form

$$L(\nu) \propto \begin{cases} \nu^{-0.5} & (1100\text{\AA} < \lambda < 2500\text{\AA}) \\ \nu^{-1.7} & \lambda < 1100\text{\AA} \end{cases} \quad (5.12)$$

The spectral slopes in the above relation are from Berk et al. (2001) and Telfer et al. (2002). At any given redshift the comoving density of the hydrogen ionising rate is then estimated by integrating the X-ray luminosity function

$$\mathcal{N} = \int \phi(L_X, z) f_{esc}(L_X, z) \dot{n} d\log L_X. \quad (5.13)$$

The LDDE parametrisation of the X-ray AGN luminosity function is adopted. The integration limits are set to $L_X(2 - 10\text{ keV}) = 10^{42}$ and $10^{46} \text{ erg s}^{-1}$. The photon escaping factor, $f_{esc}(L_X, z)$, accounts for the fraction of obscured AGN, which likely depends on both redshift and accretion luminosity. In these sources the ionising photons are absorbed locally and therefore do not contribute to the cosmic ionisation radiation field. Our analysis does not constrain the distribution of AGN in obscuration. Additionally at redshift $z > 3$ there are still discrepancies among different studies on the obscured AGN fraction and its dependence on luminosity (e.g. Ueda et al., 2014; Buchner et al., 2015; Aird et al., 2015). For example, in Figure 5.13 we find evidence that the type-I AGN fraction and by proxy the obscured AGN fraction, have a different dependence on luminosity compared to relations established at lower redshift. We choose to present our baseline results for the luminosity dependent type-I AGN fraction determined by Merloni et al. (2014). In that study type-I refers to AGN with either blue UV/optical continua and/or broad emission lines. This definition is more relevant to the determination of the hydrogen ionising photon rates of AGN, compared to e.g. the standard X-ray unobscured AGN definition, $N_H < 10^{22} \text{ cm}^{-2}$. Adopting the Merloni et al. (2014) relation for $f_{esc}(L_X, z)$ also allows direct comparison with previous X-ray studies that also used monotonically increasing obscured

AGN fraction with decreasing X-ray luminosity to approximate the escape fraction of UV photons in high redshift AGN. Figure 5.14 plots \mathcal{N} as a function of redshift for our baseline model for the photon escaping fraction. In that figure we also place an upper limit in \mathcal{N} by setting $f_{esc} = 1$, i.e. the extreme case that all photons escape and contribute to the hydrogen ionising radiation. This translates to a net increase of \mathcal{N} by a factor of about 1.4 compared to the Merloni et al. (2014) type-I AGN relation for f_{esc} . We further explore how the results in Figure 5.14 change if f_{esc} is approximated by X-ray definitions of the unobscured ($N_H < 10^{22} \text{ cm}^{-2}$) AGN fraction. We adopt the luminosity dependence of the unobscured X-ray AGN fraction derived by Buchner et al., 2015 in the redshift interval $z = 2.7 - 4$ (see Fig. 5.13). This assumption yields photon rate densities that are a factor of about 1.7 smaller compared to the baseline results that use the Merloni et al. (2014) type-I AGN fraction as proxy of f_{esc} . For clarity these results are not plotted in Figure 5.14.

The constraints above should be compared to the minimum photon rate density required to keep the Universe ionised at a given redshift. This is estimated from equation (26) of Madau, Haardt, and Rees (1999) after scaling it to our cosmology. We also assume that the ionised hydrogen clumping factor in that relation, which is a measure of the inhomogeneity of the medium, evolves with redshift as

$$C = 1 + 43 \times z^{-1.71}. \quad (5.14)$$

This relation is based on cosmological simulations (Pawlik, Schaye, and Van Scherpenzeel, 2009) and is adopted by Haardt and Madau (2012) to synthesise the evolving spectrum of the diffuse UV/X-ray radiation field. Figure 5.14 plots the redshift dependence of the photon rate density required to keep the Universe ionised. This figure shows that for the baseline model Merloni et al. (i.e. 2014, as proxy of f_{esc}) the AGN contribution to the photon rate density required to keep the Universe ionised decreases from 70% at $z = 4$ to about 20% at $z = 5$. Assuming $f_{esc} = 1$ the fractions above translate to upper limits $< 100\%$ at $z = 4$ and $< 30\%$ at $z = 5$. These numbers are in broad agreement with some previous studies on the role of X-ray AGN in the ionisation of the Universe (Barger et al., 2003b; Haardt and Madau, 2012; Grissom, R. L., Ballantyne, D. R., and Wise, J. H., 2014). The decreasing ionising photon rate density fractions with increasing redshift is a direct consequence of the strong evolution of the X-ray AGN luminosity function between redshifts $z = 3$ and $z = 5$.

Finally in Figure 5.14 the model constraints on the ionising photon rate density at $z = 3 - 5$ from our X-ray sample are compared to previous works based on either UV/optical QSO surveys (Glikman et al., 2011; Masters et al., 2012; McGreer et al.,

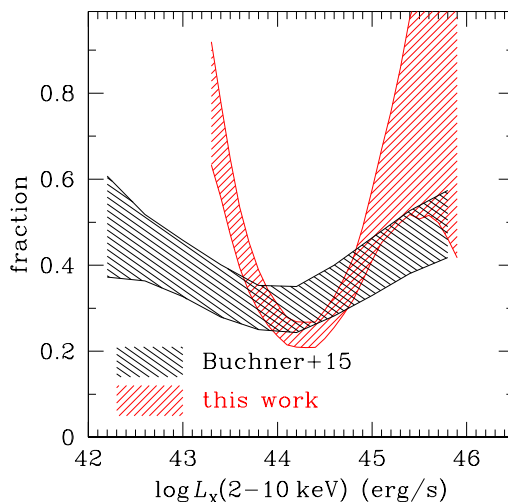


FIGURE 5.13: Type-I AGN fraction as a function of X-ray luminosity. The red shaded region is the ratio between the $z = 3.2$ luminosity functions of UV/optical QSOs from Masters et al. (2012) and our LDDE parametrisation for X-ray AGN. The black shaded region is $1 - \mathcal{F}_{\text{obscured}}$, where $\mathcal{F}_{\text{obscured}}$ is the obscured AGN fraction of Buchner et al. (2015) for the redshift interval $z = 2.7 - 4$.

2013) or UV/X-ray selected samples (Giallongo, E. et al., 2015). For the calculation of the ionising photon rate densities the luminosity functions estimated in these works are integrated between absolute magnitudes $M(1450\text{\AA}) = -18$ and -28 mag by adopting the UV spectrum of Equation 5.12. Our baseline model assuming the Merloni et al. (2014) escaping fraction is consistent with the lower-normalisation UV/optical data points in Figure 5.14.

5.5 Discussion

In this paper we explore the evolution of the AGN X-ray luminosity function in the redshift interval $z = 3 - 5$ by combining deep *Chandra* surveys with the wide-area and shallow XMM-XXL sample. This dataset provides sufficient depth and volume to determine the space density of AGN over 3 dex in luminosity, $\log L_X(2-10 \text{ keV}) \approx 43 - 46$, in the redshift intervals $z = 3 - 4$ and $4 - 5$. The analysis methodology we develop takes into account Poisson errors in the determination of X-ray fluxes and luminosities as well as uncertainties in photometric redshift measurements. We demonstrate that the latter is critical for unbiased measurements of the AGN space density at high redshift. Ignoring photometric redshift errors overestimates the X-ray luminosity function. We also choose to follow a non-parametric approach and determine the space density of AGN in luminosity and redshift bins. This allows us

TABLE 5.4: Parametric model best-fit parameters and uncertainties

Model	$\log K$ (Mpc ⁻³)	$\log L_*$ (erg/s)	γ_1	γ_2	p	q	β	$\log_{10} \mathcal{Z}$	$\Delta \log_{10} \mathcal{Z}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PDE	$-4.84^{+0.13}_{-0.16}$	$44.33^{+0.13}_{-0.12}$	$0.16^{+0.15}_{-0.11}$	$2.03^{+0.21}_{-0.20}$	—	$-6.59^{+0.90}_{-0.93}$	—	9894.95	0.0
LDDE	$-4.79^{+0.14}_{-0.17}$	$44.31^{+0.13}_{-0.11}$	$0.21^{+0.15}_{-0.13}$	$2.15^{+0.24}_{-0.21}$	—	$-7.46^{+1.03}_{-1.12}$	$2.30^{+1.60}_{-1.53}$	9894.70	-0.25
LADE	$-4.78^{+0.14}_{-0.17}$	$44.27^{+0.14}_{-0.13}$	$0.18^{+0.16}_{-0.12}$	$2.10^{+0.26}_{-0.22}$	$-0.85^{+0.17}_{-0.17}$	$2.25^{+1.47}_{-1.50}$	—	9893.62	-1.33
PLE	$-5.05^{+0.14}_{-0.20}$	$44.42^{+0.18}_{-0.14}$	$0.19^{+0.17}_{-0.13}$	$1.73^{+0.19}_{-0.15}$	$-4.98^{+0.80}_{-0.80}$	—	—	9890.93	-4.02

TABLE 5.5: Listed are the best-fit parameters for each of the four parametric models considered in this paper. The listed values are the median of the probability distribution function of each parameter. The errors correspond to the 16th and 84th percentiles around the median. The columns are: (1) parametric model. Models are listed in order of decreasing Bayesian evidence, (2) X-ray luminosity function normalisation (see equation 5.5), (3) break luminosity of the X-ray luminosity function (see equation 5.5), (4) faint-end slope, (5) bright-end slope, (6) density evolution parameter (see equations 5.7, 5.8, 5.9), (7) luminosity evolution parameters (see equations 5.6, 5.8), (8) β parameter for the LDDE model, (9) base 10 logarithm of the Bayesian evidence for each model, (10) the difference between the $\log_{10} \mathcal{Z}$ of each model and the PDE that has the highest evidence. The bayes factor of the PDE model relative to each of the other three is $\exp(\Delta \ln \mathcal{Z})$.

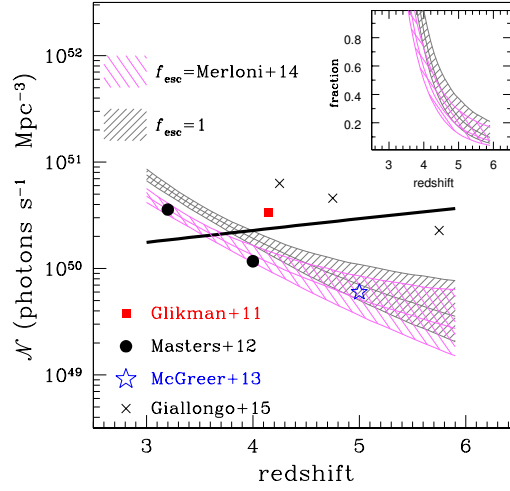


FIGURE 5.14: Hydrogen Ionising photon rate density as a function of redshift. The shaded regions are the constraints from our analysis using the LDDE parametrisation for the X-ray luminosity function and under different assumptions on the escape fraction of AGN photons. The grey-shaded region assumes an escaping fraction of unity, i.e. ignoring obscuration effects close to the supermassive black hole. The pink-shaded region assumes the luminosity-dependent Type-1 AGN fraction of Merloni et al. (2014). We caution that beyond $z = 5$ the shaded curves are extrapolations. The data points correspond to results in literature. The thick black line in the plot shows the photon rate density required to keep the Universe ionised at any given redshift. The ratio between the shaded regions and the black line are presented in the inset plot.

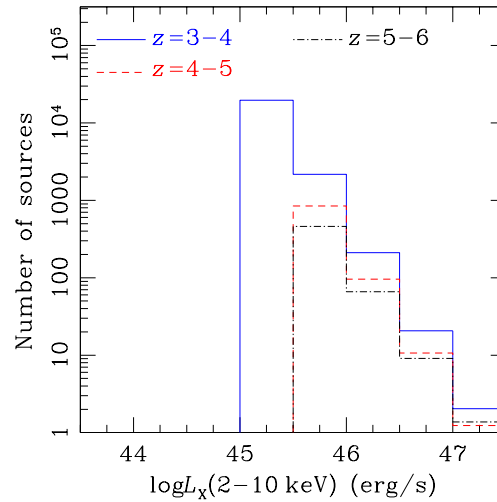


FIGURE 5.15: Number of high redshift AGN as a function of 2-10 keV X-ray luminosity that the eROSITA 4-year All Sky Survey is expected to detect. The LDDE parametrisation of the X-ray luminosity function is used for the predictions. Results for 3 redshift intervals are plotted, $z = 3 - 4$ (blue solid), $z = 4 - 5$ (red dashed) and $z = 5 - 6$ (black dot-dashed). The predictions for the latter redshift bin are extrapolations of the LDDE model. The numbers are for 0.5 dex wide luminosity bins.

to explore the evolution of the AGN population independent of model assumptions. Additionally when a model parametrisation is applied to AGN at all redshifts the fit may be driven by the redshift and luminosity intervals that contain most data. This can introduce biases at high redshift, $z \gtrsim 3$, where AGN samples are typically small as a result of the rapid evolution of the AGN population as well as survey sensitivity and volume limitations. We account for this potential issue by splitting the luminosity function into independent redshift components, i.e. $z < 3$, $z = 3 - 5$.

We confirm previous studies for a strong evolution of the AGN population in the redshift interval $z = 3 - 5$ (Brusa, M. et al., 2009; Vito et al., 2013; Civano et al., 2012; Kalfountzou et al., 2014). We also find suggestive evidence for luminosity dependent evolution of the X-ray luminosity function. The space density of AGN in the luminosity interval $\log L_X(2 - 10 \text{ keV}) \approx 43 - 45 \text{ erg s}^{-1}$ decreases faster than more luminous sources between redshifts $z = 3.5$ and $z = 4.5$, albeit at the 90% significance level. A similar evolution pattern is also observed in the optical luminosity function of QSOs between redshifts $z = 3$ and $z = 4$ (Masters et al., 2012). Our finding can be interpreted as evidence that the formation epoch of the most powerful QSOs [$L_X(2 - 10 \text{ keV}) \gtrsim 10^{45} \text{ erg s}^{-1}$] precedes that of lower luminosity systems. This is similar to AGN downsizing trends established at lower redshifts (Ueda et al., 2003; Hasinger, G., 2008; Aird et al., 2010; Miyaji et al., 2015). A strong evolution of the faint-end of the AGN luminosity function with increasing redshift is consistent with the absence of X-ray selected AGN at $z \gtrsim 5$ in the CANDELS (Grogin et al., 2011; Koekemoer et al., 2011) subregion of the Chandra Deep Field South (Weigel et al., 2015). Extrapolating the LDDE model of Table 5.4 we predict < 1 AGN at redshift $z > 5$ in that field.

Our analysis also places limits on the contribution of AGN to the UV photon field needed to keep the hydrogen ionised at high redshift. Using empirical relations for the type-1 AGN fraction as a function of luminosity Merloni et al. (2014), we show that AGN dominate or at least contribute a sizable fraction of the required UV photons to redshift $z \approx 4$. At higher redshift the evolution of the X-ray luminosity function translates to a decreasing contribution of X-ray AGN to the UV photon field required to keep the hydrogen ionised. The extreme assumption of a photon escaping fraction of unity for all AGN sets an upper limit of 30% to the contribution of AGN to the UV photon rate density required to keep the hydrogen ionised at $z = 5$. Barger et al. (2003b) use multicolour optical data in the 2 Ms *Chandra Deep Field North* (Barger et al., 2003a) and conclude that the X-ray selected AGN candidates at $z = 5 - 6.5$ are too few to ionise the intergalactic medium at those redshifts.

Haardt and Madau (2012) estimated the contribution of AGN to hydrogen ionisation rate using the Ueda et al. (2003) X-ray luminosity function and AGN obscuration distribution. They found that AGN do not play an important role as a source of ionising photons above redshifts ≈ 4 . Grissom, R. L., Ballantyne, D. R., and Wise, J. H. (2014) determine the contribution of AGN to the ionisation of the hydrogen in the Universe by taking into account secondary collisional ionisations from the X-ray radiation. They extrapolate to high redshift ($z \gtrsim 6$) the Hiroi et al. (2012) hard X-ray luminosity function and conclude that AGN only contribute a small fraction of the photon rate densities required to ionise the Universe at these redshifts. Our results are in agreement with the above studies and do not support claims for a dominant role of AGN to the ionisation of the hydrogen in the Universe at redshift $z \gtrsim 4$ (Glikman et al., 2011; Giallongo, E. et al., 2015). This discrepancy is likely related to the way different groups select their samples and subsequently account for this selection in the analysis. It also highlights the need for further research to better constrain the impact of AGN radiation to the ionisation of the Universe. Glikman et al. (2011) estimate type-I QSO space densities at $z \approx 4$ that are a factor of 3-4 higher than those determined by Masters et al. (2012) or Ikeda et al. (2011) at similar redshifts and luminosities. Giallongo, E. et al. (2015) combined X-ray and HST optical/near-IR data in the CANDELS GOODS-S region to identify optical sources with photometric or spectroscopic redshifts $z > 4$ and then study their X-ray properties following methods described by Fiore, F. et al. (2012). Their approach allows them to identify faint AGN with X-ray luminosities as low as $L_X \approx 10^{43} \text{ erg s}^{-1}$ in the redshift interval $z = 4 - 6.5$. Strictly speaking the Giallongo, E. et al. (2015) photon-rate densities in Figure 5.14 are upper limits. The UV photon escape fraction is set to one, part of the observed UV radiation of individual sources may be associated with the AGN host galaxy, the photometric redshift uncertainties of at least some sources in the sample are large. The increase of the X-ray depth in the CANDELS GOODS-S from 4 to 7 Ms (PI Brandt) region will help better constrain the faint-end of the AGN luminosity function at high redshift and their role in the ionisation of the Universe.

Finally the parametric X-ray luminosity functions derived in this chapter are used to make predictions on the number of $z > 3$ AGN that eROSITA (Merloni et al., 2012) surveys will detect. Our LDDE parametrisation is convolved with the expected X-ray sensitivity of the 4-year eROSITA All Sky Survey. The number of AGN in logarithmic luminosity bins of size $\Delta \log L_X = 0.5$ is plotted as a function of 2-10 keV luminosity in Figure 5.15. Predictions are presented for 3 redshift intervals, $z = 3-4$, $4-5$ and $5-6$. This shows that surveys by eROSITA will provide tight constraints on the

evolution of bright AGN and will allow us to explore with high statistical significance the evidence for luminosity dependence of the AGN population at high redshift. This however, would also require a dedicated follow-up program to identify high redshift AGN among the eROSITA population. High multiplex spectroscopic facilities that are able to simultaneously observe large number of targets over a wide field of view are essential for eROSITA X-ray source follow-ups. The SDSS/BOSS spectrographs (Smee et al., 2013) at the Apache Point SDSS telescope (Gunn et al., 2006) and in the future the ESO/4MOST facility (4-metre Multi-Object Spectroscopic Telescope, Jong et al., 2014) are well suited for follow-up observations of the eROSITA sky.

5.6 Conclusions

X-ray data from *Chandra* deep surveys and the shallow/wide *XMM*-XXL sample are combined to explore the evolution of the X-ray luminosity function at high redshift, $z = 3 - 5$. Our analysis accounts for Poisson errors in the calculation of fluxes and luminosities as well as photometric redshift uncertainties. We also show that the latter point is crucial for unbiased AGN space density measurements. The sample used in the paper consists of nearly 340 sources with either photometric (212) or spectroscopic (128) redshift in the redshift range $z = 3 - 5$. The luminosity baseline of the sample is $L_X(2 - 10 \text{ keV}) \approx 10^{43} - 10^{46} \text{ erg s}^{-1}$ at $z > 3$. Our main findings are as follows.

- (i) The AGN population evolves strongly between the redshift intervals $z = 3 - 4$ and $z = 4 - 5$
- (ii) There is evidence, significant at the 90% level, that the amplitude of the AGN evolution depends on X-ray luminosity. Sources with luminosities $L_X(2 - 10 \text{ keV}) < 10^{45} \text{ erg s}^{-1}$ appear to evolve faster than brighter ones.
- (iii) The faint-end slope of UV/optical QSO luminosity functions is steeper than that of the X-ray selected AGN samples. This implies an increasing fraction of type-I AGN with decreasing X-ray luminosity at $z > 3$.
- (iv) X-ray AGN may dominate or at least contribute substantially to the UV photon rate density required to keep the Universe ionised to $z = 4$. At higher redshift the contribution of AGN to UV hydrogen ionising field decreases.

Bibliography

- Aalborg, H.A., P. Molnár, and J.E. de Vries (2019). “What can explain the price, volatility and trading volume of Bitcoin?” In: *Finance Research Letters* 29, pp. 255–265. DOI: [10.1016/j.frl.2018.08.010](https://doi.org/10.1016/j.frl.2018.08.010).
- Aggarwal, D. (2019). “Do Bitcoins follow a random walk model?” In: *Research in Economics* 73.1, pp. 15–22. DOI: [10.1016/j.rie.2019.01.002](https://doi.org/10.1016/j.rie.2019.01.002).
- Aihara, Hiroaki et al. (2011). “The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III”. In: *The Astrophysical Journal Supplement Series* 193.2, p. 29. DOI: [10.1088/0067-0049/193/2/29](https://doi.org/10.1088/0067-0049/193/2/29).
- Aird, J. et al. (Jan. 2010). “The evolution of the hard X-ray luminosity function of AGN”. In: *Monthly Notices of the Royal Astronomical Society* 401.4, pp. 2531–2551. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2009.15829.x](https://doi.org/10.1111/j.1365-2966.2009.15829.x).
- Aird, J. et al. (June 2015). “The evolution of the X-ray luminosity functions of unabsorbed and absorbed AGNs out to $z \sim 5$ ”. In: *Monthly Notices of the Royal Astronomical Society* 451.2, pp. 1892–1927. ISSN: 0035-8711. DOI: [10.1093/mnras/stv1062](https://doi.org/10.1093/mnras/stv1062).
- Aird, James et al. (2012). “PRIMUS: The Dependence of AGN Accretion on Host Stellar Mass and Color”. In: *The Astrophysical Journal* 746.1, p. 90. DOI: [10.1088/0004-637x/746/1/90](https://doi.org/10.1088/0004-637x/746/1/90).
- Akaike, Hirotogu, BN Petrov, and F Csaki (1973). “Information theory and an extension of the maximum likelihood principle”. In: *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281.
- Akylas, A. et al. (2006). “XMM-Newton and Chandra measurements of the AGN intrinsic absorption: dependence on luminosity and redshift”. In: *Astronomy & Astrophysics* 459.3, pp. 693–701. DOI: [10.1051/0004-6361:20054632](https://doi.org/10.1051/0004-6361:20054632).
- Alexander, D. M. et al. (2003). “The Chandra Deep Field North Survey. XIII. 2 Ms Point-Source Catalogs”. In: *The Astronomical Journal* 126.2, pp. 539–574. DOI: [10.1086/376473](https://doi.org/10.1086/376473).
- Andersen, Torben G. et al. (2003). “Modeling and Forecasting Realized Volatility”. In: *Econometrica* 71.2, pp. 579–62. DOI: [10.1111/1468-0262.00418](https://doi.org/10.1111/1468-0262.00418).

- Ang, Andrew and Allan Timmermann (2012). “Regime Changes and Financial Markets”. In: *Annual Review of Financial Economics* 4.1, pp. 313–337. doi: [10.1146/annurev-financial-110311-101808](https://doi.org/10.1146/annurev-financial-110311-101808).
- Angelidis, Timotheos and Nikolaos Tassaromatis (2009). “Idiosyncratic risk matters! A regime switching approach”. In: *International Review of Economics & Finance* 18.1. Foreign direct Investment: Greenfield Investments, Mergers and Acquisitions, and Joint Ventures, pp. 132 –141. ISSN: 1059-0560. doi: [10.1016/j.iref.2007.09.003](https://doi.org/10.1016/j.iref.2007.09.003).
- Ardia, David, Keven Bluteau, and Maxime Rüede (2019). “Regime changes in Bitcoin GARCH volatility dynamics”. In: *Finance Research Letters* 29, pp. 266 – 271. ISSN: 1544-6123. doi: [10.1016/j.frl.2018.08.009](https://doi.org/10.1016/j.frl.2018.08.009).
- Assef, R. J. et al. (2013). “Mid-infrared Selection of Active Galactic Nuclei with the Wide-field Infrared Survey Explorer. II. Properties of WISE-selected Active Galactic Nuclei in the NDWFS Bootes Field”. In: *The Astrophysical Journal* 772.1, p. 26. doi: [10.1088/0004-637x/772/1/26](https://doi.org/10.1088/0004-637x/772/1/26).
- Assef, R. J. et al. (2015). “Half of the most luminous quasars may be obscured: Investigating the nature of wise-selected hot dust-obscured galaxies”. In: *The Astrophysical Journal* 804.1, p. 27. doi: [10.1088/0004-637x/804/1/27](https://doi.org/10.1088/0004-637x/804/1/27).
- Aste, Tomaso (2019). “Cryptocurrency market structure: connecting emotions and economics”. In: *Digital Finance* 1.1. ISSN: 2524-6186. doi: [10.1007/s42521-019-00008-9](https://doi.org/10.1007/s42521-019-00008-9).
- Baillie, Richard T. et al. (2019). “Long Memory, Realized Volatility and Heterogeneous Autoregressive Models”. In: *Journal of Time Series Analysis*. doi: [10.1111/jtsa.12470](https://doi.org/10.1111/jtsa.12470).
- Balcilar, M. et al. (2017). “Can volume predict Bitcoin returns and volatility? A quantiles-based approach”. In: *Economic Modelling* 64, pp. 74–81. doi: [10.1016/j.econmod.2017.03.019](https://doi.org/10.1016/j.econmod.2017.03.019).
- Barbieri, Maria Maddalena and James O. Berger (2004). “Optimal predictive model selection”. In: *The Annals of Statistics* 32.3, pp. 870–897. doi: [10.1214/009053604000000238](https://doi.org/10.1214/009053604000000238).
- Barger, A. J. et al. (2003a). “Optical and Infrared Properties of the 2 Ms Chandra Deep Field North X-Ray Sources”. In: *The Astronomical Journal* 126.2, pp. 632–665. doi: [10.1086/376843](https://doi.org/10.1086/376843).
- Barger, A. J. et al. (2003b). “Very High Redshift X-Ray–selected Active Galactic Nuclei in the Chandra Deep Field–North”. In: *The Astrophysical Journal* 584.2, pp. L61–L64. doi: [10.1086/368407](https://doi.org/10.1086/368407).

- Baum, Leonard E. et al. (1970). "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". In: *The Annals of Mathematical Statistics* 41.1, pp. 164–171. doi: [10.1214/aoms/1177697196](https://doi.org/10.1214/aoms/1177697196).
- Baur, D. G. and T. Dimpfl (2018). "Asymmetric volatility in cryptocurrencies". In: *Economics Letters* 173, pp. 148–151. issn: 0165-1765. doi: [10.1016/j.econlet.2018.10.008](https://doi.org/10.1016/j.econlet.2018.10.008).
- Baur, Dirk G., Thomas Dimpfl, and Konstantin Kuck (2018). "Bitcoin, gold and the US dollar – A replication and extension". In: *Finance Research Letters* 25, pp. 103–110. issn: 1544-6123. doi: [10.1016/j.frl.2017.10.012](https://doi.org/10.1016/j.frl.2017.10.012).
- Beckmann, J. and R. Schüssler (2016). "Forecasting exchange rates under parameter and model uncertainty". In: *Journal of International Money and Finance* 60, pp. 267–288. doi: [10.1016/j.jimonfin.2015.07.001](https://doi.org/10.1016/j.jimonfin.2015.07.001).
- Bergman, U. Michael and Jesper Hansson (2005). "Real exchange rates and switching regimes". In: *Journal of International Money and Finance* 24.1, pp. 121–138. issn: 0261-5606. doi: [10.1016/j.jimonfin.2004.10.002](https://doi.org/10.1016/j.jimonfin.2004.10.002).
- Berk, Daniel E. Vanden et al. (Aug. 2001). "Composite Quasar Spectra from the Sloan Digital Sky Survey". In: *The Astronomical Journal* 122.2, pp. 549–564. doi: [10.1086/321167](https://doi.org/10.1086/321167).
- Besag, Julian et al. (1995). "Bayesian Computation and Stochastic Systems". In: *Statistical Science* 10.1, pp. 3–41. doi: [10.1214/ss/1177010123](https://doi.org/10.1214/ss/1177010123).
- Billio, M., A. Monfort, and C.P. Robert (1999). "Bayesian estimation of switching ARMA models". In: *Journal of Econometrics* 93.2, pp. 229–255. issn: 0304-4076. doi: [10.1016/S0304-4076\(99\)00010-X](https://doi.org/10.1016/S0304-4076(99)00010-X).
- Blau, Benjamin M. (2017). "Price dynamics and speculative trading in bitcoin". In: *Research in International Business and Finance* 41, pp. 493–499. issn: 0275-5319. doi: [10.1016/j.ribaf.2017.05.010](https://doi.org/10.1016/j.ribaf.2017.05.010).
- Bleher, Johannes and Thomas Dimpfl (2019). "Today I got a million, tomorrow, I don't know: On the predictability of cryptocurrencies by means of Google search volume". In: *International Review of Financial Analysis* 63, pp. 147–159. issn: 1057-5219. doi: [10.1016/j.irfa.2019.03.003](https://doi.org/10.1016/j.irfa.2019.03.003).
- Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2011). "Scoring rules and survey density forecasts". In: *International Journal of Forecasting* 27.2, pp. 379–393. doi: [10.1016/j.ijforecast.2010.04.003](https://doi.org/10.1016/j.ijforecast.2010.04.003).

- Bollen, Nicolas P.B., Stephen F. Gray, and Robert E. Whaley (2000). “Regime switching in foreign exchange rates:: Evidence from currency option prices”. In: *Journal of Econometrics* 94.1, pp. 239 –276. ISSN: 0304-4076. DOI: [10.1016/S0304-4076\(99\)00022-6](https://doi.org/10.1016/S0304-4076(99)00022-6).
- Bongiorno, A. et al. (Dec. 2012). “Accreting supermassive black holes in the COSMOS field and the connection to their host galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 427.4, pp. 3103–3133. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2012.22089.x](https://doi.org/10.1111/j.1365-2966.2012.22089.x).
- Borri, N. (2019). “Conditional tail-risk in cryptocurrency markets”. In: *Journal of Empirical Finance* 50, pp. 1–19. DOI: [10.1016/j.jempfin.2018.11.002](https://doi.org/10.1016/j.jempfin.2018.11.002).
- Bouri, E., G. Azzi, and A. H. Dyhrberg (2017). “On the return-volatility relationship in the Bitcoin market around the price crash of 2013”. In: *Economics: The Open-Access, Open-Assessment E-Journal* 11.2017-2, pp. 1–16. DOI: [10.5018/economics-ejournal.ja.2017-2](https://doi.org/10.5018/economics-ejournal.ja.2017-2).
- Bouri, E. et al. (2017). “On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?” In: *Finance Research Letters* 20, pp. 192–198. DOI: [10.1016/j.frl.2016.09.025](https://doi.org/10.1016/j.frl.2016.09.025).
- Bouri, E. et al. (2019a). “Trading volume and the predictability of return and volatility in the cryptocurrency market”. In: *Finance Research Letters* 29, pp. 340–346. DOI: [10.1016/j.frl.2018.08.015](https://doi.org/10.1016/j.frl.2018.08.015).
- Bouri, Elie, Rangan Gupta, and David Roubaud (2019). “Herding behaviour in cryptocurrencies”. In: *Finance Research Letters* 29, pp. 216 –221. ISSN: 1544-6123. DOI: [0.1016/j.frl.2018.07.008](https://doi.org/10.1016/j.frl.2018.07.008).
- Bouri, Elie et al. (2019b). “Modelling long memory volatility in the Bitcoin market: Evidence of persistence and structural breaks”. In: *International Journal of Finance & Economics* 24.1, pp. 412–426. DOI: [10.1002/ijfe.1670](https://doi.org/10.1002/ijfe.1670).
- Bouri, Elie et al. (2020). “Bitcoin, gold, and commodities as safe havens for stocks: New insight through wavelet analysis”. In: *The Quarterly Review of Economics and Finance*. ISSN: 1062-9769. DOI: [0.1016/j.qref.2020.03.004](https://doi.org/10.1016/j.qref.2020.03.004).
- Brandt, W. N. and D. M. Alexander (Jan. 2015). “Cosmic X-ray surveys of distant active galaxies”. In: *The Astronomy and Astrophysics Review* 23.1. ISSN: 1432-0754. DOI: [10.1007/s00159-014-0081-z](https://doi.org/10.1007/s00159-014-0081-z).
- Brauneis, A. and R. Mestel (2018). “Price discovery of cryptocurrencies: Bitcoin and beyond”. In: *Economics Letters* 165, pp. 58–61. DOI: [10.1016/j.econlet.2018.02.001](https://doi.org/10.1016/j.econlet.2018.02.001).

- Brightman, Murray and Kirpal Nandra (Apr. 2011a). “An XMM–Newton spectral survey of 12 μm selected galaxies -I. X-ray data”. In: *Monthly Notices of the Royal Astronomical Society* 413.2, pp. 1206–1235. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2011.18207.x](https://doi.org/10.1111/j.1365-2966.2011.18207.x).
- (July 2011b). “An XMM–Newton spectral survey of 12 μm selected galaxies – II. Implications for AGN selection and unification”. In: *Monthly Notices of the Royal Astronomical Society* 414.4, pp. 3084–3104. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2011.18612.x](https://doi.org/10.1111/j.1365-2966.2011.18612.x).
- Brooks, Stephen P. and Andrew Gelman (1998). “General Methods for Monitoring Convergence of Iterative Simulations”. In: *Journal of Computational and Graphical Statistics* 7.4, pp. 434–455. doi: [10.1080/10618600.1998.10474787](https://doi.org/10.1080/10618600.1998.10474787).
- Brusa, M. et al. (2009). “Black hole growth and starburst activity at $z = 0.6\text{--}4$ in the Chandra Deep Field South* - Host galaxies properties of obscured AGN”. In: *A&A* 507.3, pp. 1277–1289. doi: [10.1051/0004-6361/200912261](https://doi.org/10.1051/0004-6361/200912261).
- Buchner, Johannes et al. (2015). “Obscuration-dependent Evolution of Active Galactic Nuclei”. In: *The Astrophysical Journal* 802.2, p. 89. doi: [10.1088/0004-637x/802/2/89](https://doi.org/10.1088/0004-637x/802/2/89).
- Buchner, J. et al. (2014). “X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue”. In: *Astronomy & Astrophysics* 564, A125. doi: [10.1051/0004-6361/201322971](https://doi.org/10.1051/0004-6361/201322971).
- Budavári, Tamás and Alexander S. Szalay (2008). “Probabilistic Cross-Identification of Astronomical Sources”. In: *The Astrophysical Journal* 679.1, pp. 301–309. doi: [10.1086/587156](https://doi.org/10.1086/587156).
- Burlon, D. et al. (2011). “Three-year Swift-BAT Survey of Active Galactic Nuclei: Reconciling Theory and Observations?” In: *The Astrophysical Journal* 728.1, p. 58. doi: [10.1088/0004-637x/728/1/58](https://doi.org/10.1088/0004-637x/728/1/58).
- Buterin, V. (2014). *A Next-Generation Smart Contract and Decentralized Application Platform*. Available at [GitHub/Ethereum/Wiki](https://github.com/Ethereum/Wiki).
- Buterin, Vitalik et al. (2020). “Incentives in Ethereum’s hybrid Casper protocol”. In: *International Journal of Network Management*, e2098. doi: [10.1002/nem.2098](https://doi.org/10.1002/nem.2098).
- Canela, M. A. and E. P. Collazo (2007). “Portfolio selection with skewness in emerging market industries”. In: *Emerging Markets Review* 8.3, pp. 230–250. doi: [10.1016/j.ememar.2006.03.001](https://doi.org/10.1016/j.ememar.2006.03.001).

- Caporale, Guglielmo Maria and Timur Zekokh (2019). “Modelling volatility of cryptocurrencies using Markov-Switching GARCH models”. In: *Research in International Business and Finance* 48, pp. 143 –155. ISSN: 0275-5319. DOI: [10.1016/j.ribaf.2018.12.009](https://doi.org/10.1016/j.ribaf.2018.12.009).
- Cappé, Olivier, Eric Moulines, and Tobias Ryden (2005). *Inference in Hidden Markov Models* (Springer Series in Statistics). Springer-Verlag New York, Inc. ISBN: 0387402640.
- Cappé, Olivier, Eric Moulines, and Tobias Rydén (2006). *Inference in hidden Markov models*. Springer Science & Business Media.
- Cardamone, Carolin N. et al. (2010). “Dust-corrected Colors Reveal Bimodality in the Host-galaxy Colors of Active Galactic Nuclei at $z \sim 1$ ”. In: *The Astrophysical Journal* 721.1, pp. L38–L42. DOI: [10.1088/2041-8205/721/1/L38](https://doi.org/10.1088/2041-8205/721/1/L38).
- Catania, Leopoldo, Stefano Grassi, and Francesco Ravazzolo (2019). “Forecasting cryptocurrencies under model and parameter instability”. In: *International Journal of Forecasting* 35.2, pp. 485 –501. ISSN: 0169-2070. DOI: [10.1016/j.ijforecast.2018.09.005](https://doi.org/10.1016/j.ijforecast.2018.09.005).
- Chaim, P. and M. P. Laurini (2018). “Volatility and return jumps in Bitcoin”. In: *Economics Letters* 173, pp. 158–163. DOI: [10.1016/j.econlet.2018.10.011](https://doi.org/10.1016/j.econlet.2018.10.011).
- Chaim, Pedro and Márcio P. Laurini (2019). “Nonlinear dependence in cryptocurrency markets”. In: *The North American Journal of Economics and Finance* 48, pp. 32 –47. ISSN: 1062-9408. DOI: [10.1016/j.najef.2019.01.015](https://doi.org/10.1016/j.najef.2019.01.015).
- Charles, Amélie and Olivier Darné (2009). “Variance-Ratio tests of random walk: An overview”. In: *Journal of Economic Surveys* 23.3, pp. 503–527. DOI: [10.1111/j.1467-6419.2008.00570.x](https://doi.org/10.1111/j.1467-6419.2008.00570.x).
- Chauvet, Marcelle and James D. Hamilton (2006). “Chapter 1 Dating Business Cycle Turning Points”. In: *Nonlinear Time Series Analysis of Business Cycles*. Ed. by Costas Milas, Philip Rothman, and Dick [van Dijk]. Vol. 276. Contributions to Economic Analysis. Elsevier, pp. 1 –54. DOI: [10.1016/S0573-8555\(05\)76001-6](https://doi.org/10.1016/S0573-8555(05)76001-6).
- Cheah, E.-T. and J. Fry (2015). “Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin”. In: *Economics Letters* 130, pp. 32–36. DOI: [10.1016/j.econlet.2015.02.029](https://doi.org/10.1016/j.econlet.2015.02.029).
- Chen, Wei et al. (2020). “Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants”. In: *International Journal of Forecasting*. ISSN: 0169-2070. DOI: [10.1016/j.ijforecast.2020.02.008](https://doi.org/10.1016/j.ijforecast.2020.02.008).
- Chen, Zheshi, Chunhong Li, and Wenjun Sun (2020). “Bitcoin price prediction using machine learning: An approach to sample dimension engineering”. In: *Journal of*

- Computational and Applied Mathematics* 365, p. 112395. ISSN: 0377-0427. DOI: [10.1016/j.cam.2019.112395](https://doi.org/10.1016/j.cam.2019.112395).
- Cheung, Yin-Wong and Ulf G Erlandsson (2005). “Exchange Rates and Markov Switching Dynamics”. In: *Journal of Business & Economic Statistics* 23.3, pp. 314–320. DOI: [10.1198/073500104000000488](https://doi.org/10.1198/073500104000000488).
- Choi, Hee Min and James P. Hobert (2013). “The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic”. In: *Electron. J. Statist.* 7, pp. 2054–2064. DOI: [10.1214/13-EJS837](https://doi.org/10.1214/13-EJS837).
- Christiansen, Charlotte, Maik Schmeling, and Andreas Schrimpf (2012). “A comprehensive look at financial volatility prediction by economic variables”. In: *Journal of Applied Econometrics* 27.6, pp. 956–977. DOI: [10.1002/jae.2298](https://doi.org/10.1002/jae.2298).
- Chunhachinda, Pornchai et al. (1997). “Portfolio selection and skewness: Evidence from international stock markets”. In: *Journal of Banking & Finance* 21.2, pp. 143–167. ISSN: 0378-4266. DOI: [10.1016/S0378-4266\(96\)00032-5](https://doi.org/10.1016/S0378-4266(96)00032-5).
- Ciaian, P., M. Rajcaniova, and A. Kancs (2016). “The economics of BitCoin price formation”. In: *Applied Economics* 48.19, pp. 1799–1815. DOI: [10.1080/00036846.2015.1109038](https://doi.org/10.1080/00036846.2015.1109038).
- Civano, F. et al. (2010). “A Runaway Black Hole in COSMOS: Gravitational Wave or Slingshot Recoil?” In: *The Astrophysical Journal* 717.1, pp. 209–222. DOI: [10.1088/0004-637x/717/1/209](https://doi.org/10.1088/0004-637x/717/1/209).
- Civano, F. et al. (Oct. 2011). “The Population of High-redshift Active Galactic Nuclei in the Chandra-COSMOS Survey”. In: *The Astrophysical Journal* 741.2, p. 91. DOI: [10.1088/0004-637x/741/2/91](https://doi.org/10.1088/0004-637x/741/2/91).
- Civano, F. et al. (2012). “The Chandra COSMOS Survey. III. Optical and Infrared Identification of X-Ray Point Sources”. In: *The Astrophysical Journal* 201.2, p. 30. DOI: [10.1088/0067-0049/201/2/30](https://doi.org/10.1088/0067-0049/201/2/30).
- Clerc, N. et al. (Sept. 2014). “The XMM-LSS survey: the Class 1 cluster sample over the extended 11 deg² and its spatial distribution”. In: *Monthly Notices of the Royal Astronomical Society* 444.3, pp. 2723–2753. ISSN: 0035-8711. DOI: [10.1093/mnras/stu1625](https://doi.org/10.1093/mnras/stu1625).
- Colianni, S. G., S. M. Rosales, and M. Signorotti (2015). *Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis*. CS229 Project. Stanford University.
- Conrad, J., R. F. Dittmar, and E. Ghysels (2013). “Ex Ante Skewness and Expected Stock Returns”. In: *The Journal of Finance* 68.1, pp. 85–124. DOI: [10.1111/j.1540-6261.2012.01795.x](https://doi.org/10.1111/j.1540-6261.2012.01795.x).

- Corbet, S., B. Lucey, and L. Yarovaya (2018). “Datestamping the Bitcoin and Ethereum bubbles”. In: *Finance Research Letters* 26, pp. 81–88. doi: [10.1016/j.frl.2017.12.006](https://doi.org/10.1016/j.frl.2017.12.006).
- Corbet, S. et al. (2018). “Exploring the dynamic relationships between cryptocurrencies and other financial assets”. In: *Economics Letters* 165, pp. 28–34. doi: [10.1016/j.econlet.2018.01.004](https://doi.org/10.1016/j.econlet.2018.01.004).
- Corbet, Shaen and Paraskevi Katsiampa (2018). “Asymmetric mean reversion of Bitcoin price returns”. In: *International Review of Financial Analysis*. ISSN: 1057-5219. doi: [10.1016/j.irfa.2018.10.004](https://doi.org/10.1016/j.irfa.2018.10.004).
- Corbet, Shaen et al. (2019). “Cryptocurrencies as a financial asset: A systematic analysis”. In: *International Review of Financial Analysis* 62, pp. 182–199. ISSN: 1057-5219. doi: [10.1016/j.irfa.2018.09.003](https://doi.org/10.1016/j.irfa.2018.09.003).
- Cowie, L. L. et al. (2003). “The Redshift Evolution of the 2–8 [CLC]ke[CLC]V X-Ray Luminosity Function”. In: *The Astrophysical Journal* 584.2, pp. L57–L60. doi: [10.1086/368404](https://doi.org/10.1086/368404).
- Croton, Darren J. et al. (Jan. 2006). “The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 365.1, pp. 11–28. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2005.09675.x](https://doi.org/10.1111/j.1365-2966.2005.09675.x).
- Crypto.com (2019). Available at Coin Market Cap. Accessed: July 30, 2019.
- Davis, M. et al. (Apr. 2007). “The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets”. In: *The Astrophysical Journal* 660.1, pp. L1–L6. doi: [10.1086/517931](https://doi.org/10.1086/517931).
- Dawson, Kyle S. et al. (2013). “The Baryon Oscillation Spectroscopic Survey of SDSS-III”. In: *The Astronomical Journal* 145.1, p. 10. doi: [10.1088/0004-6256/145/1/10](https://doi.org/10.1088/0004-6256/145/1/10).
- de la Horra, Luis P., Gabriel de la Fuente, and Javier Perote (2019). “The drivers of Bitcoin demand: A short and long-run analysis”. In: *International Review of Financial Analysis* 62, pp. 21–34. ISSN: 1057-5219. doi: [10.1016/j.irfa.2019.01.006](https://doi.org/10.1016/j.irfa.2019.01.006).
- Del Moro, A. et al. (2009). “An extreme EXO: a type 2 QSO at $z=7$ ”. In: *Astronomy & Astrophysics* 493.2, pp. 445–451. doi: [10.1051/0004-6361/200809665](https://doi.org/10.1051/0004-6361/200809665).
- Dellaportas, Petros, Jonathan J. Forster, and Ioannis Ntzoufras (2002). “On Bayesian model and variable selection using MCMC”. In: *Statistics and Computing* 12.1, pp. 27–36. doi: [10.1023/A:1013164120801](https://doi.org/10.1023/A:1013164120801).

- Dellaportas, Petros and Ioulia Papageorgiou (2006). "Multivariate mixtures of normals with unknown number of components". In: *Statistics and Computing* 16.1, pp. 57–68. doi: [10.1007/s11222-006-5338-6](https://doi.org/10.1007/s11222-006-5338-6).
- Demir, E. et al. (2018). "Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation". In: *Finance Research Letters* 26, pp. 145–149. doi: [10.1016/j.frl.2018.01.005](https://doi.org/10.1016/j.frl.2018.01.005).
- Di Matteo, Tiziana, Volker Springel, and Lars Hernquist (Feb. 2007). "Energy input from quasars regulates the growth and activity of black holes and their host galaxies". In: *Nature* 443.1 (7026). doi: [10.1038/nature03335](https://doi.org/10.1038/nature03335).
- Dias, José G., Jeroen K. Vermunt, and Sofia Ramos (2015). "Clustering financial time series: New insights from an extended hidden Markov model". In: *European Journal of Operational Research* 243.3, pp. 852–864. ISSN: 0377-2217. doi: [10.1016/j.ejor.2014.12.041](https://doi.org/10.1016/j.ejor.2014.12.041).
- Diebold, Francis X, Joon-Haeng Lee, and Gretchen C Weinbach (1994). *Regime switching with time-varying transition probabilities*. In: *Hargreaves, C.P. (ed.) Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press, pp. 282–302.
- Dittmar, Robert F. (2002). "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns". In: *The Journal of Finance* 57.1, pp. 369–403. doi: [10.1111/1540-6261.00425](https://doi.org/10.1111/1540-6261.00425).
- Dixon, Matthew F et al. (2019). "Blockchain analytics for intraday financial risk modeling". In: *Digital Finance* 1.1, pp. 67–89. doi: [10.1007/s42521-019-00009-8](https://doi.org/10.1007/s42521-019-00009-8).
- Dyhrberg, A. H. (2016). "Bitcoin, gold and the dollar – A GARCH volatility analysis". In: *Finance Research Letters* 16, pp. 85–92. doi: [10.1016/j.frl.2015.10.008](https://doi.org/10.1016/j.frl.2015.10.008).
- Dyhrberg, Anne H., Sean Foley, and Jiri Svec (2018). "How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets". In: *Economics Letters* 171, pp. 140–143. ISSN: 0165-1765. doi: [10.1016/j.econlet.2018.07.032](https://doi.org/10.1016/j.econlet.2018.07.032).
- Ebrero, J. et al. (2009). "The XMM-Newton serendipitous survey - VI. The X-ray luminosity function". In: *Astronomy & Astrophysics* 493.1, pp. 55–69. doi: [10.1051/0004-6361:200810919](https://doi.org/10.1051/0004-6361:200810919).
- Elvis, Martin et al. (2009). "The Chandra COSMOS Survey. I. Overview and Point Source Catalog". In: *The Astrophysical Journal Supplement Series* 184.1, pp. 158–171. doi: [10.1088/0067-0049/184/1/158](https://doi.org/10.1088/0067-0049/184/1/158).

- Engel, C. (1994a). "Can the Markov switching model forecast exchange rates?" In: *Journal of International Economics* 36.1, pp. 151–165. doi: [10.1016/0022-1996\(94\)90062-0](https://doi.org/10.1016/0022-1996(94)90062-0).
- Engel, Charles (1994b). "Can the Markov switching model forecast exchange rates?" In: *Journal of International Economics* 36.1, pp. 151 –165. ISSN: 0022-1996. doi: [10.1016/0022-1996\(94\)90062-0](https://doi.org/10.1016/0022-1996(94)90062-0).
- Estrada, J. C. S. (2017). *Analyzing Bitcoin price volatility*. PhD Thesis. University of California, Berkeley.
- Fabian, A. C. (Oct. 1999). "The obscured growth of massive black holes". In: *Monthly Notices of the Royal Astronomical Society* 308.4, pp. L39–L43. ISSN: 0035-8711. doi: [10.1046/j.1365-8711.1999.03017.x](https://doi.org/10.1046/j.1365-8711.1999.03017.x).
- Fama, Eugene F. and Kenneth R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3 –56. doi: [10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5).
- Feroz, F. and M. P. Hobson (Jan. 2008). "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses". In: *Monthly Notices of the Royal Astronomical Society* 384.2, pp. 449–463. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2007.12353.x](https://doi.org/10.1111/j.1365-2966.2007.12353.x).
- Feroz, F., M. P. Hobson, and M. Bridges (Sept. 2009). "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics". In: *Monthly Notices of the Royal Astronomical Society* 398.4, pp. 1601–1614. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x).
- Ferrarese, Laura and David Merritt (2000). "A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies". In: *The Astrophysical Journal* 539.1, pp. L9–L12. doi: [10.1086/312838](https://doi.org/10.1086/312838).
- Filardo, Andrew J. and Stephen F. Gordon (1998). "Business cycle durations". In: *Journal of Econometrics* 85.1, pp. 99 –123. doi: [10.1016/S0304-4076\(97\)00096-1](https://doi.org/10.1016/S0304-4076(97)00096-1).
- Fiore, F. et al. (2012). "Faint high-redshift AGN in the Chandra deep field south: the evolution of the AGN luminosity function and black hole demography". In: *Astronomy & Astrophysics* 537, A16. doi: [10.1051/0004-6361/201117581](https://doi.org/10.1051/0004-6361/201117581).
- Fontanot, F. et al. (2007). "The luminosity function of high-redshift quasi-stellar objects. A combined analysis of GOODS and SDSS". In: *Astronomy & Astrophysics* 461.1, pp. 39–48. doi: [10.1051/0004-6361:20066073](https://doi.org/10.1051/0004-6361:20066073).

- Frühwirth-Schnatter, Sylvia (2001). “Markov chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models”. In: *Journal of the American Statistical Association* 96.453, pp. 194–209. doi: [10.1198/016214501750333063](https://doi.org/10.1198/016214501750333063).
- Frömmel, Michael, Ronald MacDonald, and Lukas Menkhoff (2005). “Markov switching regimes in a monetary exchange rate model”. In: *Economic Modelling* 22.3, pp. 485–502. ISSN: 0264-9993. doi: [10.1016/j.econmod.2004.07.001](https://doi.org/10.1016/j.econmod.2004.07.001).
- Frühwirth-Schnatter, Sylvia (2001). “Fully Bayesian Analysis of Switching Gaussian State Space Models”. In: *Annals of the Institute of Statistical Mathematics* 53.1, pp. 31–49. doi: [10.1023/A:1017908219076](https://doi.org/10.1023/A:1017908219076).
- Frühwirth-Schnatter, Sylvia and Rudolf Frühwirth (2010). *Data Augmentation and MCMC for Binary and Multinomial Logit Models*. Physica-Verlag HD, pp. 111–132. doi: [10.1007/978-3-7908-2413-1_7](https://doi.org/10.1007/978-3-7908-2413-1_7).
- Fry, J. (2018). “Booms, busts and heavy-tails: The story of Bitcoin and cryptocurrency markets?” In: *Economics Letters* 171, pp. 225–229. doi: [10.1016/j.econlet.2018.08.008](https://doi.org/10.1016/j.econlet.2018.08.008).
- Fussl, Agnes, Sylvia Frühwirth-Schnatter, and Rudolf Frühwirth (2013). “Efficient MCMC for Binomial Logit Models”. In: *ACM Trans. Model. Comput. Simul.* 23.1, 3:1–3:21. doi: [10.1145/2414416.2414419](https://doi.org/10.1145/2414416.2414419).
- Gallagher, S. C. et al. (2006). “An Exploratory Chandra Survey of a Well-defined Sample of 35 Large Bright Quasar Survey Broad Absorption Line Quasars”. In: *The Astrophysical Journal* 644.2, pp. 709–724. doi: [10.1086/503762](https://doi.org/10.1086/503762).
- Gandal, Neil et al. (2018). “Price manipulation in the Bitcoin ecosystem”. In: *Journal of Monetary Economics* 95, pp. 86–96. ISSN: 0304-3932. doi: [10.1016/j.jmoneco.2017.12.004](https://doi.org/10.1016/j.jmoneco.2017.12.004).
- Garay, J., A. Kiayias, and N. Leonardos (2015). “The Bitcoin Backbone Protocol: Analysis and Applications”. In: *Advances in Cryptology – EUROCRYPT 2015*. Ed. by E. Oswald and M. Fischlin. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 281–310.
- Gebhardt, Karl et al. (2000). “A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion”. In: *The Astrophysical Journal* 539.1, pp. L13–L16. doi: [10.1086/312840](https://doi.org/10.1086/312840).
- Gelman, Andrew (2003). “A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-fit Testing*”. In: *International Statistical Review* 71.2, pp. 369–382. doi: [10.1111/j.1751-5823.2003.tb00203.x](https://doi.org/10.1111/j.1751-5823.2003.tb00203.x).

- Gelman, Andrew, Jessica Hwang, and Aki Vehtari (2014a). “Understanding predictive information criteria for Bayesian models”. In: *Statistics and Computing* 24.6, pp. 997–1016. doi: [10.1007/s11222-013-9416-2](https://doi.org/10.1007/s11222-013-9416-2).
- (2014b). “Understanding predictive information criteria for Bayesian models”. In: *Statistics and Computing* 24, pp. 997–1016. doi: [10.1007/s11222-013-9416-2](https://doi.org/10.1007/s11222-013-9416-2).
- Gelman, Andrew et al. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Georgakakis, A. et al. (July 2008). “A new method for determining the sensitivity of X-ray imaging observations and the X-ray number counts”. In: *Monthly Notices of the Royal Astronomical Society* 388.3, pp. 1205–1213. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2008.13423.x](https://doi.org/10.1111/j.1365-2966.2008.13423.x).
- Georgakakis, A. et al. (July 2009). “Host galaxy morphologies of X-ray selected AGN: assessing the significance of different black hole fuelling mechanisms to the accretion density of the Universe at $z \sim 1$.” In: *Monthly Notices of the Royal Astronomical Society* 397.2, pp. 623–633. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2009.14951.x](https://doi.org/10.1111/j.1365-2966.2009.14951.x).
- Georgakakis, A. et al. (Dec. 2011). “Observational constraints on the physics behind the evolution of active galactic nuclei since $z \sim 1$ ”. In: *Monthly Notices of the Royal Astronomical Society* 418.4, pp. 2590–2603. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2011.19650.x](https://doi.org/10.1111/j.1365-2966.2011.19650.x).
- Georgakakis, A. et al. (Mar. 2014). “Investigating evidence for different black hole accretion modes since redshift $z \sim 1$ ”. In: *Monthly Notices of the Royal Astronomical Society* 440.1, pp. 339–352. ISSN: 0035-8711. doi: [10.1093/mnras/stu236](https://doi.org/10.1093/mnras/stu236).
- Georgakakis, Antonis and K. Nandra (June 2011). “A serendipitous XMM survey of the SDSS: the evolution of the colour–magnitude diagram of X-ray AGN from $z = 0.8$ to 0.1 ”. In: *Monthly Notices of the Royal Astronomical Society* 414.2, pp. 992–1010. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2011.18387.x](https://doi.org/10.1111/j.1365-2966.2011.18387.x).
- Georgoula, I. et al. (2015). *Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices*. Available at SSRN. doi: [10.2139/ssrn.2607167](https://doi.org/10.2139/ssrn.2607167).
- Geweke, John and Gianni Amisano (2010). “Comparing and evaluating Bayesian predictive distributions of asset returns”. In: *International Journal of Forecasting* 26.2. Special Issue: Bayesian Forecasting in Economics, pp. 216–230. ISSN: 0169-2070. doi: [10.1016/j.ijforecast.2009.10.007](https://doi.org/10.1016/j.ijforecast.2009.10.007).
- Geweke, John and Charles Whiteman (2006). “Chapter 1 Bayesian Forecasting”. In: *Handbook of Economic Forecasting*. Ed. by G. Elliott, C.W.J. Granger, and A. Timmermann. Vol. 1. Elsevier, pp. 3–80. doi: [10.1016/S1574-0706\(05\)01001-3](https://doi.org/10.1016/S1574-0706(05)01001-3).

- Giallongo, E. et al. (2015). “Faint AGNs at $z \gtrsim 4$ in the CANDELS GOODS-S field: looking for contributors to the reionization of the Universe”. In: *Astronomy & Astrophysics* 578, A83. doi: [10.1051/0004-6361/201425334](https://doi.org/10.1051/0004-6361/201425334).
- Gibson, Robert R. et al. (2009). “A Catalog of Broad Absorption Line Quasars in Sloan Digital Sky Survey Data Release 5”. In: *The Astrophysical Journal* 692.1, pp. 758–777. doi: [10.1088/0004-637x/692/1/758](https://doi.org/10.1088/0004-637x/692/1/758).
- Gilli, R., Comastri, A., and Hasinger, G. (2007). “The synthesis of the cosmic X-ray background in the Chandra and XMM-Newton era”. In: *Astronomy & Astrophysics* 463.1, pp. 79–96. doi: [10.1051/0004-6361:20066334](https://doi.org/10.1051/0004-6361:20066334).
- Gkillas, K. and P. Katsiampa (2018). “An application of extreme value theory to cryptocurrencies”. In: *Economics Letters* 164, pp. 109–111. doi: [10.1016/j.econlet.2018.01.020](https://doi.org/10.1016/j.econlet.2018.01.020).
- Glikman, Eilat et al. (2011). “The Faint End of the Quasar Luminosity Function at $z \sim 4$: Implications for Ionization of the Intergalactic Medium and Cosmic Downsizing”. In: *The Astrophysical Journal* 728.2, p. L26. doi: [10.1088/2041-8205/728/2/L26](https://doi.org/10.1088/2041-8205/728/2/L26).
- Gültekin, Kayhan et al. (2009). “The $M-\sigma$ and $M-L$ Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter”. In: *The Astrophysical Journal* 698.1, pp. 198–221. doi: [10.1088/0004-637x/698/1/198](https://doi.org/10.1088/0004-637x/698/1/198).
- Gneiting, Tilmann (2011). “Making and Evaluating Point Forecasts”. In: *Journal of the American Statistical Association* 106.494, pp. 746–762. doi: [10.1198/jasa.2011.r10138](https://doi.org/10.1198/jasa.2011.r10138).
- Gneiting, Tilmann and Adrian E Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477, pp. 359–378. doi: [10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437).
- Gospodinov, Nikolay, Athanasia Gavala, and Deming Jiang (2006). “Forecasting volatility”. In: *Journal of Forecasting* 25.6, pp. 381–400. doi: [10.1002/for.993](https://doi.org/10.1002/for.993).
- Graham, Alister W. (2012). “Breaking the Law: The $M_{\text{bh}}-M_{\text{spheroid}}$ Relations for Core-S LaTeX command Sample Description Sérsic and Sérsic Galaxies”. In: *The Astrophysical Journal* 746.1, p. 113. doi: [10.1088/0004-637x/746/1/113](https://doi.org/10.1088/0004-637x/746/1/113).
- Graham, Alister W. et al. (2001). “A Correlation between Galaxy Light Concentration and Supermassive Black Hole Mass”. In: *The Astrophysical Journal* 563.1, pp. L11–L14. doi: [10.1086/338500](https://doi.org/10.1086/338500).
- Graham, Alister W. et al. (Apr. 2011). “An expanded $M_{\text{bh}}-\sigma$ diagram, and a new calibration of active galactic nuclei masses”. In: *Monthly Notices of the Royal*

- Astronomical Society* 412.4, pp. 2211–2228. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2010.18045.x](https://doi.org/10.1111/j.1365-2966.2010.18045.x).
- Green, Peter J. (1995). “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination”. In: *Biometrika* 82.4, pp. 711–732. doi: [10.1093/biomet/82.4.711](https://doi.org/10.1093/biomet/82.4.711).
- Grimit, E. P. et al. (2006). “The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification”. In: *Quarterly Journal of the Royal Meteorological Society* 132.621C, pp. 2925–2942. doi: [10.1256/qj.05.235](https://doi.org/10.1256/qj.05.235).
- Grissom, R. L., Ballantyne, D. R., and Wise, J. H. (2014). “On the contribution of active galactic nuclei to reionization”. In: *Astrophysics & Astronomy* 561, A90. doi: [10.1051/0004-6361/201322637](https://doi.org/10.1051/0004-6361/201322637).
- Groen, Jan J. J., Richard Paap, and Francesco Ravazzolo (2013). “Real-Time Inflation Forecasting in a Changing World”. In: *Journal of Business & Economic Statistics* 31.1, pp. 29–44. doi: [10.1080/07350015.2012.727718](https://doi.org/10.1080/07350015.2012.727718).
- Grogin, Norman A. et al. (2011). “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey”. In: *The Astrophysical Journal Supplement Series* 197.2, p. 35. doi: [10.1088/0067-0049/197/2/35](https://doi.org/10.1088/0067-0049/197/2/35).
- Gschlößl, Susanne and Claudia Czado (2007). “Spatial modelling of claim frequency and claim size in non-life insurance”. In: *Scandinavian Actuarial Journal* 2007.3, pp. 202–225. doi: [10.1080/03461230701414764](https://doi.org/10.1080/03461230701414764).
- Guidolin, M. (2011). “Markov Switching Models in Empirical Finance”. In: *Missing Data Methods: Time-Series Methods and Applications*. Ed. by David M. Drukker. Vol. 27 Part 2. Advances in Econometrics. Emerald Group Publishing Limited, pp. 1–86. doi: [10.1108/S0731-9053\(2011\)000027B004](https://doi.org/10.1108/S0731-9053(2011)000027B004).
- Guidolin, Massimo and Allan Timmermann (2006). “An econometric model of non-linear dynamics in the joint distribution of stock and bond returns”. In: *Journal of Applied Econometrics* 21.1, pp. 1–22. doi: [10.1002/jae.824](https://doi.org/10.1002/jae.824).
- (2008). “International asset allocation under regime switching, skew, and kurtosis preferences”. In: *The Review of Financial Studies* 21.2, pp. 889–935. doi: [10.1093/rfs/hhn006](https://doi.org/10.1093/rfs/hhn006).
- Gunn, James E. et al. (2006). “The 2.5 m Telescope of the Sloan Digital Sky Survey”. In: *The Astronomical Journal* 131.4, pp. 2332–2359. doi: [10.1086/500975](https://doi.org/10.1086/500975).
- Guo, Yicheng et al. (2013). “CANDELS Multi-wavelength Catalogs: Source Detection and Photometry in the GOODS-South Field”. In: *The Astrophysical Journal Supplement Series* 207.2, p. 24. doi: [10.1088/0067-0049/207/2/24](https://doi.org/10.1088/0067-0049/207/2/24).

- Haardt, Francesco and Piero Madau (2012). “Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background”. In: *The Astrophysical Journal* 746.2, p. 125. doi: [10.1088/0004-637x/746/2/125](https://doi.org/10.1088/0004-637x/746/2/125).
- Hamilton, James D. (1989). “A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle”. In: *Econometrica* 57.2, pp. 357–384.
- Hasinger, G. (2008). “Absorption properties and evolution of active galactic nuclei”. In: *Astronomy & Astrophysics* 490.3, pp. 905–922. doi: [10.1051/0004-6361:200809839](https://doi.org/10.1051/0004-6361:200809839).
- Hasinger, G., Miyaji, T., and Schmidt, M. (2005). “Luminosity-dependent evolution of soft X-ray selected AGN - New Chandra and XMM-Newton surveys”. In: *Astronomy & Astrophysics* 441.2, pp. 417–434. doi: [10.1051/0004-6361:20042134](https://doi.org/10.1051/0004-6361:20042134).
- Hastie, David I. and Peter J. Green (2011). “Model choice using reversible jump Markov chain Monte Carlo”. In: *Statistica Neerlandica* 66.3, pp. 309–338. doi: [10.1111/j.1467-9574.2012.00516.x](https://doi.org/10.1111/j.1467-9574.2012.00516.x).
- Hayes, A.S. (2017). “Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin”. In: *Telematics and Informatics* 34.7, pp. 1308–1321. doi: [10.1016/j.tele.2016.05.005](https://doi.org/10.1016/j.tele.2016.05.005).
- Häring, Nadine and Hans-Walter Rix (2004). “On the Black Hole Mass-Bulge Mass Relation”. In: *The Astrophysical Journal* 604.2, pp. L89–L92. doi: [10.1086/383567](https://doi.org/10.1086/383567).
- Hersbach, Hans (2000). “Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems”. In: *Weather and Forecasting* 15.5, pp. 559–570. doi: [10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Hewett, Paul C. and Craig B. Foltz (2003). “The Frequency and Radio Properties of Broad Absorption Line Quasars”. In: *The Astronomical Journal* 125.4, pp. 1784–1794. doi: [10.1086/368392](https://doi.org/10.1086/368392).
- Hiroi, Kazuo et al. (2012). “Comoving Space Density and Obscured Fraction of High-redshift Active Galactic Nuclei in the Subaru/XMM-Newton Deep Survey”. In: *The Astrophysical Journal* 758.1, p. 49. doi: [10.1088/0004-637x/758/1/49](https://doi.org/10.1088/0004-637x/758/1/49).
- Holmes, Chris C. and Leonhard Held (2006). “Bayesian auxiliary variable models for binary and multinomial regression”. In: *Bayesian Analysis* 1.1, pp. 145–168. doi: [10.1214/06-BA105](https://doi.org/10.1214/06-BA105).

- Holsclaw, Tracy et al. (2017). “Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling”. In: *The Annals of Applied Statistics* 11.1, pp. 393–426. doi: [10.1214/16-AOAS1009](https://doi.org/10.1214/16-AOAS1009).
- Hopkins, Philip F., Gordon T. Richards, and Lars Hernquist (2007). “An Observational Determination of the Bolometric Quasar Luminosity Function”. In: *The Astrophysical Journal* 654.2, pp. 731–753. doi: [10.1086/509629](https://doi.org/10.1086/509629).
- Hotz-Behofsits, C., F. Huber, and T. O. Zörner (2018). “Predicting crypto-currencies using sparse non-Gaussian state space models”. In: *Journal of Forecasting* 37.6, pp. 627–640. doi: [10.1002/for.2524](https://doi.org/10.1002/for.2524).
- Hsieh, Bau-Ching et al. (2012). “The Taiwan ECDFS Near-Infrared Survey: Ultra-deep J and KS Imaging in the Extended Chandra Deep Field-South”. In: *The Astrophysical Journal Supplement Series* 203.2, p. 23. doi: [10.1088/0067-0049/203/2/23](https://doi.org/10.1088/0067-0049/203/2/23).
- Hsu, Li-Ting et al. (2014). “CANDELS/GOODS-S, CDFS, and ECDFS: Photometric Redshifts for Normal and X-Ray-Detected Galaxies”. In: *The Astrophysical Journal* 796.1, p. 60. doi: [10.1088/0004-637x/796/1/60](https://doi.org/10.1088/0004-637x/796/1/60).
- Hu, Albert S., Christine A. Parlour, and Uday Rajan (2019). “Cryptocurrencies: Stylized facts on a new investible instrument”. In: *Financial Management* 48.4, pp. 1049–1068. doi: [10.1111/fima.12300](https://doi.org/10.1111/fima.12300).
- Ikeda, H. et al. (Jan. 2011). “Probing the Faint End of the Quasar Luminosity Function at $z \sim 4$ in the COSMOS Field”. In: *The Astrophysical Journal* 728.2, p. L25. doi: [10.1088/2041-8205/728/2/L25](https://doi.org/10.1088/2041-8205/728/2/L25).
- Jahnke, Knud and Andrea V. Macciò (May 2011). “The Non-causal Origin of the Black-hole-galaxy Scaling Relations”. In: *The Astrophysical Journal* 734.2, p. 92. doi: [10.1088/0004-637x/734/2/92](https://doi.org/10.1088/0004-637x/734/2/92).
- Jasra, A., C. C. Holmes, and D. A. Stephens (2005). “Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling”. In: *Statist. Sci.* 20.1, pp. 50–67. doi: [10.1214/088342305000000016](https://doi.org/10.1214/088342305000000016).
- Jeffreys, Harold (1961). *Theory of Probability*, 3rd edn. Oxford University Press.
- Jiang, Yonghong, He Nie, and Weihua Ruan (2018). “Time-varying long-term memory in Bitcoin market”. In: *Finance Research Letters* 25, pp. 280–284. ISSN: 1544-6123. doi: [10.1016/j.frl.2017.12.009](https://doi.org/10.1016/j.frl.2017.12.009).
- Jondeau, E. and M. Rockinger (2003). “Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements”. In: *Journal of Economic Dynamics and Control* 27.10, pp. 1699–1737. doi: [10.1016/S0165-1889\(02\)00079-9](https://doi.org/10.1016/S0165-1889(02)00079-9).

- Jong, Roelof S. de et al. (2014). “4MOST: 4-metre Multi-Object Spectroscopic Telescope”. In: *Ground-based and Airborne Instrumentation for Astronomy V*. Ed. by Suzanne K. Ramsay, Ian S. McLean, and Hideki Takami. Vol. 9147. International Society for Optics and Photonics. SPIE, pp. 188–201. doi: [10.1117/12.2055826](https://doi.org/10.1117/12.2055826).
- Just, D. W. et al. (Aug. 2007). “The X-Ray Properties of the Most Luminous Quasars from the Sloan Digital Sky Survey”. In: *The Astrophysical Journal* 665.2, pp. 1004–1022. doi: [10.1086/519990](https://doi.org/10.1086/519990).
- Kalfountzou, E. et al. (Oct. 2014). “The largest X-ray-selected sample of $z \lesssim 3$ AGNs: C-COSMOS and ChaMP”. In: *Monthly Notices of the Royal Astronomical Society* 445.2, pp. 1430–1448. ISSN: 0035-8711. doi: [10.1093/mnras/stu1745](https://doi.org/10.1093/mnras/stu1745).
- Kang, S. H., R. P. McIver, and J. A. Hernandez (2019). “Co-movements between Bitcoin and Gold: A wavelet coherence analysis”. In: *Physica A: Statistical Mechanics and its Applications*, p. 120888. doi: [10.1016/j.physa.2019.04.124](https://doi.org/10.1016/j.physa.2019.04.124).
- Kapar, B. and J. Olmo (2019). “An analysis of price discovery between Bitcoin futures and spot markets”. In: *Economics Letters* 174, pp. 62–64. ISSN: 0165-1765. doi: [10.1016/j.econlet.2018.10.031](https://doi.org/10.1016/j.econlet.2018.10.031).
- Kass, Robert E. et al. (1998). “Markov Chain Monte Carlo in Practice: A Roundtable Discussion”. In: *The American Statistician* 52.2, pp. 93–100. doi: [10.1080/00031305.1998.10480547](https://doi.org/10.1080/00031305.1998.10480547).
- Katsiampa, P. (2017a). “Volatility estimation for Bitcoin: A comparison of GARCH models”. In: *Economics Letters* 158, pp. 3–6. doi: [10.1016/j.econlet.2017.06.023](https://doi.org/10.1016/j.econlet.2017.06.023).
- Katsiampa, Paraskevi (2017b). “Volatility estimation for Bitcoin: A comparison of GARCH models”. In: *Economics Letters* 158, pp. 3–6. ISSN: 0165-1765. doi: [10.1016/j.econlet.2017.06.023](https://doi.org/10.1016/j.econlet.2017.06.023).
- (2019a). “An empirical investigation of volatility dynamics in the cryptocurrency market”. In: *Research in International Business and Finance* 50, pp. 322–335. ISSN: 0275-5319. doi: [10.1016/j.ribaf.2019.06.004](https://doi.org/10.1016/j.ribaf.2019.06.004).
- (2019b). “Volatility co-movement between Bitcoin and Ether”. In: *Finance Research Letters* 30, pp. 221–227. ISSN: 1544-6123. doi: [10.1016/j.frl.2018.10.005](https://doi.org/10.1016/j.frl.2018.10.005).
- Khuntia, Sashikanta and J.K. Pattanayak (2020). “Adaptive long memory in volatility of intra-day bitcoin returns and the impact of trading volume”. In: *Finance Research Letters* 32, p. 101077. ISSN: 1544-6123. doi: [10.1016/j.frl.2018.12.025](https://doi.org/10.1016/j.frl.2018.12.025).
- King, Andrew (2003). “Black Holes, Galaxy Formation, and the $M_{\text{BH}}-\sigma$ Relation”. In: *The Astrophysical Journal* 596.1, pp. L27–L29. doi: [10.1086/379143](https://doi.org/10.1086/379143).
- (2005). “The AGN-Starburst Connection, Galactic Superwinds, and $M_{\text{BH}}-\sigma$ ”. In: *The Astrophysical Journal* 635.2, pp. L121–L123. doi: [10.1086/499430](https://doi.org/10.1086/499430).

- Klein, T., H.P. Thu, and T. Walther (2018). “Bitcoin is not the New Gold – A comparison of volatility, correlation, and portfolio performance”. In: *International Review of Financial Analysis* 59, pp. 105–116. doi: [10.1016/j.irfa.2018.07.010](https://doi.org/10.1016/j.irfa.2018.07.010).
- Koekemoer, Anton M. et al. (Jan. 2004). “A Possible New Population of Sources with Extreme X-Ray/Optical Ratios”. In: *The Astrophysical Journal* 600.2, pp. L123–L126. doi: [10.1086/378181](https://doi.org/10.1086/378181).
- Koekemoer, Anton M. et al. (Dec. 2011). “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics”. In: *The Astrophysical Journal Supplement Series* 197.2, p. 36. doi: [10.1088/0067-0049/197/2/36](https://doi.org/10.1088/0067-0049/197/2/36).
- Koki, C., S. Leonardos, and G. Piliouras (2019). “Do Cryptocurrency Prices Camouflage Latent Economic Effects? A Bayesian Hidden Markov Approach”. In: *MDPI Proceedings of The 3rd Annual Decentralized Conference on Blockchain and Cryptocurrency* 28.1. ISSN: 2504-3900. doi: [10.3390/proceedings2019028005](https://doi.org/10.3390/proceedings2019028005).
- Koki, Constandina, Loukia Meligkotsidou, and Ioannis Vrontos (2020). “Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Pölya-Gamma data augmentation”. In: *Journal of Forecasting* 39.4, pp. 580–598. doi: [10.1002/for.2645](https://doi.org/10.1002/for.2645).
- Konno, Hiroshi, Hiroshi Shirakawa, and Hiroaki Yamazaki (1993). “A mean-absolute deviation-skewness portfolio optimization model”. In: *Annals of Operations Research* 45.1, pp. 205–220. ISSN: 1572-9338. doi: [10.1007/BF02282050](https://doi.org/10.1007/BF02282050).
- Kormendy, John and Luis C. Ho (2013). “Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies”. In: *Annual Review of Astronomy and Astrophysics* 51.1, pp. 511–653. doi: [10.1146/annurev-astro-082708-101811](https://doi.org/10.1146/annurev-astro-082708-101811).
- Koutmos, D. (2018). “Liquidity uncertainty and Bitcoin’s market microstructure”. In: *Economics Letters* 172, pp. 97–101. ISSN: 0165-1765. doi: [10.1016/j.econlet.2018.08.041](https://doi.org/10.1016/j.econlet.2018.08.041).
- Koutmos, Dimitrios (2019). “Market risk and Bitcoin returns”. In: *Annals of Operations Research*. doi: [10.1007/s10479-019-03255-6](https://doi.org/10.1007/s10479-019-03255-6).
- Koutmos, Dimitrios and James E. Payne (2020). “Intertemporal asset pricing with bitcoin”. In: *Review of Quantitative Finance and Accounting*. doi: [10.1007/s11156-020-00904-x](https://doi.org/10.1007/s11156-020-00904-x).
- Kurka, Josef (2019). “Do cryptocurrencies and traditional asset classes influence each other?” In: *Finance Research Letters* 31, pp. 38–46. doi: doi.org/10.1016/j.frl.2019.04.018.

- Laird, E. S. et al. (Dec. 2009). “AEGIS-X: the Chandra Deep Survey of the Extended Groth Strip”. In: *The Astrophysical Journal Supplement Series* 180.1, pp. 102–116. doi: [10.1088/0067-0049/180/1/102](https://doi.org/10.1088/0067-0049/180/1/102).
- Lee, H.-Y. and S.-L. Chen (2006a). “Why use Markov-switching models in exchange rate prediction?” In: *Economic Modelling* 23.4, pp. 662–668. issn: 0264-9993. doi: [10.1016/j.econmod.2006.03.007](https://doi.org/10.1016/j.econmod.2006.03.007).
- Lee, Hsiu-Yun and Show-Lin Chen (2006b). “Why use Markov-switching models in exchange rate prediction?” In: *Economic Modelling* 23.4, pp. 662 –668. issn: 0264-9993. doi: [10.1016/j.econmod.2006.03.007](https://doi.org/10.1016/j.econmod.2006.03.007).
- Lehmer, B. D. et al. (Nov. 2005). “The Extended Chandra Deep Field–South Survey: Chandra Point-Source Catalogs”. In: *The Astrophysical Journal Supplement Series* 161.1, pp. 21–40. doi: [10.1086/444590](https://doi.org/10.1086/444590).
- Lilly, Simon J. et al. (Sept. 2009). “The zCOSMOS 10k-Bright Spectroscopic Sample”. In: *The Astrophysical Journal Supplement Series* 184.2, pp. 218–229. doi: [10.1088/0067-0049/184/2/218](https://doi.org/10.1088/0067-0049/184/2/218).
- Liu, Jia and John M. Maheu (2018). “Improving Markov switching models using realized variance”. In: *Journal of Applied Econometrics* 33.3, pp. 297–318. doi: [10.1002/jae.2605](https://doi.org/10.1002/jae.2605).
- Liu, Zhu et al. (Apr. 2016). “X-ray spectral properties of the AGN sample in the northern XMM-XXL field”. In: *Monthly Notices of the Royal Astronomical Society* 459.2, pp. 1602–1625. issn: 0035-8711. doi: [10.1093/mnras/stw753](https://doi.org/10.1093/mnras/stw753).
- Luo, B. et al. (Sept. 2014). “Weak Hard X-Ray Emission from Broad Absorption Line Quasars: Evidence for Intrinsic X-Ray Weakness”. In: *The Astrophysical Journal* 794.1, p. 70. doi: [10.1088/0004-637x/794/1/70](https://doi.org/10.1088/0004-637x/794/1/70).
- Lusso, E. et al. (Apr. 2015). “The first ultraviolet quasar-stacked spectrum at $z \approx 2.4$ from WFC3”. In: *Monthly Notices of the Royal Astronomical Society* 449.4, pp. 4204–4220. issn: 0035-8711. doi: [10.1093/mnras/stv516](https://doi.org/10.1093/mnras/stv516).
- Lusso, E. et al. (2010). “The X-ray to optical-UV luminosity ratio of X-ray selected type 1 AGN in XMM-COSMOS*”. In: *Astronomy & Astrophysics* 512, A34. doi: [10.1051/0004-6361/200913298](https://doi.org/10.1051/0004-6361/200913298).
- Machete, Reason L. (2013). “Contrasting probabilistic scoring rules”. In: *Journal of Statistical Planning and Inference* 143.10, pp. 1781 –1790. doi: [10.1016/j.jspi.2013.05.012](https://doi.org/10.1016/j.jspi.2013.05.012).
- Madau, Piero, Francesco Haardt, and Martin J. Rees (Apr. 1999). “Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources”. In: *The Astrophysical Journal* 514.2, pp. 648–659. doi: [10.1086/306975](https://doi.org/10.1086/306975).

- Magorrian, John et al. (Jan. 1998). "The Demography of Massive Dark Objects in Galaxy Centers". In: *The Astronomical Journal* 115.6, pp. 2285–2305. doi: [10.1086/300353](https://doi.org/10.1086/300353).
- Mamon, R.S. and R.J. Elliott, eds. (2014). *Hidden Markov Models in Finance: Further Developments and Applications, Volume II*. International Series in Operations Research & Management Science. Springer.
- Marconi, A. et al. (June 2004). "Local supermassive black holes, relics of active galactic nuclei and the X-ray background". In: *Monthly Notices of the Royal Astronomical Society* 351.1, pp. 169–185. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2004.07765.x](https://doi.org/10.1111/j.1365-2966.2004.07765.x).
- Marconi, Alessandro and Leslie K. Hunt (Apr. 2003). "The Relation between Black Hole Mass, Bulge Mass, and Near-Infrared Luminosity". In: *The Astrophysical Journal* 589.1, pp. L21–L24. doi: [10.1086/375804](https://doi.org/10.1086/375804).
- Marin, Jean-Michel, Kerrie Mengersen, and Christian P. Robert (2005). "Bayesian Modelling and Inference on Mixtures of Distributions". In: *Bayesian Thinking*. Ed. by D.K. Dey and C.R. Rao. Vol. 25. Handbook of Statistics. Elsevier, pp. 459–507. doi: [10.1016/S0169-7161\(05\)25016-2](https://doi.org/10.1016/S0169-7161(05)25016-2).
- Marshall, H. L. et al. (June 1983). "Analysis of complete quasar samples to obtain parameters of luminosity and evolution functions". In: *The Astrophysical Journal* 269, pp. 35–41. doi: [10.1086/161016](https://doi.org/10.1086/161016).
- Masters, D. et al. (Aug. 2012). "Evolution of the Quasar Luminosity Function over $3 < z < 5$ in the COSMOS Survey Field". In: *The Astrophysical Journal* 755.2, p. 169. doi: [10.1088/0004-637x/755/2/169](https://doi.org/10.1088/0004-637x/755/2/169).
- McAleer, Michael and Marcelo C. Medeiros (2008). "Realized Volatility: A Review". In: *Econometric Reviews* 27.1-3, pp. 10–45. doi: [10.1080/07474930701853509](https://doi.org/10.1080/07474930701853509).
- McGreer, Ian D. et al. (Apr. 2013). "The $z = 5$ Quasar Luminosity Function from SDSS Stripe 82". In: *The Astrophysical Journal* 768.2, p. 105. doi: [10.1088/0004-637x/768/2/105](https://doi.org/10.1088/0004-637x/768/2/105).
- McLure, R. J. and J. S. Dunlop (Apr. 2002). "On the black hole-bulge mass relation in active and inactive galaxies". In: *Monthly Notices of the Royal Astronomical Society* 331.3, pp. 795–804. ISSN: 0035-8711. doi: [10.1046/j.1365-8711.2002.05236.x](https://doi.org/10.1046/j.1365-8711.2002.05236.x).
- McMillan, David G. (2020). "Forecasting U.S. stock returns". In: *The European Journal of Finance* 0.0, pp. 1–24. doi: [10.1080/1351847X.2020.1719175](https://doi.org/10.1080/1351847X.2020.1719175).
- McNally, S., J. Roche, and S. Caton (2018). "Predicting the Price of Bitcoin Using Machine Learning". In: *26th Euromicro International Conference on Parallel,*

- Distributed and Network-based Processing (PDP)*, pp. 339–343. doi: [10.1109/PDP2018.2018.00060](https://doi.org/10.1109/PDP2018.2018.00060).
- Meligkotsidou, Loukia and Petros Dellaportas (2011). “Forecasting with Non-homogeneous Hidden Markov Models”. In: *Statistics and Computing* 21.3, pp. 439–449. doi: [10.1007/s11222-010-9180-5](https://doi.org/10.1007/s11222-010-9180-5).
- Meligkotsidou, Loukia et al. (2019). “Quantile forecast combinations in realised volatility prediction”. In: *Journal of the Operational Research Society* 0.0, pp. 1–14. doi: [10.1080/01605682.2018.1489354](https://doi.org/10.1080/01605682.2018.1489354).
- Mensi, Walid, Khamis Hamed Al-Yahyaee, and Sang Hoon Kang (2019). “Structural breaks and double long memory of cryptocurrency prices: A comparative analysis from Bitcoin and Ethereum”. In: *Finance Research Letters* 29, pp. 222–230. ISSN: 1544-6123. doi: [10.1016/j.frl.2018.07.011](https://doi.org/10.1016/j.frl.2018.07.011).
- Menzel, M.-L. et al. (Jan. 2016). “A spectroscopic survey of X-ray-selected AGNs in the northern XMM-XXL field”. In: *Monthly Notices of the Royal Astronomical Society* 457.1, pp. 110–132. ISSN: 0035-8711. doi: [10.1093/mnras/stv2749](https://doi.org/10.1093/mnras/stv2749).
- Merloni, A. et al. (Sept. 2012). “eROSITA Science Book: Mapping the Structure of the Energetic Universe”. In: *arXiv e-prints*, arXiv:1209.3114, arXiv:1209.3114. arXiv: [1209.3114](https://arxiv.org/abs/1209.3114) [astro-ph.HE].
- Merloni, A. et al. (Nov. 2014). “The incidence of obscuration in active galactic nuclei”. In: *Monthly Notices of the Royal Astronomical Society* 437.4, pp. 3550–3567. ISSN: 0035-8711. doi: [10.1093/mnras/stt2149](https://doi.org/10.1093/mnras/stt2149).
- Mittnik, Stefan, Nikolay Robinzonov, and Martin Spindler (2015). “Stock market volatility: Identifying major drivers and the nature of their impact”. In: *Journal of Banking & Finance* 58, pp. 1–14. doi: [10.1016/j.jbankfin.2015.04.003](https://doi.org/10.1016/j.jbankfin.2015.04.003).
- Miyaji, T. et al. (May 2015). “Detailed Shape and Evolutionary Behavior of the X-Ray Luminosity Function of Active Galactic Nuclei”. In: *The Astrophysical Journal* 804.2, p. 104. doi: [10.1088/0004-637x/804/2/104](https://doi.org/10.1088/0004-637x/804/2/104).
- Miyaji, T., Hasinger, G., and Schmidt, M. (2001). “Soft X-ray AGN luminosity function from ROSAT surveys - II. Table of the binned soft X-ray luminosity function”. In: *Astronomy & Astrophysics* 369.1, pp. 49–56. doi: [10.1051/0004-6361:20010102](https://doi.org/10.1051/0004-6361:20010102).
- Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available at Bitcoin.org.
- Nandra, K. and K. A. Pounds (May 1994). “Ginga observations of the X-ray spectra of Seyfert galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 268.2, pp. 405–429. ISSN: 0035-8711. doi: [10.1093/mnras/268.2.405](https://doi.org/10.1093/mnras/268.2.405).

- Nandra, K. et al. (Oct. 2007). “An XMM–Newton survey of broad iron lines in Seyfert galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 382.1, pp. 194–228. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2007.12331.x](https://doi.org/10.1111/j.1365-2966.2007.12331.x).
- Nandra, K. et al. (Sept. 2015). “AEGIS-X: Deep Chandra Imaging of the Central Groth Strip”. In: *The Astrophysical Journal Supplement Series* 220.1, p. 10. doi: [10.1088/0067-0049/220/1/10](https://doi.org/10.1088/0067-0049/220/1/10).
- Narisetty, Naveen N., Juan Shen, and Xuming He (2018). “Skinny Gibbs: A Consistent and Scalable Gibbs Sampler for Model Selection”. In: *Journal of the American Statistical Association* 0.0, pp. 1–13. doi: [10.1080/01621459.2018.1482754](https://doi.org/10.1080/01621459.2018.1482754).
- Narisetty, Naveen Naidu and Xuming He (2014). “Bayesian variable selection with shrinking and diffusing priors”. In: *Ann. Statist.* 42.2, pp. 789–817. doi: [10.1214/14-AOS1207](https://doi.org/10.1214/14-AOS1207).
- Neal, Radford M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Tech. rep. t RG-TR-93-1, University of Toronto, Department of Computer Science. URL: <https://www.cs.toronto.edu/~radford/ftp/review.pdf>.
- Nelson, Daniel B. (1991). “Conditional Heteroskedasticity in Asset Returns: A New Approach”. In: *Econometrica* 59.2, pp. 347–370. doi: [10.2307/2938260](https://doi.org/10.2307/2938260).
- Nikolsko-Rzhevskyy, Alex and Ruxandra Prodan (2012). “Markov switching and exchange rate predictability”. In: *International Journal of Forecasting* 28.2, pp. 353–365. ISSN: 0169-2070. doi: [10.1016/j.ijforecast.2011.04.007](https://doi.org/10.1016/j.ijforecast.2011.04.007).
- O’Brien, Sean M. and David B. Dunson (2004). “Bayesian Multivariate Logistic Regression”. In: *Biometrics* 60.3, pp. 739–746. doi: [10.1111/j.0006-341X.2004.00224.x](https://doi.org/10.1111/j.0006-341X.2004.00224.x).
- O’Hara, R.B. and M. J. Sillanpää (2009). “A review of Bayesian variable selection methods: what, how and which”. In: *Bayesian Analysis* 4.1, pp. 85–117. doi: [10.1214/09-BA403](https://doi.org/10.1214/09-BA403).
- Page, M. J. and F. J. Carrera (Jan. 2000). “An improved method of constructing binned luminosity functions”. In: *Monthly Notices of the Royal Astronomical Society* 311.2, pp. 433–440. ISSN: 0035-8711. doi: [10.1046/j.1365-8711.2000.03105.x](https://doi.org/10.1046/j.1365-8711.2000.03105.x).
- Palanque-Delabrouille, N. et al. (2011). “Variability selected high-redshift quasars on SDSS Stripe 82”. In: *Astronomy & Astrophysics* 530, A122. doi: [10.1051/0004-6361/201016254](https://doi.org/10.1051/0004-6361/201016254).
- Panopoulou, Ekaterini and Theologos Pantelidis (2015). “Regime-switching models for exchange rates”. In: *The European Journal of Finance* 21.12, pp. 1023–1069. doi: [10.1080/1351847X.2014.904240](https://doi.org/10.1080/1351847X.2014.904240).

- Panopoulou, Ekaterini and Ioannis Souropanis (2019). “The role of technical indicators in exchange rate forecasting”. In: *Journal of Empirical Finance* 53, pp. 197–221. ISSN: 0927-5398. DOI: [10.1016/j.jempfin.2019.07.004](https://doi.org/10.1016/j.jempfin.2019.07.004).
- Paper, Libra White (2019). Available at Libra Association.
- Paroli, Roberta and Luigi Spezia (2008). “Bayesian inference in non-homogeneous Markov mixtures of periodic autoregressions with state-dependent exogenous variables”. In: *Computational Statistics & Data Analysis* 52.5, pp. 2311–2330. DOI: [10.1016/j.csda.2007.09.020](https://doi.org/10.1016/j.csda.2007.09.020).
- Paul, Rajib, Steven N. MacEachern, and L. Mark Berliner (2012). “Assessing Convergence and Mixing of MCMC Implementations via Stratification”. In: *Journal of Computational and Graphical Statistics* 21.3, pp. 693–712. DOI: [10.1080/10618600.2012.663293](https://doi.org/10.1080/10618600.2012.663293).
- Pawlik, Andreas H., Joop Schaye, and Eveline Van Scherpenzeel (Apr. 2009). “Keeping the Universe ionized: photoheating and the clumping factor of the high-redshift intergalactic medium”. In: *Monthly Notices of the Royal Astronomical Society* 394.4, pp. 1812–1824. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2009.14486.x](https://doi.org/10.1111/j.1365-2966.2009.14486.x).
- Paye, Bradley S. (2012). “Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables”. In: *Journal of Financial Economics* 106.3, pp. 527–546. DOI: [10.1016/j.jfineco.2012.06.005](https://doi.org/10.1016/j.jfineco.2012.06.005).
- Pereira, Marcelo, Sofia B. Ramos, and José G. Dias (2017). “The cyclical behaviour of commodities”. In: *The European Journal of Finance* 23.12, pp. 1107–1128. DOI: [10.1080/1351847X.2016.1205505](https://doi.org/10.1080/1351847X.2016.1205505).
- Persio], Luca [Di and Matteo Frigo (2016). “Gibbs sampling approach to regime switching analysis of financial time series”. In: *Journal of Computational and Applied Mathematics* 300, pp. 43–55. ISSN: 0377-0427. DOI: [10.1016/j.cam.2015.12.010](https://doi.org/10.1016/j.cam.2015.12.010).
- Phillip, A., J. Chan, and S. Peiris (2018a). “On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin”. In: *Econometrics and Statistics*. DOI: [10.1016/j.ecosta.2018.10.003](https://doi.org/10.1016/j.ecosta.2018.10.003).
- Phillip, A., J. S.K. Chan, and S. Peiris (2018b). “A new look at Cryptocurrencies”. In: *Economics Letters* 163, pp. 6–9. ISSN: 0165-1765. DOI: [10.1016/j.econlet.2017.11.020](https://doi.org/10.1016/j.econlet.2017.11.020).

- Phillips, R. C. and D. Gorse (2017). “Predicting cryptocurrency price bubbles using social media data and epidemic modelling”. In: *IEEE Symposium Series on Computational Intelligence*, pp. 1–7. doi: [10.1109/SSCI.2017.8280809](https://doi.org/10.1109/SSCI.2017.8280809).
- Pichl, L. and T. Kaizoji (2017). “Volatility Analysis of Bitcoin Price Time Series”. In: *Quantitative Finance and Economics* 1.QFE-01-00474, pp. 474–485. doi: [10.3934/QFE.2017.4.474](https://doi.org/10.3934/QFE.2017.4.474).
- Pinto, Cecilia and Luigi Spezia (2016). “Markov switching autoregressive models for interpreting vertical movement data with application to an endangered marine apex predator”. In: *Methods in Ecology and Evolution* 7.4, pp. 407–417. doi: [10.1111/2041-210X.12494](https://doi.org/10.1111/2041-210X.12494).
- Platanakis, Emmanouil, Athanasios Sakkas, and Charles Sutcliffe (2019). “The role of transaction costs and risk aversion when selecting between one and two regimes for portfolio models”. In: *Applied Economics Letters* 26.6, pp. 516–521. doi: [10.1080/13504851.2018.1486984](https://doi.org/10.1080/13504851.2018.1486984).
- Platanakis, Emmanouil and Andrew Urquhart (2020). “Should investors include Bitcoin in their portfolios? A portfolio theory approach”. In: *The British Accounting Review* 52.4, p. 100837. ISSN: 0890-8389. doi: [10.1016/j.bar.2019.100837](https://doi.org/10.1016/j.bar.2019.100837).
- Polpo, Adriano et al., eds. (May 2002). *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Vol. 617. American Institute of Physics Conference Series.
- Polson, Nicholas G., James G. Scott, and Jesse Windle (2013). “Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables”. In: *Journal of the American Statistical Association* 108.504, pp. 1339–1349. doi: [10.1080/01621459.2013.829001](https://doi.org/10.1080/01621459.2013.829001).
- Poyser, O. (Mar. 2019). “Exploring the dynamics of Bitcoin’s price: a Bayesian structural time series approach”. In: *Eurasian Economic Review* 9.1, pp. 29–60. doi: [10.1007/s40822-018-0108-2](https://doi.org/10.1007/s40822-018-0108-2).
- Press, William H. et al. (1992). *Numerical Recipes in FORTRAN (2nd Ed.): The Art of Scientific Computing*. USA: Cambridge University Press. ISBN: 052143064X.
- Rangel, C. et al. (Nov. 2013). “X-ray properties of BzK-selected galaxies in the deepest X-ray fields”. In: *Monthly Notices of the Royal Astronomical Society* 428.4, pp. 3089–3103. ISSN: 0035-8711. doi: [10.1093/mnras/sts256](https://doi.org/10.1093/mnras/sts256).
- Reichard, Timothy A. et al. (2003). “A Catalog of Broad Absorption Line Quasars from the Sloan Digital Sky Survey Early Data Release”. In: *The Astronomical Journal* 125.4, pp. 1711–1728. doi: [10.1086/368244](https://doi.org/10.1086/368244).

- Richards, Gordon T. et al. (June 2006). “The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3”. In: *The Astronomical Journal* 131.6, pp. 2766–2787. doi: [10.1086/503559](https://doi.org/10.1086/503559).
- Richards, Gordon T. et al. (Dec. 2008). “Efficient Photometric Selection of Quasars from the Sloan Digital Sky Survey. II. $\sim 1,000,000$ Quasars from Data Release 6”. In: *The Astrophysical Journal Supplement Series* 180.1, pp. 67–83. doi: [10.1088/0067-0049/180/1/67](https://doi.org/10.1088/0067-0049/180/1/67).
- Rodighiero, G. et al. (2007). “Unveiling the oldest and most massive galaxies at very high redshift*”. In: *Astronomy & Astrophysics* 470.1, pp. 21–37. doi: [10.1051/0004-6361:20066497](https://doi.org/10.1051/0004-6361:20066497).
- Rodríguez, Carlos E. and Stephen G. Walker (2014). “Label Switching in Bayesian Mixture Models: Deterministic Relabeling Strategies”. In: *Journal of Computational and Graphical Statistics* 23.1, pp. 25–45. doi: [10.1080/10618600.2012.735624](https://doi.org/10.1080/10618600.2012.735624).
- Ross, Nicholas P. et al. (Feb. 2012). “The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine”. In: *The Astrophysical Journal Supplement Series* 199.1, p. 3. doi: [10.1088/0067-0049/199/1/3](https://doi.org/10.1088/0067-0049/199/1/3).
- Ross, Nicholas P. et al. (July 2013). “The SDSS-III Baryon Oscillation Spectroscopic Survey: The Quasar Luminosity Function from Data Release Nine”. In: *The Astrophysical Journal* 773.1, p. 14. doi: [10.1088/0004-637x/773/1/14](https://doi.org/10.1088/0004-637x/773/1/14).
- Salvato, M. et al. (Dec. 2009). “Photometric Redshift and Classification for the XMM-COSMOS Sources”. In: *The Astrophysical Journal* 690.2, pp. 1250–1263. doi: [10.1088/0004-637x/690/2/1250](https://doi.org/10.1088/0004-637x/690/2/1250).
- Salvato, M. et al. (2011). “Dissecting Photometric Redshift for Active Galactic Nucleus Using XMM- and Chandra-COSMOS Samples”. In: *The Astrophysical Journal* 742.2, p. 61. doi: [10.1088/0004-637x/742/2/61](https://doi.org/10.1088/0004-637x/742/2/61).
- Savorgnan, G. et al. (July 2013). “The supermassive black hole mass–Sérsic index relations for bulges and elliptical galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 434.1, pp. 387–397. issn: 0035-8711. doi: [10.1093/mnras/stt1027](https://doi.org/10.1093/mnras/stt1027).
- Schaerer, D. et al. (2007). “EROs found behind lensing clusters - I. Stellar populations and dust properties of optical dropout EROs and comparison with related objects”. In: *Astronomy & Astrophysics* 476.1, p. 97. doi: [10.1051/0004-6361:20077127e](https://doi.org/10.1051/0004-6361:20077127e).

- Schmidt, Maarten, Donald P. Schneider, and James E. Gunn (July 1995). “Spectroscopic CCD Surveys for Quasars at Large Redshift.IV.Evolution of the Luminosity Function from Quasars Detected by Their Lyman-Alpha Emission”. In: *Astronomical Journal* 110, p. 68. doi: [10.1086/117497](https://doi.org/10.1086/117497).
- Schwarz, Gideon (1978). “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2, pp. 461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).
- Scott, Steven L. (2002). “Bayesian Methods for Hidden Markov Models”. In: *Journal of the American Statistical Association* 97.457, pp. 337–351. doi: [10.1198/016214502753479464](https://doi.org/10.1198/016214502753479464).
- Shemmer, O. et al. (Sept. 2005). “The X-Ray Spectral Properties and Variability of Luminous High-Redshift Active Galactic Nuclei”. In: *The Astrophysical Journal* 630.2, pp. 729–739. doi: [10.1086/432050](https://doi.org/10.1086/432050).
- Shen, Dehua, Andrew Urquhart, and Pengfei Wang (n.d.). “Forecasting the volatility of Bitcoin: The importance of jumps and structural breaks”. In: *European Financial Management* n/a.n/a (). doi: [10.1111/eufm.12254](https://doi.org/10.1111/eufm.12254).
- Shorish, Jamsheed (2018). “Hedonic pricing of cryptocurrency tokens”. In: *Digital Finance* 1, pp. 163 –189. ISSN: 254-6186. doi: [10.1007/s42521-019-00005-y](https://doi.org/10.1007/s42521-019-00005-y).
- Silk, Joseph and Martin J. Rees (Mar. 1998). “Quasars and galaxy formation”. In: *Astronomy & Astrophysics* 331, pp. L1–L4. arXiv: [astro-ph/9801013](https://arxiv.org/abs/astro-ph/9801013) [astro-ph].
- Simpson, Chris (June 2005). “The luminosity dependence of the type 1 active galactic nucleus fraction”. In: *Monthly Notices of the Royal Astronomical Society* 360.2, pp. 565–572. ISSN: 0035-8711. doi: [10.1111/j.1365-2966.2005.09043.x](https://doi.org/10.1111/j.1365-2966.2005.09043.x).
- Smee, Stephen A. et al. (July 2013). “The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey”. In: *The Astronomical Journal* 146.2, p. 32. doi: [10.1088/0004-6256/146/2/32](https://doi.org/10.1088/0004-6256/146/2/32).
- Soltan, Andrzej (Sept. 1982). “Masses of quasars”. In: *Monthly Notices of the Royal Astronomical Society* 200.1, pp. 115–122. ISSN: 0035-8711. doi: [10.1093/mnras/200.1.115](https://doi.org/10.1093/mnras/200.1.115).
- Spezia, Luigi (2006). “Bayesian analysis of non-homogeneous hidden Markov models”. In: *Journal of Statistical Computation and Simulation* 76.8, pp. 713–725. doi: [10.1080/10629360500108798](https://doi.org/10.1080/10629360500108798).
- (2009). “Reversible jump and the label switching problem in hidden Markov models”. In: *Journal of Statistical Planning and Inference* 139.7, pp. 2305 –2315. ISSN: 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2008.10.016>.

- Spiegelhalter, David J. et al. (2002). “Bayesian measures of model complexity and fit”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.4, pp. 583–639. doi: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353).
- Stalin, C. S. et al. (Dec. 2010). “Optical identification of XMM sources in the Canada – France – Hawaii Telescope Legacy Survey”. In: *Monthly Notices of the Royal Astronomical Society* 401.1, pp. 294–306. issn: 0035-8711. doi: [10.1111/j.1365-2966.2009.15636.x](https://doi.org/10.1111/j.1365-2966.2009.15636.x).
- Stavroyiannis, Stavros and Vassilios Babalos (2019). “Herding behavior in cryptocurrencies revisited: Novel evidence from a TVP model”. In: *Journal of Behavioral and Experimental Finance* 22, pp. 57 –63. issn: 2214-6350. doi: [10.1016/j.jbef.2019.02.007](https://doi.org/10.1016/j.jbef.2019.02.007).
- Steffen, A. T. et al. (June 2006). “The X-Ray-to-Optical Properties of Optically Selected Active Galaxies over Wide Luminosity and Redshift Ranges”. In: *The Astronomical Journal* 131.6, pp. 2826–2842. doi: [10.1086/503627](https://doi.org/10.1086/503627).
- Stern, D. et al. (Sept. 2014). “NuSTAR and XMM-Newton Observations of Luminous, Heavily Obscured, WISE-selected Quasars at $z \sim 2$ ”. In: *The Astrophysical Journal* 794.2, p. 102. doi: [10.1088/0004-637x/794/2/102](https://doi.org/10.1088/0004-637x/794/2/102).
- Strüder, L. et al. (2001). “The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera”. In: *Astronomy & Astrophysics* 365.1, pp. L18–L26. doi: [10.1051/0004-6361:20000066](https://doi.org/10.1051/0004-6361:20000066).
- Sutherland, Will and Will Saunders (Dec. 1992). “On the likelihood ratio for source identification”. In: *Monthly Notices of the Royal Astronomical Society* 259.3, pp. 413–420. issn: 0035-8711. doi: [10.1093/mnras/259.3.413](https://doi.org/10.1093/mnras/259.3.413).
- Székely, Gábor J. and Maria L. Rizzo (2005a). “A new test for multivariate normality”. In: *Journal of Multivariate Analysis* 93.1, pp. 58 –80. doi: [10.1016/j.jmva.2003.12.002](https://doi.org/10.1016/j.jmva.2003.12.002).
- (2005b). “A new test for multivariate normality”. In: *Journal of Multivariate Analysis* 93.1, pp. 58 –80. issn: 0047-259X. doi: [10.1016/j.jmva.2003.12.002](https://doi.org/10.1016/j.jmva.2003.12.002).
- Telfer, Randal C. et al. (Feb. 2002). “The Rest-Frame Extreme-Ultraviolet Spectral Properties of Quasi-stellar Objects”. In: *The Astrophysical Journal* 565.2, pp. 773–785. doi: [10.1086/324689](https://doi.org/10.1086/324689).
- Thies, Sven and Peter Molnár (2018). “Bayesian change point analysis of Bitcoin returns”. In: *Finance Research Letters* 27, pp. 223 –227. issn: 1544-6123. doi: [10.1016/j.frl.2018.03.018](https://doi.org/10.1016/j.frl.2018.03.018).

- Tian, Ruilin and Gang Shen (2019). “Predictive power of Markovian models: Evidence from US recession forecasting”. In: *Journal of Forecasting* 38.6, pp. 525–551. doi: [10.1002/for.2579](https://doi.org/10.1002/for.2579).
- Timmermann, Allan (2000). “Moments of Markov switching models”. In: *Journal of Econometrics* 96.1, pp. 75–111. ISSN: 0304-4076. doi: [10.1016/S0304-4076\(99\)00051-2](https://doi.org/10.1016/S0304-4076(99)00051-2).
- Trump, Jonathan R. et al. (Apr. 2009). “The COSMOS Active Galactic Nucleus Spectroscopic Survey. I. XMM-Newton Counterparts”. In: *The Astrophysical Journal* 696.2, pp. 1195–1212. doi: [10.1088/0004-637x/696/2/1195](https://doi.org/10.1088/0004-637x/696/2/1195).
- Turner, M. J. L. et al. (2001). “The European Photon Imaging Camera on XMM-Newton: The MOS cameras”. In: *Astronomy & Astrophysics* 365.1, pp. L27–L35. doi: [10.1051/0004-6361:20000087](https://doi.org/10.1051/0004-6361:20000087).
- Ueda, Yoshihiro et al. (Dec. 2003). “Cosmological Evolution of the Hard X-Ray Active Galactic Nucleus Luminosity Function and the Origin of the Hard X-Ray Background”. In: *The Astrophysical Journal* 598.2, pp. 886–908. doi: [10.1086/378940](https://doi.org/10.1086/378940).
- Ueda, Yoshihiro et al. (Apr. 2014). “Toward the Standard Population 3. Synthesis Model of the X-Ray Background: Evolution of X-Ray Luminosity and Absorption Functions of Active Galactic Nuclei Including Compton-thick Populations”. In: *The Astrophysical Journal* 786.2, p. 104. doi: [10.1088/0004-637x/786/2/104](https://doi.org/10.1088/0004-637x/786/2/104).
- Urquhart, A. (2018). “What causes the attention of Bitcoin?” In: *Economics Letters* 166, pp. 40–44. doi: [10.1016/j.econlet.2018.02.017](https://doi.org/10.1016/j.econlet.2018.02.017).
- Urquhart, Andrew and Hanxiong Zhang (2019). “Is Bitcoin a hedge or safe haven for currencies? An intraday analysis”. In: *International Review of Financial Analysis* 63, pp. 49–57. ISSN: 1057-5219. doi: [10.1016/j.irfa.2019.02.009](https://doi.org/10.1016/j.irfa.2019.02.009).
- Van Wijk, D. (2013). *What can be expected from the BitCoin*. PhD Thesis. Erasmus Universiteit Rotterdam.
- Vats, Dootika, James M Flegal, and Galin L Jones (Apr. 2019). “Multivariate output analysis for Markov chain Monte Carlo”. In: *Biometrika* 106.2, pp. 321–337. doi: [10.1093/biomet/asz002](https://doi.org/10.1093/biomet/asz002).
- Vignali, C. et al. (2005). “X-Ray Lighthouses of the High-Redshift Universe. II. Further Snapshot Observations of the Most Luminous $z \gtrsim 4$ Quasars with Chandra”. In: *The Astronomical Journal* 129.6, pp. 2519–2530. doi: [10.1086/430217](https://doi.org/10.1086/430217).
- Vito, F. et al. (Oct. 2013). “The high-redshift ($z \gtrsim 3$) active galactic nucleus population in the 4-Ms Chandra Deep Field-South”. In: *Monthly Notices of the Royal Astronomical Society* 428.1, pp. 354–369. ISSN: 0035-8711. doi: [10.1093/mnras/sts034](https://doi.org/10.1093/mnras/sts034).

- Vito, F. et al. (Oct. 2014). “The hard X-ray luminosity function of high-redshift ($3 < z \lesssim 5$) active galactic nuclei”. In: *Monthly Notices of the Royal Astronomical Society* 445.4, pp. 3557–3574. ISSN: 0035-8711. DOI: [10.1093/mnras/stu2004](https://doi.org/10.1093/mnras/stu2004).
- Watanabe, Sumio (2013). “A Widely Applicable Bayesian Information Criterion”. In: *J. Mach. Learn. Res.* 14.1, pp. 867–897. ISSN: 1532-4435. URL: <http://dl.acm.org/citation.cfm?id=2567709.2502609>.
- Weigel, Anna K. et al. (Mar. 2015). “The systematic search for $z \geq 5$ active galactic nuclei in the Chandra Deep Field South”. In: *Monthly Notices of the Royal Astronomical Society* 448.4, pp. 3167–3195. ISSN: 0035-8711. DOI: [10.1093/mnras/stv184](https://doi.org/10.1093/mnras/stv184).
- Welch, Ivo and Amit Goyal (2008). “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction”. In: *The Review of Financial Studies* 21.4, pp. 1455–1508. DOI: [10.1093/rfs/hhm014](https://doi.org/10.1093/rfs/hhm014).
- Wright, Jonathan H. (2009). “Forecasting US inflation by Bayesian model averaging”. In: *Journal of Forecasting* 28.2, pp. 131–144. DOI: [10.1002/for.1088](https://doi.org/10.1002/for.1088).
- Wu, Shu and Yong Zeng (2014). “An Econometric Model of the Term Structure of Interest Rates Under Regime-Switching Risk”. In: *Hidden Markov models in Finance. International Series in Operations Research & Management Science*. Ed. by Rogemar S. Mamon and Robert J. Elliott. Vol. 209. Springer, Boston, MA, pp. 55–83.
- Xue, Y. Q. et al. (June 2011). “The Chandra Deep Field-South Survey: 4 Ms source catalogs”. In: *The Astrophysical Journal Supplement Series* 195.1, p. 10. DOI: [10.1088/0067-0049/195/1/10](https://doi.org/10.1088/0067-0049/195/1/10).
- Yenko, B. et al. (May 2009). “The OPTX Project. II. Hard X-Ray Luminosity Functions of Active Galactic Nuclei for $z \lesssim 5$ ”. In: *The Astrophysical Journal* 698.1, pp. 380–396. DOI: [10.1088/0004-637x/698/1/380](https://doi.org/10.1088/0004-637x/698/1/380).
- Yermack, D. (2015). “Chapter 2 – Is Bitcoin a Real Currency? An Economic Appraisal”. In: *Handbook of Digital Currency*. Ed. by D. L.K. Chuen. San Diego: Academic Press, pp. 31–43. DOI: [10.1016/B978-0-12-802117-0.00002-3](https://doi.org/10.1016/B978-0-12-802117-0.00002-3).
- Yi, S., Z. Xu, and G.-J. Wang (2018). “Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency?” In: *International Review of Financial Analysis* 60, pp. 98–114. DOI: [10.1016/j.irfa.2018.08.012](https://doi.org/10.1016/j.irfa.2018.08.012).
- Yu, Miao (2019). “Forecasting Bitcoin volatility: The role of leverage effect and uncertainty”. In: *Physica A: Statistical Mechanics and its Applications* 533, p. 120707. ISSN: 0378-4371. DOI: [10.1016/j.physa.2019.03.072](https://doi.org/10.1016/j.physa.2019.03.072).

- Yuan, Chunming (2011). “Forecasting exchange rates: The multi-state Markov-switching model with smoothing”. In: *International Review of Economics & Finance* 20.2. Heckscher-Ohlin Theory: A Modern Approach, pp. 342 –362. doi: [10.1016/j.iref.2010.09.002](https://doi.org/10.1016/j.iref.2010.09.002).
- Zargar, F. N. and D. Kumar (2019a). “Informational inefficiency of Bitcoin: A study based on high-frequency data”. In: *Research in International Business and Finance* 47, pp. 344 –353. ISSN: 0275-5319. doi: [10.1016/j.ribaf.2018.08.008](https://doi.org/10.1016/j.ribaf.2018.08.008).
- Zargar, Faisal Nazir and Dilip Kumar (2019b). “Long range dependence in the Bitcoin market: A study based on high-frequency data”. In: *Physica A: Statistical Mechanics and its Applications* 515, pp. 625 –640. ISSN: 0378-4371. doi: [10.1016/j.physa.2018.09.188](https://doi.org/10.1016/j.physa.2018.09.188).