Robust inference for random fields and latent models

MOLINARI, Roberto Carlo

Abstract

This thesis delivers a new framework for the robust parametric estimation of random fields and latent models through the use of the wavelet variance. By proposing a new M-estimation approach for the latter quantity and delivering results on the identifiability of a wide class of latent models, the thesis finally delivers a computationally efficient and statistically sound method to estimate complex models even when the data is contaminated. The results of this work are then implemented within a new statistical software which is also presented in this thesis, with a focus on its usefulness for inertial sensor calibration.

Reference


URN : urn:nbn:ch:unige-868996
DOI : 10.13097/archive-ouverte/unige:86899
ROBUST INFERENCE FOR RANDOM FIELDS AND LATENT MODELS

by

Roberto MOLINARI

A thesis submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,
in fulfillment of the requirements for the degree of
PhD in Statistics

Members of the thesis committee:
Prof. Maria-Pia VICTORIA-FESER, Co-Adviser, University of Geneva
Prof. Stéphane GUERRIER, Co-Adviser, University of Illinois at Urbana Champaign
Prof. Fabio TROJANI, Chair, University of Geneva
Prof. Roger KOENKER, External Member, University of Illinois at Urbana Champaign
Prof. Stephen PORTNOY, External Member, University of Illinois at Urbana Champaign

Thesis No. 32
August 2016
Acknowledgements

Writing the acknowledgments probably represents the hardest part of this research. I would probably need a manuscript of an equivalent length to this one to adequately thank each and every person who has helped, supported and accompanied me through this chapter of my life. However, given the time and practical constraints that such a task would require, I am forced to use all my capacity of synthesis to at least mention the people who have mainly contributed to making this experience as fulfilling as it has been.

First of all, my gratitude goes to my two supervisors Prof. Maria-Pia Victoria-Feser and Prof. Stéphane Guerrier. They both made the mistake of believing that I had the capacity of doing research and they were lucky enough that I was able to complete it. Maria-Pia encouraged me from the start and, after being a great professor during my master and teaching me how to lose at Metzle (or a game with a similar name), she gave me all the bases on which my PhD research was built, providing the inputs and insight without which I could not have completed this work. Despite the time-demanding constraints that forced her to put research on the back-burner over the last stretch of my PhD, she always managed to find time to discuss research problems with me and give me the support I needed to reach the final goal. As for Stéphane, I’ve always thought of him as my Sherpa. To put it in football terms: he’s the player who dribbles all the opponents from one end of the pitch to the other and then delivers the ball with a beautiful assist only for me to easily push it into the net from less than one metre away. It has been a true pleasure to collaborate with him and see his mind at work, as well as to appreciate his basketball and football skills at late hours in the university corridors. I owe a vast part of the results of this thesis to him and I deeply admire him. Maria-Pia and Stéphane have therefore been the driving force behind my research without whom I would not have been able to deliver this manuscript. So thank you.

I would then like to thank Prof. Fabio Trojani, Prof. Stephen Portnoy and Prof. Roger Koenker for giving me the honour of being part of my PhD committee. Fabio was kind enough to discuss some research aspects with me concerning the third chapter in this thesis and then kindly accepted to be the president of the committee, a role that he fulfilled by giving me important and interesting insights during the private defense. Stephen and Roger have been able to follow my research from a closer standpoint during my experiences as a visiting scholar at UIUC. Being kind enough to accept the presence of an inexperienced PhD student at their research discussions during coffee in the morning, they have always shown interest in my research and given me precious suggestions to achieve some of the results that are contained in this thesis. Moreover, as members of my PhD committee they contributed to making the private defense a very interesting discussion and provided me with corrections that allowed this manuscript to be improved. Another person that I would like to thank is Prof. Yanyuan Ma with whom I had the pleasure to collaborate on a couple of projects and I deeply appreciate all her talent, expertise, joyfulness and patience with me. Finally, I would also like to thank Prof.
Douglas Simpson and Melissa Banks for allowing me to be a visiting scholar at UIUC and being patient enough to help me with administrative and immigration issues! So thank you too.

As a research and teaching assistant, there are many other people that have allowed my PhD experience to be as enjoyable and fulfilling as it has been. I would first of all like to thank Prof. Elvezio Ronchetti who, as director of the PhD program, has formally allowed me to attain this goal after having inspired me through his lectures during my master. Moreover, I would like to thank Prof. Eva Cantoni and Prof. Diego Kuonen for having been patient enough with me as their teaching assistant and for organizing the courses in such a way as to minimize the amount of work I needed to carry out for the seminars and exams. I must also thank a series of other people who, although not being direct collaborators, have always been there to help me through administrative or other problems. Sandra Vuadens was the first (and last) person who helped me apply to the PhD and carry out all the necessary steps for my status to remain on track until the end. She always did so with a smile on her face which made even complicated situations seem unproblematic. Karen Longden, Karine Rama and Sylvie Huber were always extremely kind and available to help me for any issue I had to deal with at any point throughout my experience as teaching assistant. Sylvie Dreyfus has shown me great support since she joined and has always given me great encouragement by believing in my capabilities. So, to all of you, thank you so much.

Now it’s time for the people who have given me the strength, affection and support to come this far...

My parents: there are so many things that would need to be said and no good enough way to say them. You raised me to be a joyful and respectful human being with ambition to fulfill my aspirations and with kindness towards other people. I hope that I have become what you hoped I would be and I promise I will continue to follow the road you paved for me hoping to make you proud. You’re the reason I’m a happy person and you are present in all my achievements that are therefore also yours. And just in case my dad hasn’t understood: “papozzi, siete li mejo e ve vojo bene!” Grazie per tutto quello che mi avete dato, che mi date e che mi darete. Sono veramente fortunato ad avervi come genitori e la vostra gioia e il vostro sorriso mi danno la forza per andare avanti ed essere felice.

My family: not that my parents aren’t part of the family but I just needed a term to include the wider circle of people who have contributed to my upbringing and education outside the household environment. My brother Armando and my zia Giuliana have always made me feel loved and given me their moral (and material) support to reach the goals I gave myself in life. Armando allowed me to start my experience at the University of Geneva in a light-hearted manner by accompanying me from Rome and then getting lost together on the Salève while zia Giuliana has helped me through all my studies giving me affection, financial support when needed (as Armando too) and extra motivation each time she cooks for us when I’m back in Rome. Zia Emma, zia Armandina and Manuela have given me support and laughter teaching me that to be happy you should never take yourself (and life) too seriously. Lucia and Livia are the most recent members of the family but nonetheless have given me (as well as my brother of course...) great joy since they have entered our lives. Zio Toto, zia Silvana, zia Valeria, Angelo, Simona, Catalin, zio Marcello, zia Marta, zio Franco, zia Marilena, zio Giorgio, zia Marisa and zia Rossana have all encouraged me since I was young in various phases of my life and have given me affection and serenity when needed. Uncle Chris, Anthea, Bronwyn and Harj are the
Aussie/English family who, despite the distances, have always made me feel part of their family knowing that I can always count on them for help and support. Along with mum, uncle Chris has inspired me through his love for teaching and his candid enthusiasm for life. Finally, I’d like to thank Emanuele who, although not strictly part of the family following formal definitions, has given me and my parents strength and support, along with the wonderful Simona, allowing me to pursue my goals knowing that his generosity is a gift we can always count on. To all of you: thanks from the bottom of my heart.

My friends: this category is just to avoid making one long and tedious block of acknowledgments since I consider my friends as part of my family as well (although this block is long and tedious nonetheless). They are all true friends because despite my incapacity of staying in contact with them, they give me the same amount of affection each time we see each other and are extremely important to me even if I don’t show it. With no specific importance given to the order, I would first like to thank Chris and Colin who have been my ad-hoc parents since I started my experience in Geneva. With extreme kindness they have been great friends to my parents and have helped me and shown me their support in every occasion possible. I would then like to thank Anna Maria, Lorella and Mirella (aka the Trio) as well as Sheila, Lesley, Bridget, Fiona, Mary, Maria, Barbara, Wilma and Antonella for having been such good friends to me and my parents over the past years. A special thanks to Floriana who as my high school teacher has always believed in my capacities, shown deep interest in my life and pushed me to fulfill my dreams. Luisa Tibaldi was my last maths teacher at high school and it’s also thanks to her if I felt confident enough to change my career path towards Statistics. Silvia is the friend who supported me and encouraged me to do what I wanted, no matter how unreasonable it may sound. She has always shown interest, believed in me and been close to me (and my family) when needed. My highschool friends are those who brought me out of my shell, made me part of an incredible group and allowed me develop my personality (so they’re to blame for it). Federico and Mauro are those who have always been close to me and given me reasons to smile and laugh, making every problem I have fade away and allowing me to see everything in a positive light. They are acquired brothers and I thank them for being there for me. Angelo, Alessio, Massimo, Antonella, Cristiano, Eleonora, Federico (Turi) and Valentina have all been amazing friends and given me motivation by making me feel as if I can always rely on them no matter the time passed or the distance between us. Omar is another acquired brother and probably one of the main reasons I came this far. Since university in Rome, he has given me conviction in my skills and belief in the possibility of fulfilling my aspirations while giving me light-hearted serenity. I owe him a lot and I cherish his friendship. Along with Omar, Valentina and Claudia have given me support and laughter since I’ve known them, giving me unconditional friendship and encouragement. Marta and Rosita have supported me and given me affection despite all my shortcomings. My (Geneva) university friends have made the journey that has led to this manuscript the best I could wish for and I hope they will continue to accompany me in different ways for my following chapters. Stéphane, Samuel, Elise, Marco, Daniel, Kushtrim, Rose, Sezen, Marc-Olivier, Johanna, Caroline, Ozgu, Lea, Jonathan, Haotian, Rami, Gaetan: you have all made me smile, laugh, feel supported and opened my mind to different ways of thinking, helping me overcome problems and highlighting new ways to face upcoming challenges. I owe a lot to all of you for having made my PhD experience so wonderful. Walid has been the friend with whom I seem to share the most and the least at the same time: despite my annoying habits and our occasional clash of personalities, he has always offered me undivided and unconditional friendship by helping me in every
occasion possible and I hope I will be able to repay him notwithstanding my bizarre way of keeping my friends close. Amira is my soul sister who, no matter what happens, I know I will always be able to count on and I thank her for bringing so much warmth and joy to my life. Mattia and Sanda have been a hub of serenity and a mini-family for me over the past years and I thank them for the many free and tasty lunches and dinners that have given me energy to complete this PhD. Julien, Marie, Jivko, Laura, Antonio: you have been wonderful classmates and friends giving me laughs and extra motivation to continue on the academic path in Statistics. James is the first person I met at my arrival at UIUC and is a friend that has definitely left an indelible trace in my life: aside from being a driving force behind the fourth and fifth chapter of this manuscript due to his computational prowess, he has been that friend who can drive you nuts but is so adorable since he is also willing to do anything for you if you need help. Bo is a friend who, despite being a professor and a great researcher, has always made me feel at ease and given me the opportunity to discuss research issues without fear of being judged. I thank her for all her help and availability as well as her joy and bowling skills. Rami offered me true friendship from the start, giving me help and allowing me to take my mind off work from time to time (even though I don’t seem to have trouble finding distractions as Stéphane probably knows . . . ). Alessandro and Roberto provided me with some Italian getaways while in the US and have always given me enthusiasm to continue working when I was tired. Avinash, RagHAV, Yuming: thank you for helping me through the different projects we worked on. Jill and Lianne are acquired sisters: despite my absence they have always given me happiness and comfort when I needed it. I thank them for all the moments we passed together and I hope there will be others to look forward to. Stefano (Kiwi), Stefano (Crepaz), Marco (Finessi): you have always made me feel well and given me smiles and light-heartedness which has helped overcome problems and look towards the future without a shadow of worry. I also thank the Trulli who have given a background to our friendship, made me feel constantly welcome and kept me nourished. Konark is another of my acquired brothers: we started our experience in Geneva together and grew and matured together (to be taken with a pinch of salt) over these past years, sharing emotions, fears and laughter. You’re a person who I know I can count on always and I hope you know it will always be the same on my side. Nicoletta (Zio) is my acquired sister and one of the most wonderful people I know. She has given me joy and a reason to smile each time I see her, making everything look simpler. Alessia and Anna are also very important to me. Although we basically have the same age, Alessia has been an elder sister to me and given me lots of affection and support while Anna has always been there for me and made me laugh with her unique personality. Rose and Alessandra have always been close to me and I hope they know how much I appreciate their friendship. I would also like to thank Annalisa, Sara, Arianna and all my friends from the Ghiribizzo who have accompanied me through my experience in Geneva by providing me a sense of belonging and tons of laughter. Katia, Vaso, Pilar and Gustavo: we became close during my first experience in Geneva back in 2006 and I was lucky enough to find you again when I came back. You represent a group of friends with whom I always feel at home and who have helped me through many difficult moments. I consider you as my mini-family and I hope we will always be able to find occasions to be all together many times in the future. Brisilda, Manuela, Flo, Giorgio, Jordan: you’re wonderful people and I thank you for all the good times that have made this experience so great. There are many other people whose paths crossed mine and have given me support and happiness: to them I apologise if I haven’t mentioned them in these tedious acknowledgments but I will always
appreciate the moments we spent together (and may spend in the future). To all of you: thanks for your friendship and for making me the person I am today.

To Elena: you have accompanied me through the final stretch of this work. Although you don’t think so, you have given me a lot and I thank you for it from the bottom of my heart. No matter where our paths will lead, I want you to know that you have made me incredibly happy and I hope I can benefit from this for just a little longer (until Novembre or slightly beyond).
Abstract

The focus of this thesis is twofold. Firstly it delivers a new M-estimation framework for the quantity called Wavelet Variance (WV) which clarifies the conditions needed for this estimator to respect certain asymptotic properties for time series, spatial processes and multidimensional random fields. Based on these results, it is possible to ensure that the growing use of the WV also outside the natural sciences and engineering is made in a suitable manner. Indeed, the WV is increasingly being used in dependent data settings for statistical inference going from Portmanteau and isotropy tests to parametric estimation of time series and spatial models. Moreover, this extension to an M-estimation framework allows this estimator to be made robust by choosing a bounded score function. Therefore, aside from generalizing the standard estimator of WV, this M-estimator provides a robust solution that overcomes the limitations of an existing robust estimator of WV by comparing considerably better in finite sample and benefiting from clear asymptotic properties. The robustness properties of this estimator consequently transfer to any estimator or statistical test which makes use of this quantity thereby considerably extending the possibility of performing statistical inference in different dependent data settings that suffer from some form of contamination.

Secondly, this thesis has a specific focus on latent models that can be defined as models made by the sum of different underlying models. This class of models includes a wide variety of state-space and structural models for time series and delivers a flexible framework for the modelling of spatial and higher-dimensional dependent data in general. The use of these models is widespread in time series analysis for different fields going from economics to engineering and hydrology. However, these models are generally used under the assumption that they are identifiable, meaning that the methods used to estimate the parameters of these models are assumed to asymptotically have a unique solution that corresponds to the true parameter value. In this thesis we investigate this assumption for a wide class of intrinsically second-order stationary latent time series and spatial models, proving the identifiability of these different models for some extremum estimators and underlining those latent models that don’t respect this assumption. These results allow to considerably reduce the conditions needed to achieve consistency for different commonly used estimators and create a flexible modelling framework for random fields, in particular for time series and spatial models.

These two subjects of research allow this thesis to finally deliver a robust framework for the parametric estimation of time series and spatial (latent) models based on the Generalized Method of Wavelet Moments (GMWM). This method uses the WV as an auxiliary parameter that provides a computationally efficient approach to estimate the parameters of complex time series and spatial models. Using the new M-estimation framework for the WV, the GMWM can be made robust in a straightforward manner overcoming some practical limitations of existing robust methods. This new approach for robust inference on parametric random fields is implemented within a new R package called gmwm that
provides the first readily-available and computationally efficient tool for robust inference on time series models going from autoregressive moving-average to state-space and structural models. In addition, specific features are added to the package to model the complex stochastic error signals coming from inertial sensors allowing to efficiently carry out this task for their calibration.
Résumé

L’objectif de cette thèse est double. Tout d’abord, il offre un nouveau cadre de M-estimation de la quantité appelée Wavelet Variance (WV) qui précise les conditions nécessaires à cet estimateur afin de respecter certaines propriétés asymptotiques pour les séries chronologiques, processus spatiaux et random fields multidimensionnels. Sur la base de ces résultats, il est possible de faire en sorte que l’utilisation croissante de la WV au-delà des domaines des sciences naturelles et de l’ingénierie est faite d’une manière appropriée. En effet, la WV est de plus en plus utilisé pour l’inference statistique dans les cas de traitement de données dépendantes allant des tests de Portmanteau et d’isotropie à l’estimation paramétrique des séries chronologiques et des modèles spatiaux. De plus, l’extension à un cadre M-estimation permet à cet estimateur d’être robuste en choisissant une fonction de score bornée. Par conséquent, en dehors de la généralisation de l’estimateur standard de WV, ce M-estimateur fournit une solution robuste qui surmonte les limites d’un estimateur robuste de WV déjà existant en comparant beaucoup mieux dans l’échantillon fini et bénéficiant de propriétés asymptotiques claires. De cette manière, les propriétés de robustesse de cet estimateur se transfèrent à un estimateur ou à un test statistique qui fait usage de cette quantité, ce qui étend considérablement la possibilité d’effectuer l’inference statistique dans des différents contextes de données dépendantes qui souffrent d’une certaine forme de contamination.

D’autre part, cette thèse porte l’accent aussi sur les modèles latentes qui peuvent être définis comme composés par la somme des différents modèles sous-jacents. Cette classe de modèles en comprend une grande variété de comme les state-space ainsi que les structural pour les séries chronologiques. En général, il fournit un cadre flexible pour la modélisation des données dépendantes spatiales et de dimensions supérieures. Par ailleurs, l’utilisation de ces modèles est très répandue dans l’analyse des séries chronologiques dans des différents domaines, tels que l’économie, l’ingénierie et l’hydrologie. Cependant, ces modèles sont généralement utilisés sous l’hypothèse qu’ils soient identifiables, ce qui signifie que les méthodes utilisées pour estimer les paramètres de ces modèles sont supposés avoir asymptotiquement une solution unique qui correspond à la vraie valeur du paramètre. Dans cette thèse, nous étudions cette hypothèse pour une large classe de modèles de séries chronologiques et spatiales latentes, en démontrant l’identifiabilité de ces différents modèles pour certains “extremum estimators” et en soulignant, par conséquent, les modèles latents qui ne respectent pas cette hypothèse. Ces résultats permettent de réduire considérablement les conditions nécessaires pour assurer la consistence des différents estimateurs couramment utilisés et de créer un cadre de modélisation flexible pour des random fields, notamment pour les séries chronologiques et les modèles spatiaux.

Enfin, ces deux sujets de recherche permettent à cette thèse d’offrir un cadre solide pour l’estimation paramétrique des modèles de séries temporelles et spatiales (latents) sur la base du Generalized Method of Wavelet Moments (GMWM). Cette méthode utilise le WV comme un paramètre auxiliaire qui fournit une approche computationellement efficace.
pour estimer les paramètres des séries temporelles complexes et des modèles spatiaux. En utilisant le nouveau cadre de M-estimation pour le WV, le GMWM peut être rendu robuste d’une manière simple en surmontant certaines limites pratiques des méthodes robustes existantes. Cette nouvelle approche pour l’inférence robuste sur les random fields paramétriques est rendu disponible dans un nouveau package R appelé gmwm qui fournit le premier outil efficace et directement disponible pour l’inférence robuste sur les modèles de séries chronologiques allant des modèles ARMA aux modèles state-space et structural. En outre, des caractéristiques spécifiques enrichissent cet outil dans la modelisation des signaux d’erreur stochastiques complexes provenant de capteurs inertiels, ce qui permet de réaliser la procédure de calibration de ces instruments de manière efficace.
## Contents

Acknowledgements

Abstract

Résumé

Introduction

1 Wavelet Variance for Random Fields: an M-Estimation Framework
  1.1 Introduction ............................................. 5
  1.2 Multidimensional Wavelet Variance ......................... 7
  1.3 M-Estimation of Wavelet Variance .......................... 9
    1.3.1 Comparison with existing results ...................... 14
  1.4 Simulation Studies ....................................... 18
  1.5 Applications ........................................... 22
    1.5.1 Surface Albedo .................................. 22
    1.5.2 Cloud Images .................................. 24
    1.5.3 Digital Elevation Models .......................... 26
  1.6 Conclusion ............................................. 27

2 On the Identifiability of Latent Models for Dependent Data
  2.1 Introduction ............................................. 29
  2.2 Injection of Covariance Models ............................. 31
  2.3 Consistency of Extremum Estimators for Latent Models ....... 34
    2.3.1 MLE and GMM Estimators ............................. 34
    2.3.2 GMWM Estimators .................................. 35
    2.3.3 Numerical Example ................................ 38
  2.4 Conclusion ............................................. 40

3 Fast and Robust Parametric Estimation for Time Series and Spatial Models
  3.1 Introduction ............................................. 41
  3.2 Robust GMWM for Random Fields ............................ 43
  3.3 Simulation Studies ....................................... 45
    3.3.1 Time Series Model Estimation ......................... 46
    3.3.2 Spatial Model Estimation ............................. 50
  3.4 Applications ........................................... 53
    3.4.1 Application to Inertial Sensors ....................... 53
    3.4.2 Application to Personal Saving Rates .................. 55
B.1.5 Unique mapping from the covariance function to the spectral density 130
B.1.6 Proof of Lemma 2.3.2 ...................................................... 131

C Appendices for Chapter 3 ......................................................... 133
   C.1 Short literature review .................................................. 133
   C.2 Proofs .................................................. 135
      C.2.1 RGMWM Asymptotic Properties .................... 135
   C.3 Additional Results from Simulation Studies .................. 136
   C.4 Additional Results from Application on Inertial Sensor Data 137

D Appendices for Chapter 5 ......................................................... 141
   D.1 Benchmarks .................................................. 141
   D.2 How to Import IMU Data ........................................ 142
   D.3 Optimization Starting Values Algorithm ....................... 142
   D.4 Code Used in Chapter ........................................ 142
To my family and friends (who to me are the same thing).
Introduction

The initial goal of this thesis was to develop a new method for the robust estimation of time series models which it indeed finally achieved. However, as many research projects, different questions and issues were raised throughout this work thereby widening the research spectrum and delivering additional results which are either weakly or strongly related to the content of this thesis. Having said this, there are two building blocks to this work:

- the proposal of an M-estimation framework for the Wavelet Variance (WV); and
- the study of the identifiability of a wide class of latent time series and spatial models.

These two blocks constitute the first two chapters of this thesis while the final three generally make use of these results to provide a new method for the robust parametric estimation of time series and spatial models, with a specific implementation of this method for time series in the \texttt{gmwm R} package as well as a focus on its usefulness in the domain of engineering for inertial sensor calibration. It must be emphasized that each of these chapters represents a separate and stand-alone research paper which however are closely connected since the last three papers (chapters) bring together and use the results of the first two. The connections between chapters are duly highlighted in each paper when this is needed.

With this premise, Chapter 1 presents the first paper that proposes a new framework for the estimation of the WV which is a quantity that is recovered from the wavelet decomposition of a stochastic process. The latter issues the wavelet coefficients whose variance gives exactly this quantity and, based on this, an analysis of variance of the process can be carried out. The WV is frequently used in a variety of domains such as the natural sciences and engineering where the variance of a process can be interpreted over different “scales”. In addition, this quantity is able to adequately summarize the “information” contained in the spectral density of second-order stationary processes and can provide similar information also for intrinsically stationary processes (i.e. non-stationary processes with stationary backward differences). Using this property, many methods have been recently proposed to deliver a variety of tests for dependent data settings such as, for example, unit-root tests, Portmanteau tests and isotropy tests. Moreover, it can be used as an auxiliary quantity to estimate the parameters of time series and spatial models through the Generalized Method of Wavelet Moments (GMWM) which is a main focus of this thesis. Given its widespread and growing use, Chapter 1 introduces a new M-estimation framework which generalizes and/or improves over existing estimators of the WV, extending it to multidimensional random fields and delivering different joint asymptotic results that are needed for an appropriate use of this quantity in the different estimation and inference settings mentioned earlier. Furthermore, choosing a bounded score-function for this M-estimator allows it to be robust to the presence of outliers or other forms of
contamination. The latter aspect is of particular importance since it delivers a robust estimator which improves over an existing robust estimator both in terms of finite sample performance as in terms of appropriate asymptotic properties. Due to this last feature, this M-estimator can easily be used within all methods which make use of this quantity to carry out robust statistical inference for different dependent data settings. In particular, this advantage is investigated specifically for the robust parametric estimation of time series and spatial models in Chapter 3.

The second paper dealing with the identifiability of latent models is included in Chapter 2. Within this thesis, the term latent models refers to those models that are the result of the sum of different underlying models. Although this is probably not a commonly used definition, these models are actually widely used and include many state-space and structural time series models that can be represented or interpreted in this manner. Given the frequent use of these models, their identifiability is generally assumed for convenience although this may not be obvious for certain simple model combinations and is even less evident when their complexity grows. In general terms, identifiability consists in the property of an estimator to asymptotically have a unique solution for a given estimation problem and that this solution corresponds to the true value of the quantity that is being estimated. This is an essential condition for any estimator to be consistent and is often extremely difficult to verify, although it would be appropriate to prove this property in some cases in order to show that this assumption is not too strong. In this sense, Chapter 2 delivers results on the identifiability for a wide class of second-order stationary latent time series and spatial models and discusses what these results imply for the consistency of different extremum estimators such as the Maximum Likelihood Estimators (MLE), Generalized Method of Moments (GMM) estimators and GMWM estimators. These results considerably reduce the conditions for consistency of these estimators, highlight the latent models that can (and cannot) be consistently estimated and also deliver a flexible framework for the modelling of spatial processes.

Finally, Chapter 3 presents a paper where a robust version of the GMWM is proposed, that we call RGMWM, thereby delivering a computationally efficient method to robustly estimate the parameters of (latent) time series and spatial models even for large sample sizes. This robust version uses the results of the M-estimator proposed in Chapter 1 within the GMWM, allowing the asymptotic properties of the RGMWM to follow directly from the properties of this new WV estimator. The good properties of the RGMWM, especially in terms of computational efficiency, are shown in different simulation studies and applications. The implementation of this estimator is described in Chapter 4 where the \texttt{gmwm} R package is presented. In the latter we give the details on how the GMWM inference framework is made available through this open-source package, describing the main features and functions it includes for the (robust) estimation of a wide range of intrinsically stationary (latent) time series models. The final chapter (Chapter 5) consists in a paper which basically replicates the contents of Chapter 4 but with a specific focus on the usefulness of the \texttt{gmwm} package for the purposes of inertial sensor calibration for which tailored functions have been developed. Indeed, the GMWM was initially put forward to estimate the complex latent models that characterize the stochastic error signals coming from inertial sensors and, given also the large sample sizes, the \texttt{gmwm} package finally provides an efficient and suitable platform for these tasks.

As mentioned at the start of this section, further research projects were carried out throughout this thesis and these lead to the following publications and conference proceedings:


Most of the papers and proceedings are related to the study of existing methods and the proposal of new ones to estimate and select models for the stochastic error signals coming from inertial sensors. On the other hand, the paper called “Estimation of Time Series Models via Robust Wavelet Variance” consists in a preliminary study on the possibility of obtaining a robust GMWM by using an already existing robust estimator of the WV. Based on the results of the simulation studies in this paper, we decided to propose a statistically more sound robust estimator of WV with appropriate asymptotic properties as highlighted in Chapters 1 and 3. The last paper is unrelated with the research topics dealt with in this thesis and proposes a new algorithm for model selection purposes which is particularly useful for the goal of gene selection based on their capacity to classify or predict certain responses. Under the assumption of sparse models, this algorithm allows to deal with high-dimensional data in a very flexible way and delivers results that generally compare favorably to existing model selection methods and algorithms.

The conclusion to this thesis manuscript highlights the advantages of its research as well as some points for improvement. Moreover, it underlines the future research paths that have already started being investigated or are planned as upcoming research projects.
Chapter 1

Wavelet Variance for Random Fields: an M-Estimation Framework

We present a general M-estimation framework for inference on the wavelet variance. This framework generalizes the results on the scale-wise properties of the standard estimator and extends them to deliver the joint asymptotic properties of the estimated wavelet variance vector. Moreover, this is achieved by extending the estimation of the wavelet variance to multidimensional random fields and by stating the necessary conditions for these properties to hold when the size of the wavelet variance vector goes to infinity with the sample size. Finally, these results generally hold when using bounded estimating functions thereby delivering a robust framework for the estimation of this quantity which improves over existing methods both in terms of asymptotic properties and in terms of its finite sample performance. The proposed estimator is investigated in simulation studies and different applications highlighting its good properties.

1.1 Introduction

The Wavelet Variance (WV) is a quantity that is frequently used in many domains, from finance to the natural sciences, since it analyses the variance of a stochastic process by decomposing it across scales. This provides the basis for an analysis of variance in dependent data settings which is extremely useful given that many natural phenomena can be interpreted from this perspective and different tests can be derived from this quantity. For an overview of the importance of the WV we refer the reader to Percival and Walden (2006). A few examples of more recent uses of the WV can be found in Gallegati (2012) where the WV is employed to study the dynamics of financial markets and in Jia et al. (2015) where this quantity is used to study the behavior of crude oil prices. In medicine and genetics the WV was used in Xie and Krishnan (2013) to detect electroencephalogram seizure and support epilepsy diagnosis while in J. Sanderson et al. (2015) it is used to characterize ancestry block structures from observed genomic patterns. Moreover, the WV has a widespread use in geophysics as highlighted, for example, in Foufoula-Georgiou and Kumar (2014). In particular, the Haar WV is directly linked to the Allan Variance (AV) which is a quantity that is frequently used in many domains for the scale-based analysis and parametric estimation of stochastic processes for example (see Percival, 2015, and references therein for a more detailed overview). In addition, when dealing with second-order intrinsically stationary processes (see R. Christensen, 1991), the WV is a quantity that well summarizes the “information” contained in the spectral density
function into a reduced number of parameters by taking averages of the latter over octave bands. Using this compact information, Guerrier et al. (2013b) proposed the Generalized Method of Wavelet Moments (GMWM) that uses the WV to estimate the parameters for complex time series models which, for example, are often employed in engineering for inertial sensor calibration purposes (see Titterton and Weston, 2004). These advantages are preserved also in the two-dimensional case which was already studied in Mondal and Percival (2012a) where they mention how the two-dimensional WV is used, for example, in texture classification and segmentation problems (Unser, 1995) as well as for the analysis of the thickness in soil (Lark and Webster, 2004).

Considering the widespread use of this quantity in different settings, the goal of this chapter is to propose a WV estimator for regularly spaced data with desirable properties. For this purpose, an M-estimation framework is put forward based on Huber’s Proposal 2 (HP2), thereby delivering the following advantages:

- It generalizes the standard estimator of WV proposed by Percival (1995) and preserves its scale-wise properties which were studied in Serroukh et al. (2000). Moreover it extends these properties to the case where the number of scales goes to infinity with the sample size and delivers the joint asymptotic properties of this estimator across all scales. Indeed, the joint asymptotic normality of the WV vector had never been clearly stated for a fixed number of scales and it had never been studied when the scales go to infinity for a one-dimensional stochastic process (or for multidimensional random fields). This therefore allows to better understand the properties of all the methods that make use of the WV vector for estimation and inference.

- It allows to obtain robust estimates of the WV if a bounded function is chosen for the proposed M-estimator. This provides a robust estimator that preserves the asymptotic properties mentioned in the previous point and compares favorably in finite samples to the one proposed by Mondal and Percival (2012b) under relatively weaker assumptions.

In addition to the above advantages, we trivially extend the WV to the multidimensional case using the definitions given in Mondal and Percival (2012a) and allow for non-Gaussian random fields. These advantages are important under many aspects. Indeed, the joint asymptotic properties are essential for any approach that uses the WV estimator at different scales for statistical tests or parametric inference. Recent examples include the methods proposed in Fan and Gençay (2010) and in Thon et al. (2015) who respectively use the WV vector to obtain a unit-root test and a test for isotropy in random fields. Another example is given by Gençay and Signori (2015) who use the WV vector to develop a Portmanteau test. Moreover, a robust estimator of the WV and its joint asymptotic properties can deliver a robust framework for wavelet-based inference and, for example, can deliver a robust approach for the parametric estimation of time series and spatial models if used within the GMWM (see Chapter 3).

Given the above motivation, this chapter is organized as follows. Section 1.2 introduces the straightforward extension of the WV to multidimensional random fields, thereby giving the definitions that will be used throughout the chapter. Section 1.3 defines the proposed M-estimator, delivers its asymptotic properties and highlights how it relates to existing results. Section 3.3 presents a series of simulation studies where the robustness properties of the new estimator are shown and finally Section 3.4 presents three applied examples
highlighting the usefulness of the new robust framework for stochastic processes and two-dimensional data analysis. Section 3.5 concludes.

1.2 Multidimensional Wavelet Variance

A wavelet decomposition can theoretically be applied to multidimensional random fields as emphasized in Kugarajah and Q. Zhang (1995), although in practice this is usually limited to one- and two-dimensional processes (in rare cases also three-dimensional processes). The WV can therefore be obtained from the wavelet coefficients resulting from this multidimensional decomposition and, for example, the two-dimensional extension of the WV was already studied in Mondal and Percival (2012a). In this section, we use the latter to extend the definition of the WV to multidimensional random fields and, for this reason, in the following paragraphs we formalize the notation and definitions used for this extension. To start, let \((X_k)_{k \in K}\) denote the random field on the integer lattice \(\mathbb{Z}^D\), with \(D < \infty\) and \(k \in K\) defined further on, and let \(K_d\) (with \(1 \leq k_d \leq K_d\)) be the length of the random field along dimension \(d\) \((d = 1, \ldots, D)\). Given these preliminary definitions, below we provide a list of notations that are going to be used throughout the chapter:

- \(k \in K = \{k \in \mathbb{N}_+^D | k_1 \leq K_1, \ldots, k_D \leq K_D\}\): the coordinates of the points on \(\mathbb{Z}^D\) for each observation, with \(K_d\) being the maximum number of observations along dimension \(d\). Based on this we define \(N = |K| = \prod_{d=1}^D K_d\) as the cardinality of the set \(K\) (i.e. the number of observations in the random field). To abbreviate notation, we will from this point denote a random field as \((X_k)\).

- \(\delta_{k,k^*}\): the Euclidean distance between two points \(k\) and \(k^*\).

- \(j \in J \subset J^* = \{j \in \mathbb{N}_+^D | j_1 \leq J_1, \ldots, j_D \leq J_D\}\): the sets containing the combination of scales of decomposition in each dimension where \(J_d = \lfloor \log_2(K_d) \rfloor - a_d\) and
  \[a_d = \log_2 \left( \frac{2K_d(L_1 - 1)}{K_d + 2L_1 - 4} \right),\]
  with \(L_1\) being the length of the wavelet filter at the first scale. Based on this we define \(J = |J| = \prod_{d=1}^D J_d\) as being the cardinality of the set \(J\) (i.e. the number of scales of decomposition).

- \(M_{jd} = K_d - L_{jd} + 1\): the number of wavelet coefficients issued from a decomposition along dimension \(d\) where \(L_{jd}\) represents the length of the wavelet filter at scale \(j_d\).

- \(k_j \in K_j = \{k_j \in \mathbb{N}_+^D | k_1 \leq M_{j1}, \ldots, k_D \leq M_{jd}\}\): the coordinates of the points on \(\mathbb{Z}^D\) for each wavelet coefficient belonging to the decomposition at the set of scales \(j\). Based on this we define \(M_j = |K_j| = \prod_{d=1}^D M_{jd}\) as the cardinality of the set \(K_j\) (i.e. the number of wavelet coefficients issued from a decomposition at the set of scales \(j\)).

- \(l \in L = \{l \in \mathbb{N}_+^D | l_1 \leq L_{j1}, \ldots, l_D \leq L_{jd}\}\): the sets containing the combination of positions along a wavelet filter in each dimension.

- \((h_{jd,l_d})\): the Daubechies wavelet filter for scale \(j_d\) which satisfies

  \[\sum_{l_d=0}^{L_{jd}-1} h_{jd,l_d} = 0 \quad \text{and} \quad \sum_{l_d=0}^{L_{jd}-1} h_{jd,l_d}^2 = \frac{1}{2j_d}.\]
• (W_{k,j}) : the vectorized process of wavelet coefficients issued from the D-dimensional wavelet filtering at the set of scales j (see further on). These coefficients are placed in descending order according to \( \delta_{k,k^*} \).

**Remark 1.2.1.** The definition of the subset \( \mathcal{J} \) allows the wavelet coefficients at the last scale of decomposition (in a general dimension \( d \)) to go to infinity as the sample size goes to infinity. Notice that we have \( \lim_{K_d \to \infty} a_d = \log_2(2(2L_1 - 1)) \) for any \( L_1 < \infty \) and, in particular, \( a_d = 1 \) exactly when using the Haar wavelet filter.

To clarify this notation, let us give the following examples.

**Example 1.2.1.** Suppose we have a one-dimensional random field (i.e. a stochastic process or time series) which we observe on a regular grid in time \( t = 1, \ldots, T \). In this case the sets would reduce to scalars where \( k \) would correspond to \( t \) while \( N \) would correspond to \( T \), with the distance between time points being \( \delta_{t,t^*} = \sqrt{(t - t^*)^2} \). The wavelet coefficients would therefore be defined as \( (W_{t,j}) \), naturally ordered according to \( \delta_{t,t^*} \), for \( j = 1, \ldots, J \) and, using the Haar wavelet filter, \( J = \lfloor \log_2(T) \rfloor - 1 \).

**Example 1.2.2.** Suppose we have a two-dimensional random field which we observe on a regular grid with coordinates \( k = (k_1, k_2) \), where \( k_1 = 1, \ldots, K_1 \) are the coordinates in dimension \( d = 1 \) and \( k_2 = 1, \ldots, K_2 \) are the coordinates in dimension \( d = 2 \), which gives a random field of size \( N = K_1K_2 \) with \( \delta_{k,k^*} = \sqrt{(k_1 - k_1^*)^2 + (k_2 - k_2^*)^2} \). We therefore have \( j = (j_1, j_2) \), with \( j_1 = 1, \ldots, J_1 \) and \( j_2 = 1, \ldots, J_2 \), where, using the Haar wavelet filter, we have \( J_1 = \lfloor \log_2(K_1) \rfloor - 1 \) and \( J_2 = \lfloor \log_2(K_2) \rfloor - 1 \).

Keeping this notation in mind and considering the results of Kugarajah and Q. Zhang (1995) and Mondal and Percival (2012a), we start by defining the \( D \)-fold wavelet coefficients as

\[
W_{k,j} = \sum_{l \in \mathcal{L}} h_{j_1,l_1} \cdots h_{j_D,l_D} X_{k_1-l_1, \ldots, k_D-l_D}. \tag{1.2.1}
\]

Given this definition it can be noticed that \( \mathbb{E}[W_{k,j}] = 0 \) since we generally assume that \( (X_k) \) (or its centered differencing) is a second-order stationary random field. Having defined the wavelet coefficients for the set of scales \( j \), we can now define the corresponding WV, \( \nu_j^2 \), as

\[
\nu_j^2 \equiv \text{var} \left( W_{k,j} \right)
\]

which, along with (1.2.1), allows to obtain an expression for the model-implied WV \( \nu_j^2(\theta) \). For this purpose, referring to the notation used in Mondal and Percival (2012a), let \( \delta_u \) represent the Euclidean norm of a vector \( \mathbf{u} \) and let \( \varphi(\delta_u) \) denote the theoretical covariance function of the random field \( F_\theta \). Then we have

\[
\nu_j^2(\theta) = \sum_{l \in \mathcal{L}} h_{j_1,l_1} \cdots h_{j_D,l_D} \varphi(\delta_{l_j}). \tag{1.2.2}
\]

To simplify notation, from now on we will denote this quantity as \( \nu_j^2 \). In the one-dimensional setting, known analytic expressions are available for this theoretical WV based on the Haar wavelet filter by using the results of N. F. Zhang (2008). However, if the theoretical covariance function of the random field \( (X_k) \) is known, it is possible to obtain analytic expressions for any WV based on (1.2.2). An unbiased estimator of this quantity was proposed by Percival (1995) and is given by

\[
\hat{\nu}_j^2 = \frac{1}{M_j} \sum_{k_j \in \mathcal{K}_j} W_{k,j}^2. \tag{1.2.3}
\]
1.3. M-Estimation of Wavelet Variance

The theoretical properties of this estimator were studied in Serroukh et al. (2000) in which the conditions for its asymptotic properties are given. However, as highlighted by Mondal and Percival (2012b), this estimator of WV is not robust and can therefore become considerably biased in the presence of outliers or different forms of contamination. In order to deliver a WV estimator that can be robust, let us re-express (1.2.3) as an M-estimator with $\hat{\nu}_j^2$ being the implicit solution of

$$
\sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) = 0,
$$

where $\psi(\cdot)$ is a score function which can be unbounded or bounded with respect to $(W_{k,j})$. The following theorem states the sufficient condition under which the WV estimator has a bounded influence function and is therefore robust.

**Theorem 1.2.1.** Assuming that $(W_{k,j})$ is a strictly stationary and ergodic process and that contamination in dimension $d$ is independent from the contamination in dimension $d'$, $\forall d \neq d'$, the IF of the estimator of WV is bounded if $\psi(\cdot)$ is bounded.

The proof of Theorem 1.2.1 follows the standard proofs for these settings and can be found in Appendix A.3.1. The definition of a multidimensional contamination model and relative IF is a topic of wider investigation (see, for example, Alqallaf et al., 2009) but an intuitive argument based on Theorem 1.2.1 would suggest that a bounded $\psi(\cdot)$ function would nevertheless be required to ensure robustness also in different multidimensional contamination settings. Given these considerations, the next section proposes a new M-estimator for the WV based on HP2.

1.3 M-Estimation of Wavelet Variance

Following the form of the estimator defined in (1.2.4), this section generalizes the standard estimator of WV in (1.2.3) to an M-estimator and, hence, the properties of this new M-estimator are valid also for the standard estimator. The proposed approach, as mentioned earlier, is based on HP2 (Huber, 1981) that was put forward for the estimation of the scale parameter of the residuals in the linear regression framework. Since in general we suppose that $E[W_{k,j}] = 0$, we use HP2 by defining $r_{k,j} = W_{k,j}/\nu_j$ as the standardized wavelet coefficients, and $\hat{\nu}_j^2$ is defined implicitly as the solution in $\nu_j^2$ of

$$
\frac{1}{M_j} \sum_{k_j \in K_j} \omega^2(r_{k,j}; \nu_j^2, c) r_{k,j}^2 = a(\nu_j^2, c) = 0,
$$

where $r_{k,j}^2$ is the score function, $\omega(\cdot)$ represent the weights given to this score function and $a(\nu_j^2, c)$ is a correction term to ensure Fisher consistency at the unconditional distribution of the wavelet coefficients $(W_{k,j})$. The tuning constant $c$ regulates the trade-off between robustness and efficiency, where $c \to \infty$ corresponds to the classical estimator (with $\omega(r_{k,j}; \nu_j^2, c) = 1, \forall k_j$, and $a(\nu_j^2, c) = 1$). A discussion about the choice of this constant can be found in Appendix A.1. The term $a(\nu_j^2, c)$ depends on the distribution of the stationary process which is assumed for the wavelet coefficients $(W_{k,j})$ and on the specific weight function $\omega(\cdot)$. The exact analytic form of this term therefore may be complicated to derive depending on the distributional assumption. Nevertheless this problem can be overcome, for example, by using indirect inference which allows this estimator to be adapted to different distributional assumptions for $(W_{k,j})$. 
Remark 1.3.1. In the case where the wavelet coefficients are assumed to follow a Gaussian process, then the correction term \( a(\nu_j^2, c) \) can be expressed as \( a_\psi(c) \), since it only depends on the value of the tuning constant \( c \), and can be found explicitly using the results of Dhrymes (2005).

Having defined the M-estimator of WV in (1.3.1), let us now define \( \nu_0^2 \) as being the true WV for a general scale \( j \) and list a set of conditions which allows us to derive the asymptotic properties of this estimator:

(C1) \( \nu_0^2 \in \{ x \in \mathbb{R} \mid 0 < x < \infty \} \).

(C2) \( \mathbb{E}[\psi(W_{k,j}, \nu_j^2)] = 0 \) if and only if \( \nu_j^2 = \nu_0^2 \).

(C3) The spectral density \( S_\psi(\cdot) \) of the process \( (\psi(W_{k,j}, \nu_j^2)) \) is strictly positive at zero-frequency (i.e. \( S_\psi(0) > 0 \))

(C4) \( (W_{k,j}) \) is stationary with mixing coefficient \( \alpha_{W,k_j} \) such that \( \sum_{k_j=1}^{\infty} (\alpha_{W,k_j})^{\nu_j + \delta} < \infty \) for some \( \delta > 0 \).

(C5) \( \mathbb{E}[W_{k,j}] = 0 \) and \( \mathbb{E}[W_{k,j}^4] < \infty \).

(C6) The function \( \psi(W_{k,j}, \nu_j^2) \) is Bouligand-differentiable and \( |\partial / \partial \nu_j^2 \mathbb{E}[\psi(W_{k,j}, \nu_j^2)]| < \infty \).

Condition (C1) requires the true WV \( \nu_0^2 \) to belong to a compact support and is a very mild condition given the results of Huber (1967). However, a strong assumption is made by Condition (C2) which requires the WV to be identifiable based on the function \( \psi(W_{k,j}, \nu_j^2) \) and this is not an easy condition to verify in general.

Remark 1.3.2. Condition (C2) has been proven for the Huber and Tukey \( \psi \)-functions under the assumption that the wavelet coefficients \( (W_{k,j}) \) follow a Gaussian process (see Theorem A.3.1 in Appendix A.3.2).

Condition (C3) is a standard and mild assumption to make in order to obtain an asymptotic distribution for the estimator. Conditions (C4) to (C6) allow to use specific results to prove consistency and asymptotic normality. In particular, Conditions (C4) and (C5) directly refer to the wavelet coefficient process \( (W_{k,j}) \) where the mixing coefficient \( \alpha_{W,k_j} \) allows to achieve an appropriate rate of convergence of the estimator which is useful for the joint asymptotic properties (given further on) while the conditions on the moments are helpful to prove uniform integrability for the central limit theorem to be applied. The latter condition is directly satisfied, for example, if the random field is second-order stationary and Gaussian.

Remark 1.3.3. Condition (C4) becomes stronger to assume as the dimension of the random field grows. Indeed, with larger \( D \), there could be subsets of \( (W_{k,j}) \) where the dependence structure does not decay since they will be based on the same distance \( \delta_{k,k} \) of the underlying random field \( (X_k) \). This condition is less strong in the one-dimensional setting where the filtering only occurs in one direction therefore avoiding a repeated dependence structure of the stochastic process which instead could occur when adding other dimensions in the decomposition. As a final note to this, if the random field is assumed to be isotropic, this limitation could be removed by simply removing the coefficients based on the same distance \( \delta_{k,k} \) (or by taking an average over them).
Finally, Condition (C6) allows to use the standard expansions to prove the asymptotic normality of the proposed estimator and to achieve appropriate rates of convergence for the results to hold.

**Remark 1.3.4.** Condition (C6) is respected when using the Huber or Tukey $\psi$-functions. The Bouligand differentiability of the Huber $\psi$-function is discussed in Appendix A.3.3.

Given these conditions, the scale-wise asymptotic properties of the proposed estimator are given in Theorem 1.3.1 below.

**Theorem 1.3.1.** Under Conditions (C1) to (C6) we have that

$$\sqrt{M_j}(\hat{\nu}_j^2 - \nu_j^2) \xrightarrow{D} N(0, \frac{S_\psi(0)}{m_j^2})$$

where $S_\psi(0)$ is the spectral density of $\psi(W_{k,j}, \nu_j^2)$ at zero-frequency and

$$m_j = \mathbb{E} \left[ \frac{\partial}{\partial \nu_j} \psi(W_{k,j}, \nu_j^2) \right].$$

The proof of this theorem can be found in Appendix A.3.3. Theorem 1.3.1 generalizes the results of Serroukh et al. (2000) to M-estimators under much the same conditions, given that Condition (C1) is implicitly assumed in their work. The main differences lie in Conditions (C2), (C5) and (C6) where, instead of condition (C5), Serroukh et al. (2000) use the condition that $\mathbb{E}[W_{k,j}]^{4+2\delta} < \infty$, $\forall \delta > 0$. Conditions (C2) and (C6) are simply needed to generalize their results to M-estimators based on specific $\psi$-functions. A more detailed comparison with existing conditions and results in Serroukh et al. (2000) and Mondal and Percival (2012b) can be found in Section 1.3.1. For some results in this chapter we consider the Huber and Tukey $\psi$-functions, suggesting to use the Tukey $\psi$-function based on its renown high breakdown point and its bias properties compared to the Huber function. To support the latter case, in Appendix A.2 we compare Huber’s and Tukey’s weights in the way they control the bias induced by different types of contamination on the resulting estimator. This study leads us to conclude that Tukey’s weights appear to be more appropriate for overall bias reduction.

The results of Theorem 1.3.1 apply to the estimator $\hat{\nu}_j^2$ for each scale $j$ thereby delivering the scale-wise asymptotic properties of the proposed M-estimator. However different results, such as the isotropy tests in Thon et al. (2015) or the asymptotic properties of the GMWM, rely heavily on the joint asymptotic properties of the vector of estimated WV $\hat{\nu}$. In the following paragraphs we will therefore give some results on the joint properties of the proposed M-estimator when we let $J \to \infty$ as $N \to \infty$. The cases where we keep $J$ fixed are a special case whose properties are easily derived based on these results. Moreover, the latter case has already been studied in Guerrier et al. (2013b) within the specific case of the time series setting under different conditions. For this purpose, let us define $M_j = \min_d M_{J_d}$ and

$$M_{J_d} = (M_j)^D,$$

which allow to define $k_j \in K_{J_d} = \{k_j \in \mathbb{N}_+^D | k_1 \leq M_j, \ldots, k_D \leq M_j \}$. With these definitions we intend to study the joint asymptotic properties of the proposed estimator based on the scale of decomposition which delivers the fewest wavelet coefficients. With this in mind, let us consider also the following condition.
Condition (C7) implies that $M_{js}$ goes to infinity at a faster rate than $J$. This is needed since we are considering $D$-dimensional random fields where it is possible for $N \to \infty$ even when we only have one $K_d \to \infty$. In the latter case, we would therefore have that $M_{js}$ would not go to infinity even though the sample size does. Hence, if we define $K_{ds}$ as being the smallest of the $K_d$'s, we would need $K_{ds}$ to go to infinity at a rate which is sufficiently faster than the growth of the set $J$. However, in the case where $J$ is fixed, all that is needed is simply that $K_{ds} \to \infty$.

Remark 1.3.5. In the one-dimensional setting, such as stochastic processes and time series, Condition (C7) is always verified. In higher-dimensional settings, this condition becomes stronger but can still be verified and becomes

$$\frac{\prod_{d=1}^{D} (|\log_2(K_d)| - a_d)}{\sqrt{(K_{ds} - L_{J_{ds}} + 1)^D}} \to 0$$

which implies that $K_{ds}$ should be sufficiently “close” to the other $K_d$’s and grows at a similar rate. For example, if we define $K_d = K \forall d$ (meaning that the number of observations along each dimension are the same), then Condition (C7) is respected (see proof in Appendix A.3.4).

With the above considerations, let us study the joint asymptotic properties of $\hat{\nu}$ by defining $\| \cdot \|$ as the Euclidean norm and, to simplify notation, $\nu \equiv \nu(\hat{\theta})$. This allows us to give the following corollary.

Corollary 1.3.1. Under the conditions of Theorem 1.3.1 and condition (C7) we have that

$$\| \hat{\nu} - \nu \| \xrightarrow{p} 0.$$ 

This corollary therefore states the mean-squared consistency of the estimator $\hat{\nu}$ regardless of whether $J$ is fixed or goes to infinity with the sample size $N$. Before studying the asymptotic distribution of $\hat{\nu}$, let us first consider a different scale-wise WV estimator which we call $\bar{\nu}_j^2$ defined as the implicit solution in $\nu_j^2$ of

$$\frac{1}{M_{js}} \sum_{k_j \in K_{js}} \omega^2 \left( r_{k_{j,j}}; \nu_j^2, c \right) r_{k_{j,j}}^2 - a(\nu_j^2, c) = 0.$$ 

(1.3.2)

This estimator is exactly the same as the estimator proposed in (1.3.1) except that is based on fewer observations (i.e. $M_{js}$ instead of $M_j$) and the reason for defining this estimator will become apparent further on. Let us denote the corresponding estimated vector as $\bar{\nu}$ and give the following condition.

(C8) The function $\psi(W_{k_{j,j}}, \nu_j^2)$ is twice Bouligand-differentiable.

Remark 1.3.6. The Huber $\psi$-function does not satisfy condition (C8).

This condition is useful to obtain results for some technical corollaries and lemmas in Appendix A.3.6 which are followed by the proof of Theorem 1.3.2 given below.
Theorem 1.3.2. Let $\nu$ be the implicit solution of
\[ \sum_{k_j \in K_{j,*}} \Psi(W_{k_j}, \nu) = 0 \]
where
\[ \Psi(W_{k_j}, \nu) = \begin{bmatrix} \psi(W_{k_1,1}, \nu_1^2) \\ \vdots \\ \psi(W_{k_J,1}, \nu_J^2) \end{bmatrix} \]
and with
\[ \psi(W_{k_j,1}, \nu_1^2) = \omega^2(r_{k_j,1}, \nu_1^2, c) \]

\[ \psi(W_{k_j,J}, \nu_J^2) = \omega^2(r_{k_j,J}, \nu_J^2, c) - a(\nu^2, c) \]  \hspace{1cm} (1.3.3)

defining a time-invariant function of $W_{k_j}$. Then, under the conditions of Corollary 1.3.1 and condition (C8), the asymptotic distribution of $\nu$ is given by
\[ \sqrt{M_{js}} s^T \Sigma^{-1/2} (\nu - \nu) \xrightarrow{D} \mathcal{N}(0,1) \]

where $\|s\| = 1$, $\Sigma = M^{-1} S_\psi(0) M^{-T}$ is the asymptotic covariance matrix of $\nu$ with $S_\psi(0)$ being the power spectral density of $\Psi(W_{k_j}, \nu)$ and
\[ M = E \left[ - \frac{\partial}{\partial \nu} \Psi(W_{k_j}, \nu) \right] . \]

Although only valid for a “truncated” version of the estimator $\hat{\nu}$ since it only uses $M_{js}$ observations for each scale, this theorem delivers an estimator (i.e. $\bar{\nu}$) with adequate joint asymptotic properties which are extremely useful for any method which uses the joint properties of this estimator such as, for example, the GMWM. The extension of this result to the estimator $\hat{\nu}$ holds if the number of scales $J$ remains fixed but is not necessarily straightforward if $J \to \infty$, although it can be done based on the results of Theorem A.3.2 and Corollary A.3.3 in Appendix A.3.7. Indeed, based on these results, the theorem below gives the joint asymptotic normality of the estimator $\hat{\nu}$ in the one-dimensional case (i.e. $D = 1$) when $J \to \infty$ and using the Haar wavelet filter.

Theorem 1.3.3. Suppose that we have a stochastic process $(X_k)_{k \in K \subset \mathbb{N}^+}$ which respects the conditions in Theorem 1.3.1. Moreover, using the Haar wavelet filter, let $J = \lfloor \log_2(N^\alpha) \rfloor$ with $0 < \alpha < 1/2$ and $M_{js} = N - 2^J + 1$. Defining $\|s\| = 1$, we have that
\[ \sqrt{M_{js}} s^T \Sigma^{-1/2} (\hat{\nu} - \nu) \xrightarrow{D} \mathcal{N}(0,1) \]

The proof of this Theorem can also be found in Appendix A.3.7. This result allows to select a value of $J$ which goes to infinity with $N$ and preserves the joint normality of the proposed WV estimator when applied to a stochastic process.

Remark 1.3.7. It is straightforward to see that the results in Corollary 1.3.1, Theorem 1.3.2 and Theorem 1.3.3 hold for any $D$-dimensional field ($D < \infty$) when $J$ is fixed and assuming that $M_{js} \to \infty$.

The above results have allowed to generalize the WV estimator to an M-estimator whose scale-wise and joint asymptotic properties have been extended to multidimensional random fields and to cases where $J \to \infty$. Moreover, by choosing a bounded $\psi$-function, the proposed estimator can be made robust thereby allowing for WV-based analysis when the data is contaminated. Given these properties, the following paragraphs compare these results to those of the main existing estimators of the WV, namely the standard estimator of WV ($\tilde{\nu}_j^2$) and the robust M-estimator ($\bar{\nu}_j^2$) proposed by Mondal and Percival (2012b).
1.3.1 Comparison with existing results

As stated throughout the previous sections, the results of this chapter share conditions and extend on previous results for existing estimators of WV. Although these estimators were proposed for time series (and two-dimensional random fields), the conditions used for their properties to hold are roughly the same as those for the estimator proposed in the previous section. Considering this, first of all the estimator in (1.3.1), based on HP2, corresponds to the standard estimator of WV when the weights \( \omega(\cdot) \) are all equal to one, i.e. \( \nu_2 \) is the solution for

\[
\frac{1}{M_j} \sum_{k_j \in K_j} \frac{W_{k,j}^2}{\nu_2} - 1 = 0,
\]

which obviously has an explicit solution that corresponds exactly to the estimator in (1.2.3). Hence, the results presented in Section 1.3 need to be compared to those given in Serroukh et al. (2000). Moreover, the M-estimator in (1.3.1) is a robust alternative to the existing M-estimator proposed by Mondal and Percival (2012b) which was developed for Gaussian time series and random fields specifically affected by scale-based contamination and makes use of a log-transformation of the wavelet coefficients to apply standard M-estimation theory for location parameters. More specifically, the wavelet coefficients \((W_{k,j})\) are transformed as \(Q_{k,j} = \log W_{k,j}^2\) with location parameter \(\mu\). The implicit estimator for the solution point \(\mu_0\) is given by

\[
T_N = \arg\min_{\mu_0 \in \mathbb{R}} \sum_{k_j \in K_j} \psi(Q_{k,j} - \mu_0).
\]

The estimator \(T_N\) is consistent for \(\mu_0\) which does not necessarily correspond to the location parameter of interest \(\mu\) and, consequently, \(T_N\) is corrected for bias and then an inverse transform of \(T_N\) is applied to deliver the robust estimator of WV.

These two estimators are therefore the terms of comparison for the new M-estimator and let us first compare our conditions and results with those of Serroukh et al. (2000). Without loss of generality, we will only consider \((X_k)\) as being a stochastic process (i.e. one-dimensional random field). Having stated this, let us define the following settings:

(S1) Conditions in Lemma 1 in Serroukh et al. (2000):

(a) The stochastic process \((X_k)\) is a strictly stationary process;
(b) \(\mathbb{E}[|X_k|^{4+2\delta}] < \infty\) for some \(\delta > 0\);
(c) The process \((X_k)\) has mixing coefficient \(\alpha_{X_k} = \mathcal{O}(\rho^k)\) with \(k \to \infty\) and \(0 < \rho < 1\);
(d) The spectral density of the process \((X_k)\) is strictly positive at zero frequency;
(e) \(j < \infty\).

(S2) Conditions of Theorem 1 in Serroukh et al. (2000):

(a) The stochastic process \((X_k)\) is a strictly stationary process;
(b) \(\mathbb{E}[|W_{k,j}|^{4+2\delta}] < \infty\) for some \(\delta > 0\);
(c) The process \((W_{k,j})\) has mixing coefficient \(\alpha_{W_{j,k}}\) such that \(\sum_{k_j=1}^{\infty} (\alpha_{W_{j,k}})^j/2+\delta < \infty\) for some \(\delta > 0\).
1.3. M-Estimation of Wavelet Variance

(d) The spectral density of the process \( Z_{k,j} = W_{k,j} - \mathbb{E}[W_{k,j}] \) is strictly positive at zero frequency;

(e) \( j < \infty \).

(S3) Comparable conditions for the M-estimator \( \hat{\nu}_j^2 \):

(a) (C3) to (C5)

(b) \( j \in J = \{2^{1,\ldots,J}\} \) where \( J \to \infty \) by the definition given in Section 1.2.

The first two settings, as highlighted, correspond to the conditions given in Serroukh et al. (2000) where Setting (S1) gives sufficient (but not necessary) conditions for the conditions in Setting (S2) to hold. Hence, we have that Setting (S1) implies Setting (S2). On the other hand, Setting (S3) gives the conditions for the proposed estimator \( \hat{\nu}_j^2 \) which can be considered comparable to the other settings. Indeed, Conditions (C1) and (C2) are implicitly assumed in the other settings while Condition (C6) is not needed since the \( \psi \)-function gives the standard estimator in this comparison. The conditions in Setting (S2) and Setting (S3) are roughly the same except that in the latter we let \( j \to \infty \). Having said this, Setting (S2) allows to obtain scale-wise asymptotic normality and mean-square consistency of the estimated vector \( \hat{\nu} \) for \( j < \infty \). In addition to these results, Setting (S3) also allows \( j \to \infty \) and adds the results on the joint normality of the estimated vector \( \hat{\nu} \) since, for the standard estimator in the one-dimensional setting, Conditions (C7) and (C8) automatically hold. Moreover, remaining in the domain of stochastic processes and by defining \( J = \log_2(N^\alpha) \), \( 0 < \alpha < 1/2 \), Setting (S3) also delivers the joint normality of the estimated vector \( \hat{\nu} \) based on the Haar wavelet filter.

Let us now compare the proposed estimator with the one proposed in Mondal and Percival (2012b) and, for this purpose, let us give the following settings:

(S4) The stochastic process \((X_k)\) is Gaussian.

(S5) Conditions in Theorem 1 in Mondal and Percival (2012b):

(a) \((X_k)\) is a stationary stochastic process.

(b) \( \mathbb{E}[\psi(Q_{k,j} - \tilde{\mu})] = 0 \) if and only if \( \tilde{\mu} = \mu_0 \).

(c) The function \( \mathbb{E}[\psi(Q_{k,j} - \tilde{\mu})] \) is differentiable and is continuous around \( \mu_0 \).

(d) The process \((X_k)\) has a square integrable spectral density.

Setting (S5) partially overlaps with Setting (S3) (e.g. Conditions (C4) and (C5)) and adds the other conditions stated in this chapter (i.e. Conditions (C2) and (C6)). Condition (C1) is implicitly assumed for Mondal and Percival (2012b) as well, while an additional condition for them is given by Setting (S4) where they require the random field \((X_k)\) to be Gaussian. Considering these relations, under Settings (S3) to (S5), the estimator \( T_N \) is asymptotically normally distributed with expectation \( \mu_0 \). These results are not exactly comparable to the scale-wise properties of the estimator \( \hat{\nu}_j^2 \) since they refer to different quantities of interest. Nevertheless, for their respective quantities and with the exception of the condition that \((X_k)\) be a Gaussian process, these estimators basically share the same type of conditions to obtain scale-wise asymptotic normality and mean-square consistency, assuming the convergence rate of the estimator proposed by Mondal and Percival (2012b) remains unaffected by the bias correction and inverse transform. If however the Gaussian assumption were kept and the Huber and Tukey \( \psi \)-functions
were considered for the estimator \( \hat{\nu}^2 \), then Setting (S5) would not be needed anymore for \( \hat{\nu}^2 \) since Conditions (C2) and (C6) would automatically hold (see Appendices A.3.2 and A.3.3). Moreover, as in the comparison with the settings established in Serroukh et al. (2000), the properties of the proposed estimator \( \hat{\nu}^2 \) can be extended to obtain joint asymptotic normality of the vector estimators \( \hat{\nu} \) and \( \hat{\nu}^\prime \). Figure 1.1 visually summarizes the statements made in this section, attempting to clarify the links between conditions and results of these three estimators.
1.3. M-Estimation of Wavelet Variance

\[ \hat{\nu}^2_j \] (Percival, 1995) \hspace{1cm} \hat{\nu}^2_j \quad \text{(with } \ c \to \infty) \\
\quad (S1) \hspace{6cm} (S2) \hspace{6cm} (S3) \\
\text{Mean-Square Consistency} \hspace{1cm} \text{Scale-wise Asymptotic Normality} \hspace{1cm} \text{Joint Asymptotic Normality} \\
\|\hat{\nu} - \nu\| \xrightarrow{D} 0 \hspace{1cm} \sqrt{M_j(\hat{\nu}^2_j - \nu^2_j)} \xrightarrow{D} N(0, \sigma^2) \hspace{1cm} \sqrt{M_j s^2 \Sigma^{-1/2}} (\hat{\nu} - \nu) \xrightarrow{D} N(0, 1) \\
\quad (S3)+(S4)+(S5) \hspace{1cm} \quad (S3)+(S5) \quad \rightarrow \quad (S3)+(S4) \\
\quad \text{(Mondal and Percival, 2012b)} \hspace{1cm} \quad \hat{\nu}^2_j \quad \text{(with } \ c < \infty) \\

Figure 1.1: Summary of conditions and asymptotic properties for the standard estimator \( \hat{\nu}^2_j \), the robust estimator \( T_N \) and the proposed estimator \( \hat{\nu}^2_j \). Normal arrows indicate links with conditions; double-headed arrows denote equivalence or implication; small full-headed arrow denotes the consequent property under the condition that \( j \in J < \infty \); large full-headed arrow denotes the consequent property with \( J < \lfloor \log_2(T) \rfloor \); empty-headed arrow denotes the consequent property under the condition that \( J < \lfloor \log_2(\sqrt{T}) \rfloor \). The dashed lines denote approximate equivalence.
Chapter 1. Wavelet Variance for Random Fields: an M-Estimation Framework

Considering these conditions, let us now give the following corollary.

**Corollary 1.3.2.** Assuming that \((W_{k,j})\) is a stationary Gaussian process as well as \(J\) fixed, and under Conditions (C1), (C3), (C4), we have that the estimator \(\hat{v}\) based on the Tukey \(\psi\)-function is mean-square consistent and has an asymptotic multivariate normal distribution given in Theorem 1.3.2.

This corollary is useful to state the conditions for the asymptotic properties of the estimator \(\hat{v}\) which will be used in the following sections.

1.4 Simulation Studies

To highlight the advantages of the new M-estimator of WV, in this section we briefly present the results of some simulation studies based on different Gaussian time series and spatial models and different contamination settings. The simulated time series are of length \(T = 1,000\) while the spatial data is of size \(N = 30 \times 30 = 900\). The proposed estimator of WV \(\hat{v}\) based on Tukey’s \(\psi\)-function (RWV) is compared to:

- \(\tilde{v}\): the WV estimator from Percival (1995) for time series and Mondal and Percival (2012a) for spatial data (CL);
- \(\check{v}\): the robust WV estimator proposed by Mondal and Percival (2012b) (MP).

For the proposed M-estimator we choose a tuning constant \(c = 4.97\) which delivers 60% efficiency with respect to the standard estimator under the Gaussian assumption. The level of efficiency for \(\check{v}\) therefore aims for a higher degree of robustness which approaches the degree provided by the median-type estimator used for \(\tilde{v}\) in the simulations in Mondal and Percival (2012b) hence allowing for a fairer comparison between their estimator and the one proposed in this chapter. The performance of the estimators is measured by a relative and robust version of the root mean squared error (RMSE*) defined as follows

\[
\text{RMSE}^* = \sqrt{\text{med} \left( \frac{\bar{v}_i - \nu_{i,0}}{\nu_{i,0}} \right)^2 + \text{mad} \left( \frac{\bar{v}_i}{\nu_{i,0}} \right)^2}
\]

where, for this section, \(\bar{v}_i\) represents the \(i^{th}\) element of a generic vector of estimated WV and \(\nu_{i,0}\) is the true WV implied by the model.

To compare the above estimators, we investigate their performance on the first 4 scales of WV \((j = 1, \ldots, 4)\) in the time series setting and, considering that the spatial models are isotropic, on some elements of the first two columns of the lower diagonal matrix of WV (i.e. \(j = (1,1), (1,2), (1,3), (2,2), (3,2)\)). The reason for this is that contamination becomes less important at the last scales of decomposition where the length of the filters reduce the effect of outliers by downweighting them with many more uncontaminated observations. In fact, the standard estimator is comparable to the robust estimators at the last scales and this is a phenomenon which is also seen in Mondal and Percival (2012b).

Concerning the contamination settings, different portions of contamination \(\epsilon\) were used as well as different types of contamination processes. We use \(\sigma^2\) to denote the variance of the zero-mean contaminating observations which were added to the original observations while \(\mu_{\epsilon,i}\) denotes the \(i^{th}\) mean-shift in the level when using level-shift contamination. With these definitions, the time series models used for the simulations and their relative contamination settings are as follows:
• **AR(1)**: a zero-mean first-order autoregressive model with parameter vector $[\rho_1 \ \upsilon^2]^T = [0.9 \ \upsilon]^T$, scale-based contamination at scale $j = 3$, $\epsilon = 0.01$ and $\sigma^2_\epsilon = 100$;

• **AR(2)**: a zero-mean second-order autoregressive model with parameter vector $[\rho_1 \ \rho_2 \ \upsilon^2]^T = [0.5 \ -0.3 \ 1]^T$, isolated outliers, $\epsilon = 0.05$ and $\sigma^2_\epsilon = 9$;

• **ARMA(1,2)**: a zero-mean autoregressive-moving average model with parameter vector $[\rho \ \upsilon_1 \ \upsilon_2]^T = [0.5 \ -0.1 \ 0.5 \ 1]^T$, and level-shift contamination with $\epsilon = 0.05$, $\mu_{\epsilon_1} = 5$ and $\mu_{\epsilon_2} = -3$;

• **ARMA(3,1)**: a zero-mean autoregressive-moving average model with parameter vector $[\rho_1 \ \rho_2 \ \rho_3 \ \upsilon_1 \ \upsilon_2]^T = [0.7 \ 0.3 \ -0.2 \ 0.5 \ 2]^T$, patchy outliers, $\epsilon = 0.01$ and $\sigma^2_\epsilon = 100$;

• **SSM**: a state-space model ($X_t$) interpreted as a composite (latent) process in certain engineering applications. This model is defined as

$$Y^{\text{(i)}}_t = \rho^{\text{(i)}}_t Y^{\text{(i)}}_{t-1} + W^{\text{(i)}}_t$$

$$W^{\text{(i)}}_t \sim \mathcal{N}(0, \upsilon^{2\text{(i)}}_t)$$

$$X_t = \sum_{i=1}^{2} Y^{\text{(i)}}_t + Z_t,$$

$$Z_t \sim \mathcal{N}(0, \sigma^2)$$

with parameter vector

$$[\rho^{\text{(1)}}_1 \ \upsilon^{2\text{(1)}}_1 \ \rho^{\text{(2)}}_2 \ \upsilon^{2\text{(2)}}_2 \ \sigma^2]^T = [0.99 \ 0.1 \ 0.6 \ 2 \ 3]^T,$$

isolated outliers, $\epsilon = 0.05$ and $\sigma^2_\epsilon = 9$.

Finally, the spatial models used in the simulations and their relative contamination settings are as follows:

• **Exp(1)**: a zero-mean Exponential model with parameter vector $[\phi \ \sigma^2]^T = [2 \ 1]^T$ and level-shift contamination with $\epsilon = 0.05$, $\mu_{\epsilon_1} = 5$ and $\mu_{\epsilon_2} = -3$;

• **Exp(2)**: a sum of two zero-mean Exponential models with parameter vector $[\phi_1 \ \sigma^2_1 \ \phi_2 \ \sigma^2_2]^T = [2 \ 1 \ 1.5 \ 1]^T$, isolated outliers, $\epsilon = 0.01$ and $\sigma^2_\epsilon = 100$;

• **Gauss(1)**: a zero-mean Gaussian model with parameter vector $[\phi \ \sigma^2]^T = [2 \ 1]^T$, patchy outliers, $\epsilon = 0.01$ and $\sigma^2_\epsilon = 100$;

• **Gauss(2)**: a sum of two zero-mean Gaussian models with parameter vector $[\phi_1 \ \sigma^2_1 \ \phi_2 \ \sigma^2_2]^T = [2 \ 1 \ 1.5 \ 1]^T$, isolated outliers, $\epsilon = 0.05$ and $\sigma^2_\epsilon = 9$.

The results of the simulations for these models can be seen in Figures 1.2 and 1.3 where the RMSE* is represented for each scale and the same estimator is connected through a line across scales as a visual support. Figure 1.2 shows the RMSE* for the estimators based on the time series models while Figure 1.3 shows it for the spatial models. Within each figure, the top row shows the RMSE* in the uncontaminated setting while the bottom row shows it for the contaminated setting.
The simulations highlight how the estimators all perform better at the first scales and, as expected, gradually become less precise as the scale increases given the lower number of wavelet coefficients. More specifically, however, it shows how the proposed WV estimator RWV is always the best alternative to the classical estimator CL in an uncontaminated setting whereas it is also the best estimator overall when the observed process is contaminated. Indeed, although the MP estimator was proposed for scale-based contamination as for the AR(1) simulation, the proposed estimator RWV generally performs better compared to the MP estimator no matter the contamination setting. In addition, the loss of efficiency in the uncontaminated settings with respect to the standard estimator is often visibly small, confirming that the proposed approach provides a generally reliable robust estimator of WV.
1.4. Simulation Studies

Uncontaminated Setting

Figure 1.2: RMSE* of the three WV estimators for the time series models.

Contaminated Setting

Figure 1.3: RMSE* of the three WV estimators for the spatial models.
1.5 Applications

In this section we briefly analyse three datasets, two of which were already analysed in previous papers dealing with the WV. The first is the surface albedo data which was used as an example in Serroukh et al. (2000) where a one-dimensional WV analysis is employed, while the second one is taken from Mondal and Percival (2012a) and consists in a two-dimensional random field WV analysis for the purposes of interpreting and modelling cloud images. The third dataset represents a digital elevation model where the WV is used to measure the local roughness of this model which is made available through the raster package in the R statistical software. In all cases, the following paragraphs highlight how a robust approach using the proposed estimator $\hat{\nu}$ is useful to support the overall analysis or reconsider the scientific interpretation of the outputs.

1.5.1 Surface Albedo

Here we revisit the data which was analyzed in Serroukh et al. (2000) as an example of how the WV can be important to determine the quality of a satellite instrument called the “Advanced Very High Resolution Radiometer” (AVHHR) in detecting variability in the ice. In this specific example the data is taken from the measurement of the proportion of incident light reflected (i.e. albedo) of the spring pack ice in the Beaufort Sea and is shown in Figure 1.4 along with the histograms of the first and second differences of the data.

As highlighted in Serroukh et al. (2000) as well, it can be seen from Figure 1.4 that there appear to be isolated spikes of low brightness in the data which, according to the authors, can correspond to irregular cracks in the ice or even to surfaces of water. Due to these isolated spikes the authors assume a non-Gaussian process, although they underline that otherwise the series looks rather homogeneous and that the wavelet coefficients appear to be stationary with the exception of some isolated bursts. Indeed, the histograms for the first two backward differences of the data can appear to have a roughly Gaussian distribution if not for the isolated bursts which considerably widen the tails of the distribution. For this reason, we decide to perform a robust WV analysis based on the proposed estimator $\hat{\nu}$ to understand the impact of the isolated bursts on the standard estimator if we assume the wavelet coefficients to follow a Gaussian process. Figure 1.5 compares the standard estimator of WV with the robust estimator $\hat{\nu}$ for scales $\gamma = 1, \ldots, 10$ (which correspond to distances in metres) along with their respective confidence intervals.

It can be observed how the two estimates of WV differ significantly until the 8th scale where the standard estimator varies over these scales while the robust estimator remains roughly constant. This agrees with the interpretation given in Serroukh et al. (2000) where the non-Gaussian confidence intervals were used to understand if any conclusions could be drawn on the quality of the AVHHR. In fact, referring to the non-Gaussian confidence intervals in their paper, the authors state that if the true WV were close to the minimum of these confidence intervals up to 800m (6th scale) and close to the maximum of these confidence intervals between 1,600m and 3,200m (7th and 8th scales), the AVHHR “would resolve much of the significant variation” thereby indicating that the device is of good quality. The robust WV closely follows this interpretation and therefore appears to be a good measure of quality of this device in the presence of isolated bursts in brightness which are frequent in these applications.
Figure 1.4: Top: Surface albedo data. Bottom: Histogram of the first backward difference of the surface albedo data (left); Histogram of the second backward difference of the surface albedo data (right).

Figure 1.5: Log-log plot of the estimated standard WV and robust WV on the surface albedo data.
1.5.2 Cloud Images

The cloud data has already been analysed using the standard and robust estimators of WV respectively in Mondal and Percival (2012a) and Mondal and Percival (2012b). This data consists in four cloud regions taken from a larger temperature image made available by Geo-stationary Operational Environmental Satellites (GOES) Imagery. The study of cloud variability is paramount to understand the cloud-radiation interaction since this can generate three-dimensional radiative transfer effects which need to be taken into account for the modelling of global and regional climates (see, for example Davis et al., 2002). The image under study represents the stratocumulus clouds in the Chilean coast region which play a major role in the variations of temperatures and seasonal cycles in the east Pacific ocean since they reflect the light from the sun. In Mondal and Percival (2012a) they state that the four regions which were analyzed were picked by some atmospheric scientists who identified them as being of interest for the purpose of understanding the characteristics of these cloud formations whose dynamics are considered as being complex in this specific region (see Mondal and Percival, 2012a, for the details on these images). These studies are of interest also since the region is subject to atmospheric aerosol coming from industrial activities as well as Pockets of Open-Cells (POC) which are mainly responsible for the precipitations in the region.

Let us therefore reproduce the analysis made in Mondal and Percival (2012b) using the robust estimator \( \hat{\nu} \) proposed in this chapter. The results are given in Figure 1.6 where we find the heat-maps of the cloud image itself (top row), the standard WV (top-middle row), the robust WV (bottom-middle row) and finally of the absolute difference between the standard and robust WV (bottom row). The heat-maps of the standard WV roughly correspond to the figures presented in Mondal and Percival (2012a) with some minor differences simply due to the different color-codes for representing the levels of the WV. It can be observed how the first two regions have the same WV pattern both in the standard and robust case, as confirmed by the heat-maps of the absolute difference which are mostly white. This pattern is similar also for the third region in the standard case but is considerably different in the robust case where a high variability is detected mainly in the lower-left region instead of the upper-right and a slightly different variability pattern is detected also in the fourth region. If one considers what these four regions represent, this difference between standard and robust WV appears to be reasonable. Indeed, the first two regions represent “stable” regions, in the sense that the first region represents an already formed POC while the second one represents a uniform formation of stratocumulus clouds. For these two regions therefore, one would expect to well measure the cloud variability for these specific type of formations. However, the last two regions are characterized by “unstable” settings where the third region represents a formation of broken clouds while the fourth represents a POC that is in the process of being formed. These “instabilities” appear therefore to have an impact on the standard WV, assuming that the wavelet coefficients are roughly Gaussian, and although the data was preprocessed using a median filter, it would appear that a robust analysis should nevertheless be taken into consideration. The result of these analyses confirm the conclusions made in Mondal and Percival (2012b) and would consequently have an important impact on the choice of the climate models and the way cloud-radiation interaction is interpreted in these regions (see Chapter 3).
Figure 1.6: Heat-images of the raw cloud data for the four regions (top row), of the standard WV $\tilde{\nu}$ (top-middle row), of the robust WV $\hat{\nu}$ (bottom-middle row) and of their absolute difference (bottom row). The lighter shades indicate low values and darker shades indicate high values. The colour-coding is relative to each region (i.e. the WV has the same coding within the same region but not necessarily between regions).
1.5.3 Digital Elevation Models

Digital Elevation Models (DEM) are digital maps of terrain elevation data and are increasingly being obtained through remote sensing techniques where, for example, radar satellite-based systems are used to generate digital images of surfaces. There is a considerable focus on the quality of DEMs (see for example, Wechsler, 2007; Weng, 2002) since this determines how well these models capture the morphological details of a landscape and to what degree the radar measurements are well approximating the true features of surfaces. The WV is one of the main quantities used to assess the local roughness of the DEM which, if detected, can entail quality improvement techniques such as interpolation thereby improving the geological understanding of the surface (see for example Datcu et al., 1997; Lloyd, 2010). In these cases, an analysis of the diagonal elements of the WV matrix issued from a two-dimensional image decomposition can be extremely useful to understand how well the images are being collected and if some techniques must be used to improve the quality of the DEM. For this purpose, here we carry out a WV analysis on parts of a DEM collected over the San Joaquin field site located in California (USA) and made available through the raster package in the R statistical software. Figure 1.7 shows the log-log plots of the diagonal standard and robust WV taken over specific 1,000m² areas of the DEM along with their confidence intervals. The vast majority of the WV plots show the same pattern and no significant difference between the standard and robust WV, as shown in the left plot in Figure 1.7. However, there are some areas of the DEM which appear to suffer from some specific roughness as highlighted by the right plot of Figure 1.7. Indeed, in the latter the standard WV follows the regular pattern but there appears to be a significant difference with the robust WV at the first three diagonal scales indicating that the pattern for some regions is not the same. This would require to focus on these regions and apply some quality improvement techniques, such as interpolation, in order to make the mapping more reliable.
1.6 Conclusion

We have presented a new general framework for the (robust) estimation of the WV in multidimensional random fields. This framework includes and considerably extends the theoretical and asymptotic properties of the standard estimators of WV providing the conditions for the joint properties of this estimator to hold. The latter conditions are especially important for the growing use of the WV for statistical inference, from Portmanteau tests to isotropy tests and parametric estimation of time series and spatial models. Moreover, the M-estimator on which this framework is based can be made robust by choosing a bounded score function ($\psi$-function) allowing to perform WV analysis when the data is contaminated and improving over alternative robust estimators both in terms of clear asymptotic properties as in terms of finite sample performance. The simulation studies and applied examples highlighted how this new framework is statistically sound and is extremely useful for practical purposes when, for example, the data is affected by contamination.

Figure 1.7: Left: Log-log plot of the standard and robust WV over a 1,000m$^2$ area with coordinates of central point: [4050, 4050]. Right: Log-log plot of the standard and robust WV over a 1,000m$^2$ area with coordinates of central point: [450, 450].
Chapter 2

On the Identifiability of Latent Models for Dependent Data

The condition of parameter identifiability is essential for the consistency of all estimators and is often challenging to prove. As a consequence, this condition is often assumed for simplicity although this may not be straightforward to assume for a variety of model settings. In this chapter we deal with a particular class of models that we refer to as “latent” models which can be defined as models made by the sum of underlying models, such as a variety of linear state-space models for time series. These models are of great importance in many fields, from ecology to engineering, and in this chapter we prove the identifiability of a wide class of (second-order stationary) latent time series and spatial models and discuss what this implies for some extremum estimators, thereby reducing the conditions for their consistency to some very basic regularity conditions. Finally, a specific focus is given to the Generalized Method of Wavelet Moments estimator which is also able to estimate intrinsically second-order stationary models.

2.1 Introduction

The estimation of parametric models heavily relies on the condition that the model parameters are identifiable based on a specific estimation method. Supposing we have a parametric model denoted as \( F_{\theta} \), where \( \theta \) represents the \( p \)-dimensional parameter vector of interest, this condition implies that there is a unique maximizer (or minimizer) to the objective function which defines the estimator and that this unique solution is the true parameter value \( \theta_0 \). More formally, let us consider an extremum estimator (see e.g. Newey and McFadden, 1994) defined as follows

\[
\hat{\theta} = \arg\max_{\theta \in \Theta} Q_n(\theta),
\]

where \( \Theta \) is the compact set of possible parameter values and \( Q_n(\theta) \) is an objective function based on a sample size \( n \) which we assume converges uniformly in probability to \( Q(\theta) \). Based on this, we can define identifiability as being the case where, for all parameter vectors \( \theta \neq \theta_0 \), we have that

\[
\sup_{\theta \in \Theta \setminus \mathcal{N}} Q(\theta) < Q(\theta_0),
\]

where \( \mathcal{N} \) is any open subset of \( \Theta \) containing \( \theta_0 \). Consistency and asymptotic normality of all estimators depend on this condition which is often assumed for simplicity or proven
on a case-by-case basis. The importance of this property was for example emphasized in Newey and McFadden (1994) where they state that, although difficult to prove in general, “it is important to check that it is not a vacuous assumption whenever possible, by showing identification in some special cases”.

In this chapter we prove and discuss the identifiability of a specific class of second-order stationary and regularly spaced time series and spatial models and refer to this class as “latent” models that consist in the sum of different underlying models which cannot directly be observed. These models are greatly used if we consider a broad class of linear state-space and structural time series models (see e.g. Durbin and Koopman, 2012) where the observed data can be explained by a series of components that are allowed to vary over time and have a direct interpretation. For example, in a wide variety of scientific domains many phenomena are explained by unobserved components and are also measured using specific devices which are unable to perfectly recover the values of the phenomenon of interest. A simple example of this is the case where a physical phenomenon, such as the behaviour of a chemical or biological substance, is measured in time but the measuring device is characterized by a stochastic error process which can also vary over time and/or location. The resulting measurement is therefore a combination of the realizations of a certain phenomenon plus a measurement error which the scientist would need to recover in order to have correct interpretations (see e.g. Buonaccorsi, 2010). Other examples where the interest lies in the unobserved components can be found in economics, ecology or biology where the presence of coexisting stochastic processes for different factors is quite common (see e.g. Cedersund and Roll, 2009; Harvey and Koopman, 1993; Z. Zhang et al., 2010).

Despite being frequently used in many contexts, the identifiability of these models is generally assumed, even though this assumption may appear very strong for a variety of latent models. A class of models for which identifiability results do exist are the AutoRegressive Moving-Average (ARMA) models that, as emphasized by Granger and Morris (1976), can be obtained from a sum of latent autoregressive and white noise processes. However, it is unclear if it is always possible to recover the parameters of the latent models through the parameters of the corresponding ARMA as emphasized for example in Harvey and Koopman (1993). Moreover, if this were to be the case, these results could not be used when considering latent models composed by underlying models that are different from those that usually deliver an ARMA model. Therefore, although being widely used in practice, the question of whether these models are generally identifiable remains largely open thereby casting doubts on the consistency of the estimators which are used to recover their parameters.

Considering the above motivations, this chapter proves the identifiability of a class of latent models for dependent data settings through their covariance function which is essential for any estimator to obtain the parameters of these models. Different latent models for time series as well as spatial data are investigated, consequently allowing to use this flexible class of models where a sum of simple spatial models can deliver a class of models in some way akin to ARMA models in the time series setting. Based on these findings, this chapter also discusses what these imply for some extremum estimators, namely the Maximum Likelihood Estimators (MLE), Generalized Method of Moments estimators (GMM) as well as the recently proposed Generalized Method of Wavelet Moments (GMWM) estimators (see Guerrier et al., 2013b). A specific discussion is developed for the latter class of estimators which can estimate models that include intrinsically stationary processes (i.e. non-stationary processes with stationary backward differences) such as drifts and
random walks. The structure of the chapter is therefore as follows. In Section 2.2 we present several results related to the identifiability of latent time series and spatial models while in Section 2.3 we discuss what these results imply for the mentioned extremum estimators, with a particular focus on GMWM estimators and a set of open-problems for future research, as well as a numerical example to support these findings. Finally, Section 5.5 concludes.

### 2.2 Injection of Covariance Models

The estimation of parameters of second-order stationary models for dependent data relies (almost) entirely on the covariance structure of these models. For this reason, in this section we give some results related to the identifiability for different sets of latent time series and spatial models through their respective covariance functions. To do so, let us first define the basic time series models considered in this work as \((X^{(j)}_t)\) with \(t = 1 \ldots, T\) \((T \in \mathbb{N}_+)\) and \(j\) indicating a specific model. With this notation, we have:

**Process (T1)** *White Noise* (WN) with parameter \(\sigma^2 \in \mathbb{R}^+\). We denote this process as \((X^{(1)}_t)\).

**Process (T2)** *Quantization Noise* (QN) (or rounding error, see e.g. Papoulis, 1991) with parameter \(Q^2 \in \mathbb{R}^+\). We denote this process as \((X^{(2)}_t)\).

**Process (T3)** *Drift* with parameter \(\omega \in \mathbb{R}^+\). We denote this process as \((X^{(3)}_t)\).

**Process (T4)** *Random walk* (RW) with parameter \(\gamma^2 \in \mathbb{R}^+\). We denote this process as \((X^{(4)}_t)\).

**Process (T5)** *Moving Average* (MA(1)) process with parameter \(\varrho \in (-1, +1)\) and \(\varsigma^2 \in \mathbb{R}^+\). We denote this process as \((X^{(5)}_t)\).

**Process (T6)** *Auto-Regressive* (AR(1)) process with parameters \(\rho \in (-1, +1)\) and \(\nu^2 \in \mathbb{R}^+\). We denote this process as \((X^{(j)}_t), j = 6, \ldots, K\) with \(K \in \mathbb{N}^+, 6 \leq K < \infty\).

Processes (T1), (T2), (T5) and (T6) are stationary models based on their parameter definitions. Process (T2) is particularly useful in the field of engineering where measurements are often rounded therefore introducing an error process in the procedure. In general terms, this process can be represented as a linear combination of differences of standard uniform variables where \(Q^2\) plays the role of scaling factor. Processes (T3) and (T4) are typically non-stationary processes, where process (T3) is a non-random linear function with slope \(\omega\), and will be discussed in Section 4.4.1 since their covariance functions are not defined. Having defined the time series models of interest for this chapter, let us now define the spatial models. With \(i \in \mathcal{I} \subset \mathbb{N}_+^2\) indicating the spatial coordinates of the observations, we have:
(S1) **Exponential model** with parameters $\phi \in \mathbb{R}^+$ and $\sigma^2 \in \mathbb{R}^+$. We denote this process as $(Y_{i}^{(k)})$, $k = 1, \ldots, N_1$ with $N_1 \in \mathbb{N}^+$, $1 \leq N_1 < \infty$.

(S2) **Gaussian model** with parameters $\phi \in \mathbb{R}^+$ and $\sigma^2 \in \mathbb{R}^+$. We denote this process as $(Y_{i}^{(k)})$, $k = 1, \ldots, N_2$ with $N_2 \in \mathbb{N}^+$, $1 \leq N_2 < \infty$.

These two spatial models are frequently used to approximate and describe many spatial processes. Moreover, a latent model made by the sum of either $N_1$ (S1) processes or $N_2$ (S2) processes represents a flexible solution to describe a more complex spatial process as mentioned in the introduction.

Based on the above definitions, let us now define the latent models for time series as $W_t = \sum_{j=1}^{K} X_{t}^{(j)}$ and for spatial processes as $V_i = \sum_{j=1}^{N} Y_{i}^{(j)}$, where $(V_i)$ can either be the sum of $N_1$ (S1) processes or of $N_2$ (S2) processes. Considering these models, let us denote their covariance function as $\varphi_h(\theta)$, where $h$ is the lag/distance between observations and hence $\varphi(\theta) = [\varphi_h(\theta)]_{h=1,\ldots,H}$. By finally denoting $A(\theta) = \partial/\partial \theta \varphi(\theta)$, let us set the following conditions:

(C1) All processes composing a latent model are mutually independent.

(C2) $A(\theta)$ exists and is continuous $\forall \theta \in \Theta$.

Condition (C1) allows us to have a more tractable and general problem for the identifiability of the parameters of a latent model. Indeed, introducing dependence between the processes entails an approach to proving identifiability which is necessarily on a case-by-case basis since there can be different combinations of dependence between processes and their parametric dependence needs to be specified as well. Nevertheless, the results in this section give a sound basis to obtain further identifiability results also in these cases.

As for Condition (C2), this is needed to develop an expansion of the covariance function $\varphi(\theta)$ in order to prove that this function is injective. In fact, an approach that simultaneously verifies the usual identifiability conditions is to understand if the Jacobian matrix $\partial/\partial \theta \varphi(\theta)$ is of full column rank as a consequence of the following MacLaurin expansion

\[
\varphi(\theta) = \varphi(\theta_0) + A(\theta^*) (\theta - \theta_0) \tag{2.2.1}
\]

where $\theta_0$ and $\theta_1$ are two parameter vectors and $\|\theta^* - \theta_0\| \leq \|\theta_1 - \theta_0\|$. Indeed, if $\theta_0 = \theta_1$, then we automatically have that $\varphi(\theta_0) = \varphi(\theta_1)$ but if we have $\theta_0 \neq \theta_1$ then, if the matrix $A(\theta)$ is full column rank, it means that only the vector $b = 0$ can make $\varphi(\theta_0) = \varphi(\theta_1)$ implying that the only situation where this is possible is when $\theta_0 = \theta_1$.

Knowing this, let us now define the classes of latent models studied in this section:

**Model 1** $W_t = X_{t}^{(1)} + X_{t}^{(2)} + \sum_{i=6}^{G} X_{i}^{(i)}$.

**Model 2** $W_t = X_{t}^{(5)} + \sum_{i=6}^{G} X_{i}^{(i)}$.

**Model 3** $V_i = \sum_{j=1}^{N} Y_{i}^{(j)}$, with $Y_{i}^{(j)}$ representing either model (S1) or model (S2).

The above processes are all second-order stationary ones, with Model 1 and Model 2 representing latent time series and Model 3 representing latent spatial models. Since
the covariance function is well defined for second-order stationary processes, we start by considering the class of models included within Model 1 which represents the sum of a (T1) process, a (T2) process and K (T6) processes (with \( K = G - 5 < \infty \)). Models of this class are commonly used in various disciplines going from economics to engineering for sensor calibration. Therefore, denoting \( \rho_i \) as the autoregressive parameter of the \( i^{th} \) (T6) process, the following theorem considers the properties of its autocovariance function with respect to the parameters of Model 1.

**Theorem 2.2.1.** Under Conditions (C1) and (C2), and assuming \( \rho_i \neq 0 \), \( \forall i \) and \( \rho_i < \rho_j \), with \( 1 \leq i < j \leq K \), we have that the covariance function of Model 1 is injective.

The proof of this theorem can be found in Appendix B.1.1. Theorem 2.2.1 therefore states that there is a unique mapping of the parameters \( \theta \) to the covariance function \( \varphi(\theta) \) for a very general class of second-order stationary latent models. It must be noted that process (T5) is not included in this class due to the fact that the structure of matrix \( A(\theta) \) is not clearly full rank if this process is included in a latent model with processes (T1) and/or (T2). Based on this, we can give the following corollary to Theorem 2.2.1.

**Corollary 2.2.1.** Under the conditions of Theorem 2.2.1, we have that the covariance function of Model 2 is injective.

The proof of this corollary is a direct consequence of the proof of Theorem 2.2.1 where the columns of matrix \( A(\theta) \) containing the derivatives with respect to the parameters of processes (T1) and (T2) are replaced by those of the parameters of process (T5).

**Remark 2.2.1.** From the results of Granger and Morris, 1976 we know that the sum of (T1) and (T6) models delivers different kinds of ARMA models. However, as emphasized in the introduction, it is not clear that all combinations of (T1) and (T6) models enjoy a unique mapping to a specific ARMA model as highlighted by Harvey and Koopman (1993). A simple example where this mapping is indeed satisfied is given by the sum of two (T6) processes which has a unique mapping to an ARMA(2,1) model (see Appendix B.1.2 for details).

The results of Theorem 2.2.1 and Corollary 2.2.1 state that the covariance function of a wide class of second-order stationary latent time series models is injective. Let us now consider the latent processes \( (V_i) \) based on the sum of spatial models (S1) and (S2) and give the following theorem.

**Theorem 2.2.2.** Under Condition (C1) and (C2), and assuming \( \phi_i < \phi_k \), \( \forall i < k \), , with \( 1 \leq i < k \leq N \), the covariance function of Model 3 is injective.

The result of this theorem delivers a class of latent spatial models which in some way presents some similarities with ARMA models for time series. Indeed, as mentioned earlier, a sum of (T6) models delivers an ARMA model based on the results of Granger and Morris (1976) and it is possible to suppose that a sum of Exponential or Gaussian covariance models can deliver something similar in the spatial case. Hence, Theorem 2.2.2 states that this new class of spatial models is identifiable based on their covariance function.
2.3 Consistency of Extremum Estimators for Latent Models

The results in Section 2.2 are helpful to considerably reduce the conditions for consistency of various estimators with respect to a wide class of second-order stationary models. In this section we consider some estimators that are part of the class of extremum estimators defined in the introduction and state the conditions for their consistency. The asymptotic distribution of these estimators is not discussed since it would require additional model-specific conditions but, in any case, it often relies strongly on the results of consistency given below. Having emphasized this, let us now list the other conditions (aside from identifiability) necessary to achieve consistency of extremum estimators.

(C4) $\Theta$ is compact.

(C5) $Q(\theta)$ is continuous.

(C6) $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0$.

These conditions are based on Theorem 2.1 in Newey and McFadden (1994) and are standard for the consistency of extremum estimators. Indeed, Conditions (C4) and (C5) are statements which fall under the usual regularity assumptions while Condition (C6) requires the sample objective function $Q_n(\theta)$ to converge uniformly in probability to the true function $Q(\theta)$ which is not necessarily strong if adding minor assumptions to the stationary processes considered in this chapter. With this in mind, the following paragraphs discuss these conditions for different estimators in the light of the results of Section 2.2.

2.3.1 MLE and GMM Estimators

The MLE and GMM estimators are very popular members of the class of extremum estimators and, considering these two particular estimators, let us denote $f(z|\theta)$ as the density function of the latent time series or spatial model while $g(z|\theta)$ and $W$ denote the moment conditions and the weighting matrix for the GMM estimators respectively. Considering this, let us give the following conditions:

(C7) $F_{\theta}$ is uniquely defined by its covariance structure $\varphi(\theta)$ and $\mathbb{E}[|\ln(f(z|\theta))|] < \infty, \forall \theta$.

(C8) The moment condition $g(z|\theta)$ depends on $\theta$ uniquely through the covariance function $\varphi(\theta)$ and the weighting matrix $W$ is positive semi-definite.

These conditions are based on Lemmas 2.2 and 2.3 in Newey and McFadden (1994) and list the additional properties needed for $Q(\theta)$ to be uniquely maximised at the true parameter value (i.e. $\theta = \theta_0$) based on the results in Section 2.2. This allows to give the following lemma.

Lemma 2.3.1. Under the conditions of Theorem 2.2.1, Corollary 2.2.1 and Theorem 2.2.2, the parameters of Model 1 to Model 3 are identifiable for the MLE under Condition (C7) and for the GMM under Condition (C8).
Condition (C7) is stronger to assume compared to Condition (C8) since, with a few exceptions, the former would require $f(z|\theta)$ to be the density of a Gaussian distribution. On the other hand, Condition (C8) is not necessarily too strong since GMM estimators for dependent processes are usually based on the covariance function $\varphi(\theta)$, assuming the choice of a positive semi-definite matrix $W$. These conditions are necessary to ensure that one of the basic (and typically difficult to derive) conditions for consistency of an extremum estimator holds: identifiability. Under Conditions (C7) and (C8), where the latter is easily verifiable, Section 2.2 has shown that the covariance function is injective for a wide class of second-order stationary latent time series and spatial models which therefore removes this challenging condition for them. Based on these conditions, the following corollary states the consistency of the MLE and GMM estimators which, for simplicity, we both denote as $\hat{\theta}$.

**Corollary 2.3.1.** Under the conditions of Lemma 2.3.1 and Conditions (C4) to (C6), we have that

$$\hat{\theta} \overset{p}{\to} \theta.$$

Having stated the consistency of two popular extremum estimators (i.e. MLE and GMM), the next section discusses the properties of the GMWM for latent models which include the non-stationary processes (T3) and (T4).

### 2.3.2 GMWM Estimators

The GMWM is a minimum-distance extremum estimator where the reduced form (or auxiliary) parameters are represented by the Wavelet Variance (WV) which represents weighted averages of the Spectral Density Function (SDF) over octave bands. More specifically, the GMWM is defined as follows

$$\hat{\theta} = \arg\min_{\hat{\nu} \in \Theta} (\hat{\nu} - \nu(\theta))^T \Omega (\hat{\nu} - \nu(\theta))$$

(2.3.1)

where $\hat{\nu} = [\hat{\nu}_j^2]_{j=1,...,J}$ represents the vector of estimated Haar WV for the $J$ scales of wavelet decomposition (see Percival, 1995), $\nu(\theta) = [\nu_j^2(\theta)]_{j=1,...,J}$ represents the vector of WV implied by the model of interest and $\Omega$ is a positive definite weighting matrix chosen in a suitable manner (see Guerrier et al., 2013b).

As can be observed from the form of the objective function of the GMWM, the latter depends on $\theta$ uniquely through the theoretical form of the WV $\nu(\theta)$ (i.e. the WV implied by the model $F_\theta$). Based again on Newey and McFadden (1994), we have identifiability of the parameters $\theta$ through the GMWM if:

1. $\nu(\theta) = \nu(\theta_0)$ if and only if $\theta = \theta_0$; and
2. $\Omega$ is positive definite.

Therefore, we would need to focus on whether the WV $\nu(\theta)$ is injective. Let us first focus on the latent models made by the sum of stationary processes. Using the same logic as in Section 2.2, we would need to prove that the Jacobian matrix $A(\theta) = \partial \nu(\theta)/\partial \theta$ is full column rank, keeping in mind that Condition (C2) should hold also in this case. However, this direct approach appears difficult to prove when considering the sum of (T6) processes and an approach by steps is needed. This approach discusses the unique mapping of the parameters $\theta$ to the SDF $S_{\theta(f)}$ and then from there to the WV $\nu(\theta)$. However, we are forced to assume this last step and therefore we give the following condition.
(C9) There is a unique mapping between the SDF $S_{\theta}(f)$ and the WV $\nu(\theta)$.

This condition was discussed in the continuous case for the Allan Variance (AV) by Greenhall (1998). The AV has a direct relationship with the Haar WV (i.e. $AV \equiv 2WV$) and can therefore be considered similar to the WV in that it is derived from another kind of averaging of the SDF. In Greenhall (1998) some very specific cases are discussed in which this mapping is not necessarily unique but, as the author himself claims, these are particular cases that hardly exist in reality. More specifically, if we define the signed SDF as $\Phi(f) = S_{\theta_0}(f) - S_{\theta_1}(f)$, a condition that the author gives for Condition (C9) not to be satisfied is given below.

(C10) The signed SDF $\Phi(f)$ satisfies $\Phi(2f) = \frac{1}{2}\Phi(f)$.

The reason for this condition resides in the definition of the Haar WV which is adapted from Greenhall (1998) as follows

$$\nu^2_j(\theta) = \frac{1}{2} \int_0^{\infty} \sin^4\left(\frac{\tau \pi f}{(\pi f)^2}\right) S_{\theta}(f) df,$$

where $\tau = 2^j$. As a consequence of this definition, there would not be a unique mapping from the SDF to the WV if $\Delta \equiv \nu^2_j(\theta_0) - \nu^2_j(\theta_1) = 0$ for some $\theta_0 \neq \theta_1$. Following the proof in Greenhall (1998), using trigonometric inequalities and discarding constants, this would deliver

$$\Delta = \int_0^{\infty} \frac{\sin^4(\tau \pi f)}{(\tau \pi f)^2} \Phi(f) df = \lim_{n \to -\infty} \int_{2^n}^{\infty} \frac{\sin^2(\tau \pi f) - \frac{1}{4} \sin^2(2\tau \pi f)}{(\tau \pi f)^2} \Phi(f) df.$$

Rewriting the above as a difference of integrals and by change of variable $u = 2f$ in the second term we have

$$\Delta = \lim_{n \to -\infty} \left[ \int_{2^n}^{\infty} \frac{\sin^2(\tau \pi f)}{(\tau \pi f)^2} \Phi(f) df - \frac{1}{2} \int_{2^{n+1}}^{\infty} \frac{\sin^2(\tau \pi u)}{(\tau \pi u)^2} \Phi(f) du \right]$$

where, by finally using Condition (C10), we obtain

$$\Delta = \lim_{n \to -\infty} \int_{2^n}^{2^{n+1}} \frac{\sin^2(\tau \pi f)}{(\tau \pi f)^2} \Phi(f) df = 0.$$

This implies that, in the continuous case, there is not a unique mapping from the SDF to the WV when Condition (C10) is satisfied. The following corollary gives a result regarding this condition.

Corollary 2.3.2. The signed SDF of Model 1 does not satisfy Condition (C10).

The proof of this corollary can be found in Appendix B.1.4. However, even in the discrete case, Condition (C10) gives us the “if” but not the “only if” statement which would allow identifiability. The following conjecture refers to this problem.

Conjecture 2.3.1. There is always a unique mapping between the SDF and the WV if Condition (C10) is not satisfied.
This conjecture, whose proof is left for future research, claims that Condition (C10) is therefore the only condition which must be verified for there to be a unique mapping from the SDF to the WV and consequently concludes the discussion on Condition (C9). The latter therefore does not appear to be a strong condition to assume also considering additional arguments given further on. Having discussed this condition, we now provide the following corollary.

**Corollary 2.3.3.** Under the conditions of Theorem 2.2.1 and Corollary 2.2.1, and under Condition (C9), the WV $\nu(\theta)$ of Model 1 and Model 2 is injective.

The proof of this corollary is a direct consequence of Theorem 2.2.1 by using the argument that a composition of injective functions is itself an injective function. Indeed, Lemmas B.1.1 and B.1.2 in Appendix B.1.5 show that there is a unique mapping between the covariance function and the SDF of these latent models and, assuming Conjecture 2.3.1 is true, we have that there is a unique mapping between the parameter $\theta$ and the WV $\nu(\theta)$. Hence, Condition (C9) allows us to avoid proving the full column rank of the Jacobian matrix $A(\theta)$. Nevertheless, this proof is possible for some classes of latent models given below.

**Model 4** $W_t = \sum_{i=6}^{10} X_t^{(i)}$.

**Model 5** $W_t = \sum_{i=1}^{4} X_t^{(i)}$.

**Model 6** $W_t = \sum_{i=3}^{5} X_t^{(i)}$.

**Model 4** is a stationary latent model since it consists in a sum of $K$ (T6) processes while **Model 5** and **Model 6** are classes of latent models that combine stationary and non-stationary processes. Let us start from **Model 4** for which the determinant of the Jacobian $A(\theta)$ is always positive for $G = 10$ (i.e. the sum of four (T6) processes) thereby implying that the WV $\nu(\theta)$ is injective for this model. Based on this finding, we give the following conjecture.

**Conjecture 2.3.2.** Under the conditions of Theorem 2.2.1, the determinant of the Jacobian $A = \partial/\partial \nu(\theta)$ for Model 4 ($G < \infty$) is given by

$$|A| = \frac{\prod_{i=1}^{K} v_i^2 \prod_{i<j}^{K} (\rho_i - \rho_j)^4}{\prod_{i=1}^{K} (\rho_i^2 - 1)^2} > 0.$$ 

As mentioned earlier, this conjecture is verified for a sum of up to four (T6) processes, implying that the Jacobian is of full column rank. Unfortunately, proof by induction is quite hard in this case and is therefore not considered. However this conjecture, along with the results in Greenhall (1998), gives additional support to the argument stating that Condition (C9) is not a strong one to assume. Finally, using the same approach based on the Jacobian matrix $A(\theta)$, let us give the following lemma.

**Lemma 2.3.2.** Under Conditions (C1) and (C2), the WV $\nu(\theta)$ of Model 5 and Model 6 is injective.

This lemma, whose proof is given in Appendix B.1.6, allows to combine some stationary processes with the non-stationary processes (T3) and (T4) considered in this chapter. Indeed, excluding the (T6) process, the parameters of each stationary process can be identified together with those of the non-stationary ones. With these results, we can finally give the following lemma.
Lemma 2.3.3. The GMWM is able to identify the parameters of

- **Model 1** and **Model 2** under the conditions of Corollary 2.3.3.
- **Model 4** under Condition (C9).
- **Model 5** and **Model 6** under the conditions of Lemma 2.3.2.

This lemma therefore summarizes the results of this section. Indeed, the parameters of **Model 1** and **Model 2** are identifiable for the GMWM if we assume Condition (C9) which is claimed to be true for these latent models in Conjecture 2.3.1. On the other hand, Conjecture 2.3.2 claims that the parameters of **Model 4** are also identifiable for the GMWM while those of **Model 5** and **Model 6** are generally identifiable. Considering these results, the consistency of the GMWM is given in the following lemma.

**Corollary 2.3.4.** For the processes and conditions considered in Corollary 2.3.3, and assuming Conditions (C4) to (C6), we have that

\[ \hat{\theta} \xrightarrow{p} \theta. \]

The GMWM is therefore consistent for different classes of latent models based on the above results. The case of a latent models including a sum of (T6) processes with processes (T3) and (T4) has not been investigated here since the approach used to prove Corollary 2.3.3 is based on the SDF which is not defined for non-stationary processes. However, if processes (T1) to (T5) are only included once in a latent model, the results of this section strongly suggest that the parameters of any latent model made by the combination of the time series models considered in this chapter, excluding a combination of (T1) and/or (T2) with (T5), is identifiable through the WV \( \nu(\theta) \). Corollary 2.3.3 and Lemma 2.3.2 support this idea based on the intuitive argument that the WV of the non-stationary processes (T3) and (T4) increases steadily at the larger scales which cannot be in any way approximated by the stationary processes considered here since their WV decreases at these scales.

The use of the GMWM for the estimation of the parameters of latent spatial models would undergo the same type of conditions and arguments given above for latent time series models. Indeed, a similar statement to that of Corollary 2.3.3 could be envisaged for **Model 3** consisting in the sum of either spatial models (S1) or (S2). However, it is not clear if the same arguments for the mapping of the covariance function to the SDF can be used and directly proving the full-column rank of the Jacobian matrix \( A(\theta) = \partial / \partial \theta \nu(\theta) \) is equally as challenging as for the latent time series models made by a sum of (T6) processes.

### 2.3.3 Numerical Example

To conclude this section, we provide a brief numerical example in which we compare the performance of the GMM and GMWM estimators when estimating the parameters of two latent models on different sample sizes. These two models are defined as follows:

**Latent Model 1** \( W_t = X_t^{(1)} + X_t^{(2)} + \sum_{i=6}^{7} X_t^{(i)} \) with parameter vector \([\rho_1 \, \nu_1^2 \, \rho_2 \, \nu_2^2 \, \sigma^2 \, Q^2]^T = [0.99 \, 1 \, 0.85 \, 15 \, 2 \, 4].\)
Latent Model 2 \( W_t = X_t^{(1)} + X_t^{(2)} + X_t^{(5)} + X_t^{(6)} \) with parameter vector 
\[ \begin{bmatrix} \rho_1 & \sigma_2 & \sigma_2 & Q_2 \end{bmatrix}^T = [0.99 \ 1.085 \ 15.2 \ 4]. \]

Latent Model 1 is made by the sum of two (T6) processes with a (T1) and a (T2) process while Latent Model 2 is the same except that the second autoregressive process is replaced by a (T5) process. From the discussions in Section 2.2, Latent Model 2 could have issues of identifiability since process (T5) is included in a model with processes (T1) and (T2).

These models were estimated 500 times for each sample size going from 100 to 1 million and the Mean Squared Error (MSE) of the estimators was computed for each parameter. The results are given in Figure 2.1 where the logarithmic transform of the MSE for each parameter is given along with its confidence intervals (obtained via bootstrap on the estimated parameters) for growing sample sizes.

Figure 2.1: Logarithm of the MSE for the GMM (red line) and GMWM (green line) on Latent Model 1 (top part of figure) and Latent Model 2 (bottom part of figure). The grey shaded areas represent the confidence intervals of the MSE for each estimator.

For consistent estimators we would expect the logarithmic transform of the MSE to
decrease monotonically as the sample size increases, which is roughly the case for the parameters of Latent Model 1. However, this is not the case for the MSE of Latent Model 2 where it does not have a clear monotonic decrease for the last three parameters that belong to the (T1), (T2) and (T5) processes. Indeed, in this case the GMWM seems to remain roughly constant as the sample size increases while the GMM appears to reach a lower bound for the parameters of processes (T1) and (T2). These results seem therefore to support some of the findings highlighted in the previous sections.

2.4 Conclusion

The results on the identifiability of latent models presented in this chapter are summarized in Figure 2.2 and considerably widen the spectrum of identifiable models in dependent data settings beyond standard ARMA models. As a result, the widespread use of these types of models is supported by the findings of this chapter which also highlighted the latent models which are not identifiable, thereby avoiding an inappropriate estimation procedure. Moreover, we extended the idea of latent models to a group of spatial models for which we also proved that a combination thereof is identifiable, delivering a flexible class of models for spatial data. Finally, we showed how these results considerably reduce the conditions needed for consistency for some extremum estimators, such as the MLE and GMM, while the addition of non-stationary models in this framework and their identifiability was discussed for the GMWM estimators.
Chapter 3

Fast and Robust Parametric Estimation for Time Series and Spatial Models

We present a new framework for robust estimation and inference on second-order stationary time series and random fields. This framework is based on the Generalized Method of Wavelet Moments which uses the wavelet variance to achieve parameter estimation for complex models. Using an M-estimator of the wavelet variance, this method can be made robust therefore allowing to estimate the parameters of a wide range of time series and spatial models when the data suffers from outliers or different forms of contamination. This chapter presents a series of simulation studies as well as a range of applications where this new approach can be considered as a computationally efficient, numerically stable and robust method which performs at least as well as existing methods in bounding the influence of outliers on the estimation procedure.

3.1 Introduction

Parametric inference on random fields has been a widely tackled topic, especially concerning time series and spatial modelling. However, many available methods for this purpose can require a selection of auxiliary parameters (or models) which is not always clear, can be computationally impractical when dealing with larger sample sizes or can be numerically unstable when estimating complex models. Moreover, adding robust inference to this setting is often a daunting task. Indeed, already in the time series field (i.e. one-dimensional), the limited applicability of robust estimation is testified by the lack of available methods in statistical software, even though there is an abundance of literature in this domain. This is also evident when it comes to the robust estimation of spatial models. The reader can find a short literature review of existing methods for the robust estimation of random fields in Appendix C.1.

In this setting, the goal of this chapter is to propose a robust method for the parametric estimation of time series and spatial models which is able to estimate a wide variety of second-order intrinsically stationary processes (i.e. stationary or non-stationary processes with stationary backward differences), is computationally efficient, numerically stable and compares on level, if not better in some cases, with the few already available robust methods. To do so, we first extend the idea of the Generalized Method of Wavelet Moments (GMWM) (see Guerrier et al., 2013b) to the parametric estimation of random fields,
having been initially conceived for time series model inference. The reason for this extension resides in the fact that the GMWM is based on a quantity called the Wavelet Variance (WV) which adequately summarizes the dependence structure “information” in a sample issued from a certain parametric model $F_\theta$, with $\theta$ being the parameter vector of interest. As a minimum distance estimator (MD), the GMWM uses the WV as auxiliary parameters that allow to estimate $\theta$ for a variety of models which can also be relatively complex in nature (e.g. latent models which are the result of the sum of different underlying processes). Due to these characteristics, this method presents a few advantages over traditional estimators such as, for example, the Generalized Method of Moments (GMM) estimators and Maximum Likelihood Estimators (MLE). Indeed, for GMM estimators the selection of good moment conditions or auxiliary parameters becomes an important issue to deal with, especially when the sample size is large (see, for example, Andrews, 1999). On the other hand, when dealing with state-space time series models or simple spatial models, the MLE often relies on the state estimation, inversion of the covariance matrix and/or computation of the distance matrix thereby making their computational feasibility limited even when the sample is moderately large. With these issues in mind, the GMWM overcomes these problematics, paying a reasonable price in terms of statistical efficiency, by using a wavelet decomposition which adequately condenses the information in the sample to a moderate number of auxiliary parameters which are the WV and can consequently estimate also complex models in a computationally efficient manner.

Aside from the above advantages, the reason for considering the GMWM for the robust parametric estimation of time series and spatial models resides in the fact that it can easily be made robust by using a robust estimator of the WV. In fact, Ronchetti and Trojani (2001) and M. G. Genton and Ronchetti (2003) highlight that a bounded auxiliary parameter or moment condition can guarantee the robustness of the resulting parametric estimator and this property was already investigated in the time series setting in Guerrier et al. (2014) where simulation studies hinted that this approach constituted a valid means to bound the influence of contaminated observations in a dependent data scenario. However, in the latter the authors used the robust M-estimator of WV proposed by Mondal and Percival (2012a) that, although bounding the influence of outliers, does not benefit from clear asymptotic properties which would allow for inference when estimating parameters of random fields. For this reason, in this chapter we make use of the M-estimator of WV described in Chapter 1 which does not necessarily require normality of the data, benefits from appropriate asymptotic properties and overall shows better finite sample performance. More specifically, this estimator also allows to determine the desired level of robustness and its consistency and joint asymptotic normality have been proven thereby ensuring that its properties directly transfer to the GMWM estimator.

This chapter is organized in the following manner. Section 3.2 introduces the straightforward extension of the GMWM to the (robust) estimation of random fields and highlights its asymptotic properties based on the results in Chapter 1. A simulation study comparing standard and robust estimators for the parameters of a variety of time series and spatial models is presented in Section 3.3 where the good properties of the Robust GMWM (RGMWM) are shown both in terms of estimation as in terms of computational efficiency and numerical stability. The usefulness of this new method is highlighted in Section 3.4 with a series of practical examples covering some applications in time series analysis as well as in spatial modelling. Finally, Section 3.5 concludes.
3.2 Robust GMWM for Random Fields

The GMWM was initially proposed by Guerrier et al. (2013b) as a means to estimate so-called composite (or latent) time series models where the observed process is the result of a sum of different underlying processes which are intrinsically stationary. This allows to estimate, for example, a class of basic time series models as well as many linear state-space models among which we can find the classic (Seasonal) Autoregressive (Integrated) Moving Average (SARIMA) models (see for example the results of Granger and Morris, 1976). To do so, the GMWM uses the WV as an auxiliary parameter and is defined as

\[ \hat{\theta} = \arg\min_{\theta \in \Theta} (\hat{\nu} - \nu(\theta))^T \Omega (\hat{\nu} - \nu(\theta)) \]  

(3.2.1)

where \( \theta \) represents the vector of parameters defining the stochastic process \( F_\theta \), \( \hat{\nu} = [\hat{\nu}^2_j]_{j=1,\ldots,J} \) represents the vector of estimated WV (see Percival, 1995), \( \nu(\theta) = [\nu^2_j(\theta)]_{j=1,\ldots,J} \) represents the vector of WV implied by the model of interest and \( \Omega \) is a weighting matrix, with elements \( \omega_{i,j} \), \( i, j = 1, \ldots, J \), chosen in a suitable manner (see Guerrier et al., 2013b).

Having defined the GMWM, this section combines the M-estimation framework for the WV proposed in Chapter 1 to obtain the RGMWM estimator for multidimensional second-order intrinsically stationary random fields. From now onwards, the estimator of WV \( \hat{\nu}^2 \) therefore denotes the mentioned M-estimator and, based on this, in the following paragraphs we list and discuss the conditions for the consistency and asymptotic normality of the RGMWM which are generally the same as those for GMM estimators. Denoting \( \nu(\theta) \) as \( \nu \) for simplicity and using \( N \) to represent the sample size, let us therefore start by listing the conditions for the consistency of the RGMWM estimator:

(C1) The set \( \Theta \) is compact.

(C2) The function \( \nu(\theta) \) is continuous in \( \theta \).

(C3) The function \( \nu(\theta) \) is globally identifiable.

(C4) \( \|\hat{\nu} - \nu\| \xrightarrow{p} 0 \).

(C5) \( \Omega \) is positive definite and its largest eigenvalue, denoted as \( \lambda \), is such that \( |\lambda| < \infty \).

(C6) If \( \Omega \) is estimated by \( \hat{\Omega} \), then \( \hat{\omega}_{i,j} = \omega_{i,j} + O_p(1/g(N)) \) where \( g(N) \in \mathbb{R} \) is a function of \( N \) such that \( N^{\alpha}/g(N) = o(1) \).

Condition (C1) is a standard condition which is often used for estimators to be consistent while Condition (C2) is easy to verify and is respected for most intrinsically stationary processes. However, Condition (C3) is an essential one which is often hard to verify. In this case, the identifiability of a wide class of (latent) time series models and some spatial models was shown in Chapter 2 thereby removing Condition (C3) for these cases. As for Condition (C4), this is verified under the conditions presented for the M-estimator of WV given in Chapter 1. Generally speaking, according to whether a bounded or unbounded score function is chosen to obtain \( \hat{\nu} \), these conditions require the WV \( \nu \) to be identifiable and assume that the wavelet coefficients follow a strongly mixing stationary and ergodic process with bounded fourth moments. Moreover, if using a bounded score function, this must be Bouligand differentiable and its spectral density...
must be strictly positive at zero frequency. In addition, it is required that the number of scales of decomposition $J$ grow at a suitable rate compared to the sample size $N$ (see Chapter 1). For multidimensional random fields (e.g. spatial processes), the last condition would require that the sample size goes to infinity along all dimensions at a suitable rate as well. Finally, Conditions (C5) and (C6) concern the choice of the weighting matrix $\Omega$, where the first is easy to verify while the second does not appear strong to assume. Indeed, the GMWM estimator is consistent for any $\Omega$ that is positive definite. However, the most efficient GMWM estimator is the one based on $\Omega = V^{-1}$ where $V$ is the covariance matrix of the estimator of $WV$. Defining $\Psi(W_{k_j}, \nu)$ as the vector score-function of the M-estimator proposed in Chapter 1, where $W_{k_j}$ is the vector of wavelet coefficients at coordinates $k_j$, this covariance matrix is given by

$$V = M^{-1}S_{\psi}(0)M^{-T},$$

where $S_{\psi}(0)$ is the spectral density of $\Psi(W_{k_j}, \nu)$ at zero frequency and

$$M = \mathbb{E} \left[ -\frac{\partial}{\partial \nu} \Psi(W_{k_j}, \nu) \right].$$

We propose to estimate $V$ either via parametric bootstrap or by replacing $\nu$ in the above expressions with $\hat{\nu}$ and estimating $M$ and $S_{\psi}(0)$ via their sample versions (see Iverson and Randles, 1989). With these conditions and denoting $\theta_0$ as the true parameter vector of the random field $(X_k)$, we can now state the consistency of the GMWM estimator $\hat{\theta}$.

**Proposition 3.2.1.** Under conditions (C1) to (C6) we have that

$$\|\hat{\theta} - \theta_0\| \xrightarrow{P} 0.$$

The proof of this proposition can be found in Appendix C.2.1. Having stated the consistency, we can now give the conditions for the asymptotic normality of $\hat{\theta}$:

(C7) $\theta_0$ is an interior point to $\Theta$.

(C8) $H(\theta_0) \equiv \frac{\partial}{\partial \theta} \nu(\theta) \bigg|_{\theta=\theta_0}$ exists and is non-singular.

(C9) $\sqrt{N}s^TV^{-1/2}(\hat{\nu} - \nu) \xrightarrow{D} N(0, 1)$, where $\|s\| = 1$.

Condition (C7) is also a standard condition since normality is commonly proven via expansions based on derivatives which cannot be made if the true parameter is at the bounds of the parameter space $\Theta$. Condition (C8) is also usually assumed since it depends on the specific model $F_\theta$ from which the random field $(X_k)$ is generated and cannot therefore be proven in general. On the other hand, Condition (C9) is also essential for the asymptotic normality of $\hat{\theta}$ and it has been shown that $\hat{\nu}$ verifies this condition under the specific assumptions highlighted in Chapter 1. In the latter, the conditions on mean-square consistency and joint asymptotic normality of the estimator $\hat{\nu}$ are given for the first time, thereby allowing to understand the conditions needed for the asymptotic properties of the RGMWM to hold. More specifically, aside from the conditions needed for the consistency of $\hat{\nu}$ mentioned earlier, the additional conditions needed for (C9) to be verified are that the WV score function is twice Bouligand differentiable and that, for example, $J < [\log_2(\sqrt{N})]$ when using the Haar wavelet filter in a time series setting, where $[x]$ represents the largest integer smaller than $x$. Having discussed these conditions, we can use them to state the following proposition.
Proposition 3.2.2. Under the conditions for Lemma 3.2.1 as well as Conditions (C7) to (C9), the estimator $\hat{\theta}$ has the following asymptotic distribution

$$\sqrt{N} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{D} N \left( 0, BVB^T \right)$$

where $B = H(\theta_0)^{-1}D(\theta_0)^T \Omega$ and $D(\theta_0) = \partial/\partial \theta \nu(\theta)|_{\theta = \theta_0}$.

The results of Chapter 1 are therefore essential to obtain the asymptotic properties of the RGMWM estimator stated in Propositions 3.2.1 and 3.2.2. Moreover, following M. G. Genton and Ronchetti (2003), choosing a bounded $\psi$-function for the M-estimator in Chapter 1 ensures robustness of $\hat{\nu}$ allowing the RGMWM estimator $\hat{\theta}$ to be robust due to the bounded IF of $\hat{\nu}$. This new estimator delivers a few theoretical advantages over existing robust estimators as well as considerable practical advantages as shown in Section 3.3. First of all, it can deal with non-Gaussian processes as emphasized in Chapter 1 and benefits from different parameter identifiability results given in Chapter 2. Moreover, the conditions for joint asymptotic normality are known for multidimensional random fields and can therefore be taken into account for the asymptotic properties of the RGMWM to hold. A main advantage however resides in the fact that the dimension of the auxiliary parameter vector is always reasonable since in general $J \leq \left\lfloor \log_2(N) \right\rfloor - 1$ which allows to make use of all the scales of WV without the need to select specific moments even for extremely large sample sizes. This is not the case, for example, for GMM estimators where moment-selection is an important issue since, according to the model that is being estimated, the choice should fall on all moments (which can be highly impractical) or on moments that are more “informative” than others (Andrews, 1999). The RGMWM on the other hand makes use of the WV which adequately summarizes all the information in the spectral density into a few auxiliary parameters without the need to select specific moments which contain more information.

Considering the different conditions listed above, let us state the following corollary which gives the conditions for the asymptotic properties of a specific RGMWM estimator. This estimator will then be used in the simulations and applications of the following sections.

Corollary 3.2.1. Assuming that $(X_k)$ is a Gaussian random field, that $\nu$ belongs to a compact set and that $S_\psi(0) > 0$, under Conditions (C1), (C3) and Conditions (C6) to (C8), the RGMWM based on the Tukey $\psi$-function is consistent and asymptotically normally distributed considering $J$ as fixed.

The proof simply follows from the results given in Chapter 1 and in this section. Indeed, most of the conditions in Corollary 3.2.1 are parameter- and/or model-specific and therefore have to be assumed in general or proven case-by-case. Moreover, under the Gaussian assumption for $(X_k)$, we know that $\nu$ is identifiable and that Condition (C4) is verified using the Tukey $\psi$-function (see Chapter 1). Having defined this new robust estimator, the following sections investigate its performances over a series of simulations studies and applications for time series and spatial model estimation.

3.3 Simulation Studies

The aim of this section is to show that the RGMWM estimator has a reasonable performance in settings where there is no contamination and has a better performance than the
classical (and possibly robust) alternatives when the data are contaminated. Concerning
the robust alternatives, as explained in further details in the following sections, there is a
lack of implemented robust methods for time series analysis and complete absence for the
estimation of spatial models. For this reason, we only make comparisons with methods
that we were able to obtain good results with. Moreover, we intend to show the numer-
ical stability and computational efficiency of this estimator, especially in the time series
setting with models of higher complexity than an AR(p) model.

To measure the statistical performance of the estimators we choose to use a robust
and relative version of the Root Mean Squared Error (RMSE) defined as follows

$$
\text{RMSE}^* = \sqrt{\text{med} \left( \frac{\hat{\theta}_i - \theta_{i,0}}{\theta_{i,0}} \right)^2 + \text{mad} \left( \frac{\hat{\theta}_i}{\theta_{i,0}} \right)^2}
$$

with med(·) representing the median, mad(·) the median absolute deviation and \( \hat{\theta}_i \) and \( \theta_{i,0} \)
representing the \( i \)th element of the estimated and true parameter vectors respectively. The
RMSE* is therefore related to the RMSE and can also be used to assess the accuracy of
an estimator. The classical RMSE was also used allowing to reach equivalent conclusions
but the RMSE* was preferred to better highlight the difference between methods.

### 3.3.1 Time Series Model Estimation

For the simulation studies on the estimation of time series models, 500 samples of size
1000 were generated for each type of model described further on. Different types of con-
tamination were used to study the RGMWM, going from scale-contamination to additive
and replacement outliers as well as patchy outliers and level-shifts. Innovation-type con-
tamination was not considered since it did not appear to affect the estimators much (see
Maronna et al., 2006, for an overview of different contamination settings). We denote the
proportion of contaminated observations with \( \epsilon \) and the size of contamination (i.e. the
variance of the observations which are added to the uncontaminated observations) with
\( \sigma^2_\epsilon \). Finally, when dealing with level-shifts, we denote \( \mu_\epsilon \) as the size of the
\( i \)th shift in level.

Although the RGMWM is mainly conceived for the robust estimation of latent
time series models, to compare it with other classic and robust estimators we choose to study its
behaviour mainly on standard ARMA models for which it is also a consistent estimator
based on the conditions in Chapter 2. In this perspective, we compare the proposed
RGMWM estimator with:

- the Maximum Likelihood estimator (ML);
- the M-estimator for autoregressive models proposed by Kunsch (1984) (MAR);
- the Indirect Inference estimator based on the MAR (INDI) (see de Luna and M. G.
  Genton, 2001);
- the standard GMWM estimator (GMWM);
- the GMWM estimator used in Guerrier et al. (2014) (MPWM).

Given the substantial absence of general routines in statistical software to robustly es-
timate time series models, the choice of alternative methods available for comparison was
reduced. Among the many methods proposed, most of them required the computation of the explicit form of the score function or of a specific link function for each model and, once this is done, the tuning of certain parameters such as specific weighting matrices or others. Moreover, if these steps are achieved, the numerical stability of these methods is not necessarily guaranteed requiring an additional step to fine-tune the implementation. Given this and failing to obtain readily usable code from different authors, the choice of the methods was finally based on their direct availability within statistical software or their easy implementation based on these available tools. This corresponds to the setting in which a general researcher or practitioner, with basic statistical knowledge, would like to estimate model parameters in a robust manner. In this logic, using the robust regression tools available in the statistical software R, we implemented the robust MAR estimator proposed by Kunsch (1984) and, based on this, the INDI represents the corresponding easy-to-implement estimator for ARMA models (other more computationally efficient methods have been proposed such as the method proposed in Ortelli and Trojani, 2005). Nevertheless, when the latter estimator was used, the number of simulations for indirect inference was set to $H = 30$ since otherwise the computational time was beyond 5 minutes for each estimation and this issue was not improved by modifying the order of the auxiliary AR($p$) model. The MAR estimator was used exclusively for the estimation of AR($p$) models while the INDI estimator was used in all the other settings and was unweighted (i.e. the weighting matrix for indirect inference was the identity matrix). Considering this last aspect, to make the comparison fair, the GMWM, RGMWM and MPWM also used an unweighted matrix when being compared to the results of the INDI estimator. Moreover, considering the cases where the INDI estimator was used, a preliminary simulation study was carried out to determine the order of the auxiliary AR($p$) model. As for the starting values, the INDI and MPWM used the ML estimates as starting values while the GMWM and RGMWM also use the ML within the starting-value algorithm of the gmwm package in the R statistical software. Finally, considering the tuning parameters of the robust estimators, the MPWM estimator uses the median-type WV estimator used for the simulation study in Mondal and Percival (2012a) and therefore is highly robust. For this reason, the level of efficiency chosen for the MAR and RGMWM estimators was of 0.6 in order to guarantee a comparable level of robustness. Since the RGMWM is based on the WV estimated through the Tukey biweight function, this function was chosen also for the MAR estimator and, after a preliminary simulation study to determine the level of efficiency, a tuning constant $c_{MAR} = 2.2$ was chosen.

The performance of these estimators is investigated on the following models and contamination settings:

- **AR(1):** a zero-mean first-order autoregressive model with parameter vector $[\rho_1 \, \nu_2]^T = [0.9 \, 1]^T$, scale-based contamination at scale $j = 3$, $\epsilon = 0.01$ and $\sigma_\epsilon^2 = 100$;

- **AR(2):** a zero-mean second-order autoregressive model with parameter vector $[\rho_1 \, \rho_2 \, \nu_2]^T = [0.5 \, -0.3 \, 1]^T$, isolated outliers, $\epsilon = 0.05$ and $\sigma_\epsilon^2 = 9$;

- **ARMA(1,2):** a zero-mean autoregressive-moving average model with parameter vector $[\rho \, \rho_1 \, \rho_2 \, \nu_2]^T = [0.5 \, -0.1 \, 0.5 \, 1]^T$, and level-shift contamination with $\epsilon = 0.05$, $\mu_{\epsilon_1} = 5$ and $\mu_{\epsilon_2} = -3$;

- **ARMA(3,1):** a zero-mean autoregressive-moving average model with parameter vector $[\rho_1 \, \rho_2 \, \rho_3 \, \rho_1 \, \nu_2]^T = [0.7 \, 0.3 \, -0.2 \, 0.5 \, 2]^T$, patchy outliers, $\epsilon = 0.01$ and $\sigma_\epsilon^2 = 100$;
SSM: a state-space model ($X_t$) interpreted as a composite (latent) process in certain engineering applications. This model is defined as

$$Y_{t}^{(i)} = \rho(i) Y_{t-1}^{(i)} + W_{t}^{(i)}$$

$W_{t}^{(i)} \overset{iid}{\sim} \mathcal{N}(0, \upsilon_{2}^{(i)})$

$$X_t = \sum_{i=1}^{2} Y_{t}^{(i)} + Z_t,$$

$$Z_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

with parameter vector

$$[\rho(1) \upsilon_{2}^{(1)} \rho(2) \upsilon_{2}^{(2)} \sigma^2]^T = [0.99 0.1 0.6 2 3]^T,$$

isolated outliers, $\epsilon = 0.05$ and $\sigma_\epsilon^2 = 9$.

For each simulation, the sample size is $T = 1,000$ which delivers $J = 9$ scales for the GMWM-type estimators. This is one possible limitation of the latter estimators since, for identifiability issues, one needs at least as many moments or auxiliary parameters as the number of parameters of interest. Given the sample size, the larger models will rely more on the larger scales for which the WV estimators are less efficient and we therefore expect a decreased performance of these estimators for these models given this sample size. The ML and INDI estimators were not considered for the SSM simulations since they always failed to converge or gave questionable results using available software such as the MARSS package in the R statistical software. A smaller simulation study with $T = 100$ was carried out for this model and even in this case the convergence rates for these estimators were below 30%. This is highlighted in Table 3.1 which reports summary information regarding the estimation time in seconds for the RGMWM and INDI estimators in the contaminated setting for the ARMA(1,2), ARMA(3,1) and SSM models along with their convergence rates (the results for the uncontaminated setting can be found in Appendix C.3). Having stated this, the results of the simulation studies using the RMSE* are shown in Figure 3.1 where the MAR and INDI estimators are denoted as KUNSCH since they are used in complementary settings.
Figure 3.1: Top row: RMSE* of the estimators in an uncontaminated setting. Bottom row: RMSE* of the estimators in a contaminated setting. KUNSCH represents the MAR estimator for the AR(1) and AR(2) models while it represents the INDI estimator for the other models.
Table 3.1: Sample size, median computational time in seconds and convergence rates of the RGMWM and INDI estimators for the models \textbf{ARMA}(1,2), \textbf{ARMA}(3,1) and \textbf{SSM} in a contaminated setting.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample size</th>
<th>Median computational time in seconds</th>
<th>Conv. rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGMWM</td>
<td>ARMA(1,2)</td>
<td>1,000</td>
<td>5.59 \cdot 10^{-1}</td>
</tr>
<tr>
<td></td>
<td>ARMA(3,1)</td>
<td>1,000</td>
<td>7.9 \cdot 10^{-1}</td>
</tr>
<tr>
<td></td>
<td>SSM</td>
<td>100</td>
<td>5.6 \cdot 10^{-2}</td>
</tr>
<tr>
<td>INDI</td>
<td>ARMA(1,2)</td>
<td>1,000</td>
<td>1.79 \cdot 10^{2}</td>
</tr>
<tr>
<td></td>
<td>ARMA(3,1)</td>
<td>1,000</td>
<td>2.094 \cdot 10^{2}</td>
</tr>
<tr>
<td></td>
<td>SSM</td>
<td>100</td>
<td>1.025 \cdot 10^{2}</td>
</tr>
</tbody>
</table>

The first aspect to underline is that the RGMWM generally performs better than the MPWM in all the different settings and can therefore be considered as an improvement over the robust estimator investigated in Guerrier et al. (2014). Having stated this, when considering the \textbf{AR}(1), \textbf{AR}(2) models, the RGMWM does not lose much in uncontaminated settings while it performs generally as well or better than the MAR estimator in contaminated ones. As for the \textbf{ARMA}(1,2) model, the RGMWM is not as efficient as the others (excluding the MPWM) in the uncontaminated case while it adequately bounds the influence of outliers in contaminated ones, performing roughly as well or better than the INDI estimator. As for the \textbf{ARMA}(3,1) model, the RGMWM performs as well as the other estimators in uncontaminated settings while it clearly is the best estimator overall in contaminated ones. This can be related to the ease with which the GMWM estimators can estimate more complex models from a numerical point of view. The latter is confirmed looking at the results for the \textbf{SSM} simulations where it can be seen how the RGMWM is extremely close to the GMWM in uncontaminated settings while it is overall the best estimator in the contaminated ones. This is a clear advantage of the RGMWM since it provides a computationally efficient and numerically stable method to robustly estimate the parameters of many linear state-space models which has been almost unfeasible in practice to date.

There are a few points that must be emphasized considering the above results. First of all, as Table 3.1 shows, the RGMWM can estimate all these models in under a second while the INDI estimator can take up to 300 times more, with timings going up to more than 8 minutes, and does not consider the timing for those cases in which it does not converge numerically. Secondly, the INDI estimator can eventually use the RGMWM as a starting value thereby avoiding to use an algorithm to find adequate starting values which can increase its computational times even more. Moreover, it must be emphasized again that the RGMWM (as the other GMWM estimators) does not use a full weighting matrix in these cases and consequently its performance could greatly be improved as seen in the \textbf{AR}(1) and \textbf{AR}(2) model simulations.

### 3.3.2 Spatial Model Estimation

In this section we study the RGMWM in the estimation of spatial models. For this purpose, regular lattice \( K \times M \) random fields were simulated 500 times, with \( K = M = 30 \), thereby delivering sample sizes of \( N = 900 \) and \( J = 10 \) scales of wavelet decomposition since we limit ourselves to isotropic models which therefore only require the lower triangu-
lar WV matrix. Larger sample sizes were not considered since this became computationally impractical for the maximum likelihood estimator which is already time-demanding for this problem size. In this case, the following estimators were considered in addition to the RGMWM:

- the Maximum Likelihood estimator (ML);
- the standard GMWM estimator (GMWM);
- the GMWM estimator based on the robust estimator of WV proposed by Mondal and Percival (2012a) (MPWM).

To the best of our knowledge, there is no existing proposed or implemented robust procedure for the estimation of the standard spatial models considered in this section. Therefore the comparison is only made with the ML and GMWM that are non-robust estimators.

The performance of these estimators is investigated on the following models and contamination settings:

- **Exp(1)**: a zero-mean Exponential model with parameter vector \([\phi \sigma^2]^T = [2 1]^T\) and level-shift contamination with \(\epsilon = 0.05\), \(\mu_{\epsilon_1} = 5\) and \(\mu_{\epsilon_2} = -3\);
- **Exp(2)**: a sum of two zero-mean Exponential models with parameter vector \([\phi_{(1)} \sigma_{(1)}^2 \phi_{(2)} \sigma_{(2)}^2]^T = [2 1.5 1 1.5 1]^T\), isolated outliers, \(\epsilon = 0.01\) and \(\sigma^2_{\epsilon} = 100\);
- **Gauss(1)**: a zero-mean Gaussian model with parameter vector \([\phi \sigma^2]^T = [2 1]^T\), patchy outliers, \(\epsilon = 0.01\) and \(\sigma^2_{\epsilon} = 100\);
- **Gauss(2)**: a sum of two zero-mean Gaussian models with parameter vector \([\phi_{(1)} \sigma_{(1)}^2 \phi_{(2)} \sigma_{(2)}^2]^T = [2 1.5 1 1.5 1]^T\), isolated outliers, \(\epsilon = 0.05\) and \(\sigma^2_{\epsilon} = 9\);

The only model for which the ML was used was the **Exp(1)** model since it was unable to estimate the Gaussian and the latent models due to numerical instability using either the liikfit() function in the R package geoR or a tailored implementation of the ML. With this in mind, similarly to Section 3.3.1, the RMSE* of these estimators for the above models are given in Figure 3.2.
Figure 3.2: Top row: RMSE* of the estimators in an uncontaminated setting. Bottom row: RMSE* of the estimators in a contaminated setting.
The first aspect to underline is that, in any case, the GMWM and RGMWM estimators provide computationally efficient and statistically sound solutions for the estimation of standard and latent spatial models which are often numerically challenging for existing methods such as the ML (which could only be used for the $\text{Exp}(1)$ model). Nevertheless, focusing on the simulation results for the $\text{Exp}(1)$ model, the ML is by far the best estimator in the uncontaminated setting and we would expect similar results for the $\text{Gauss}(1)$ model (and the others) if there weren’t any numerical issues. However, it is also clear that performance of the ML is worse under contamination where the RGMWM, being the only robust estimator available, is the best in bounding the influence of outliers. In all other settings, we can see that the GMWM is slightly better than the RGMWM in the uncontaminated settings whereas the performance of the latter does not change much under contamination and therefore makes it the preferred estimator in these settings.

To summarize, as the simulation studies in this section and in Section 3.3.1 have shown, the RGMWM estimator represents a good robust and general purpose estimator for time series and spatial models, even when these are of considerable complexity and the number of observations is large, delivering estimation and inference results in a computationally efficient and numerically stable manner.

3.4 Applications

The wide class of intrinsically stationary models that the RGMWM can estimate in a robust manner allows it to be used in a large variety of applications where outliers and contamination can often occur. In this section we will therefore investigate the performance of the proposed RGMWM estimator on some real data sets coming from:

- Engineering;
- Economics;
- Hydrology;
- Climatology.

In all these applications different models are used, going from simple AR(1) models (for time series) and Exponential models (for spatial data) to complex latent models which include ARMA models. To select these models, a bootstrap version of the J-test (see Hansen, 1982) was used, as suggested in Guerrier et al. (2013b), therefore selecting the models for which we could not reject the null hypothesis that they fit the data well.

3.4.1 Application to Inertial Sensors

The engineering data set consists in the angular rate signal issued from a micro-electromechanical system gyroscope in static conditions. Due also to their low cost, these sensors are very common and are being increasingly used in the field of navigation engineering. The main goal of recording this kind of data is to improve the performance of the navigation sensors by identifying and estimating the parameters of the error model coming from the accelerometers and gyroscopes that compose the sensor. Once these parameters are estimated they are inserted in a filter (usually an extended Kalman filter) which is used within a navigation system. The latter collects measurements from different sources
such as Global Positioning Systems (GPS) or the inertial sensors themselves in an optimal manner in order to improve the navigation precision. Therefore, the latter greatly depends on the estimation of the parameters of the selected error model for the inertial sensor.

Figure 3.3 shows the error signal from the gyroscope along with the outliers in a portion of the signal identified via the weights given to the observations by the RGMWM estimator. As can be observed, there are outliers that would appear to be obvious by simply looking at the plot and could be treated by fault detection algorithms for navigation systems (see further on) but there are many others that lie within the part of the signal which one would not expect to contain outliers. Despite the numerous outliers, these are extremely low in proportion to the whole dataset ($\approx 0.4\%$) which contains a little under 900,000 observations (issued from an approximately 2.5 hours-long recording sampled at 100 Hz). This may lead to think that estimations on this dataset would not be significantly influenced by outliers.

Nevertheless, to understand how influential these observations could be, we estimated the classical and robust WV from the signal represented in Figure 3.3. Using these estimates we then estimated an error model made by the sum of three latent first-order autoregressive models. This state-space model is among those suggested by Stebler et al. (2014b) as being most appropriate to describe such signals. Table C.2 in Appendix C.4 shows the estimated parameters for the GMWM and RGMWM estimators together

![Figure 3.3: Top part: Inertial sensor time series. Bottom part: zoom-in on grey part of the time series with black points indicating extreme outliers identified through the weights of RGMWM.](image)
3.4. Applications

with their confidence intervals (the ML was not considered for the same reasons given in Section 3.3 for the SSM model: the numerical stability and computational efficiency are unreasonable for this model and sample size). For both estimators the values of some autoregressive parameters are close to one, suggesting that the AR(1) model could be considered as a random walk. Indeed, a model that was commonly used to describe these signals was the sum of a white noise process with a random walk. However, Stebler et al. (2014b) show how the use of sums of AR(1) models greatly improves the navigation performance over this model and the J-tests and confidence intervals support this view by ruling out the models which included a random walk. Although the differences between the estimations do not appear to be large since the estimated level of contamination is low, a significant difference is to be noticed for the parameters of the first two autoregressive processes indicating that the contamination appears to have an impact on estimation and that robust methods should be preferred (assuming the Gaussian assumption holds). Even one (or few) slightly misestimated parameter(s) can be highly relevant in the context of navigation systems since these are fed into the filters which will progressively misestimate the position as the sensors work in “coasting mode” (i.e. without the GPS integration) and deliver the so-called “error accumulation”. Informally speaking this is due to the fact that these measurements are integrated several times and therefore their errors accumulate in time especially when no GPS observations are present to “reinitialize” the system (more details on this can be found, for example, in Titterton and Weston, 2004).

Moreover, our robust approach can be of great usefulness in the area of Fault Detection and Isolation (FDI) for inertial measurement units (see for example Guerrier et al., 2012, and references therein) as shown in Figure 3.3. In general, the task of FDI includes the detection of the presence of failures (or outliers) and the isolation of the component responsible of the irregularity. In the inertial navigation framework, FDI algorithms are used, for example, to ensure the safety of aircrafts or robots which deeply rely on inertial sensors. In fact, usual FDI methods in this area use various measurements coming from several sensors which entail a series of disadvantages. Moreover, these methods often make use of unconditional cut-off values which generally determine which observations are “unusual”. On the other hand, the proposed approach would be able to detect “unusual” observations conditioned on previous ones instead of detecting them based on a general cut-off value. Although this is left for future research as some further adjustments would need to be put in place, our approach could be used as a basis for FDI by only using one signal coming from the sensor calibration procedure. One of the advantages of this approach, in addition to those already mentioned, is that it would have important impacts in terms of costs and constraints (e.g. weight, electric consumption, etc.) for robots or small unmanned aerial vehicles which are currently a major focus of technological and mechanical research.

3.4.2 Application to Personal Saving Rates

The economics data set consists in the monthly seasonally adjusted Personal Saving Rates (PSR) data from January 1959 to May 2015 provided by the Federal Reserve Bank of St. Louis. The study of PSR is an essential part of the overall investigation on the health of national and international economies since, within more general economic models, PSR can greatly impact the funds available for investment which in turn determine the productive capacity of an economy. Understanding the behaviour of PSR is therefore an important step in correct economic policy decision making. In this sense, Slacalek and
Chapter 3. Fast and Robust Parametric Estimation for Time Series and Spatial Models

Table 3.2: Random Walk plus ARMA(2,1) model estimates for the PSR data. Estimated parameters with GMWM and RGMWM estimators with $\gamma^2$ being the random walk parameter, $\rho_i$ the $i^{th}$ autoregressive parameter, $\varrho$ the moving average parameter and $\sigma^2$ the innovation variance of the ARMA(2,1) model. Confidence intervals (CI) based on the approach used in Guerrier et al. (2013b).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GMWM Estimate CI(%, 95%)</th>
<th>RGMWM Estimate CI(%, 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma^2$</td>
<td>$7.95 \cdot 10^{-2}$ (3.67 $\cdot$ 10$^{-2}$; 1.11 $\cdot$ 10$^{-1}$)</td>
<td>$5.85 \cdot 10^{-2}$ (1.54 $\cdot$ 10$^{-2}$; 9.97 $\cdot$ 10$^{-2}$)</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>$1.64 \cdot 10^{-1}$ (5.93 $\cdot$ 10$^{-2}$; 2.89 $\cdot$ 10$^{-1}$)</td>
<td>$6.00 \cdot 10^{-1}$ (4.48 $\cdot$ 10$^{-1}$; 7.55 $\cdot$ 10$^{-1}$)</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>$3.06 \cdot 10^{-3}$ (-1.31 $\cdot$ 10$^{-1}$; 1.48 $\cdot$ 10$^{-1}$)</td>
<td>$1.84 \cdot 10^{-1}$ (3.10 $\cdot$ 10$^{-2}$; 2.46 $\cdot$ 10$^{-1}$)</td>
</tr>
<tr>
<td>$\varrho$</td>
<td>$2.43 \cdot 10^{-1}$ (2.02 $\cdot$ 10$^{-1}$; 2.81 $\cdot$ 10$^{-1}$)</td>
<td>$2.92 \cdot 10^{-1}$ (2.28 $\cdot$ 10$^{-1}$; 3.45 $\cdot$ 10$^{-1}$)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$3.14 \cdot 10^{-1}$ (2.59 $\cdot$ 10$^{-1}$; 3.85 $\cdot$ 10$^{-1}$)</td>
<td>$1.32 \cdot 10^{-1}$ (8.59 $\cdot$ 10$^{-2}$; 1.80 $\cdot$ 10$^{-1}$)</td>
</tr>
</tbody>
</table>

Sommer (2012) study the factors behind saving rates and investigate different models which, among others, are compared to the random-walk-plus-noise (local level) model (RWN). As opposed to the latter model, various time-varying models are proposed in the literature to explain precautionary PSR together with risk aversion in the light of different factors such as financial shocks or others (see, for example, Brunnermeier and Nagel, 2008; Videras and Wu, 2004). Nevertheless, as emphasized in Pankratz (2012), modelling the time series with a stationary model, or a $d^{th}$-order non-stationary model such as an ARIMA, can be useful under many aspects such as, for example, to understand if a dynamic model is needed for forecasting and, if so, what kind of model is appropriate.

In this example, we consider the RWN model and, as in Section 3.4.1, we use the WV log-log plot and a J-test to understand what kind of model could fit the time series. By doing so, we find that a random walk plus an ARMA(2,1) fits the data well and therefore, in this case, we have that the “noise” in the RWN model is an ARMA(2,1). This can be seen in Figure 3.4 where, in the top part, the saving rate time series is represented along with the identified outliers and, in the bottom part, we see the log-log representation of the classic and robust estimated and model-implied WV respectively. Indeed, for the bottom part, the diagonal plots show the classic and robust estimations respectively, each with the estimated WV and the WV implied by the estimated model. The off-diagonal plots compare the classic and robust estimated WV (upper diagonal) and the WV implied by the GMWM and RGMWM model parameter estimates (lower diagonal). It can be seen how there is a significant difference between the classic and robust WV estimates, especially at the first scales where the confidence intervals of the estimated WV do not overlap (upper diagonal plot). This leads to a difference in the model-implied WV whose parameters have been estimated through the GMWM and RGMWM (lower diagonal plot).

The estimated parameters with the GMWM and RGMWM are given in Table 3.2 along with their respective confidence intervals. There are two main differences between the two estimations: (i) the estimates of the first autoregressive parameter $\rho_1$ and innovation variance $\sigma^2$ are significantly different; (ii) the second autoregressive parameter $\rho_2$ is not significant using the GMWM. These differences highlight how the conclusions concerning parameter values and model selection can considerably change when outliers are present in the data. Indeed, the choice of the model would then affect the decisions taken towards the selection of appropriate causal and dynamic models to better explain the behaviour of saving rates and the economy as a whole. The selected model based on the robust fit in
Figure 3.4: Top figure: Saving rates time series with different types of points indicating outliers identified through the weights of the RGMWM. Bottom figure: log-log scale WV plots for saving rates series; Top left: classic estimated WV superposed with model-implied WV based on the parameters estimated through the GMWM. Top right: classic and robust estimated WV with respective confidence intervals superposed. Bottom left: classic and robust model-implied WV based on the GMWM and RGMWM estimates respectively. Bottom right: robust estimated WV superposed with model-implied WV based on the parameters estimated through the RGMWM.

In fact, be interpreted as a sum of latent models along the lines given in Slacalek and Sommer (2012) where the ARMA(2,1) can be seen as a sum of two AR(1) models where each of them represents, for example, the reaction of PSR to changes in uncertainty (affected by unemployment) and interest rates, respectively, while the random walk describes the continuous fluctuations of target wealth which also drives PSR.

The additional benefit of the RGMWM estimator, as opposed to the median-type MPWM, is also to deliver weights that allow to identify outliers which may not be visible simply by looking at the time series (as highlighted in Section 3.4.1). As shown in the top part of Figure 3.4, the outliers identified by the RGMWM can be interpreted in the
light of the national and global economic and political events. Limiting ourselves to the major identified outliers, the first one corresponds to a rise in the precautionary savings in the aftermath of the OPEC oil crisis and the 1974 stock market crash. In the months following October 1987 we can see an instability in the PSR with a rise and sudden fall linked to the “Black Monday” stock market crash which added to the savings and loans crisis which lasted to the early 1990s. This period also saw an economic recession where a rise in the saving rates, highlighted by the presence of high outliers, led to a drop in aggregate demand and bankruptcies. Finally, the various financial crises of the 21st century led to sudden and isolated rises in PSR as indicated again by the outliers.

3.4.3 Application to Precipitation Data

The hydrology data set collects the monthly precipitation data from 1907 to 1972 provided in Hipel and McLeod (1994) and is shown in the top panel of Figure 3.5. The modelling approach described in this section is the Environmental System Model (ESM) of a watershed which, despite being less used due to other more recent approaches (such as, for example, adaptive neural networks in Tokar and Johnson, 1999), can still be highly useful for practitioners who wish to have a straightforward and clear tool to describe and interpret phenomena linked to the water cycle. Moreover, the example clearly shows how our method can help detect dependence where classical methods may not due to contamination in the data (be this in the domain of hydrology or others). The goal of the ESM is to explain how water resources behave and are distributed throughout their cycles from the stage of precipitation to river flows. Salas and Smith (1981) describe how the precipitation model is the basis for the models of the following stages in the ESM. Three models are envisaged by Salas and Smith (1981) for the precipitation stage among which the independent precipitation (i.e. a white noise process) and the AR(1) model.

Taking a look at the time series, it may not appear to be Gaussian but the GMWM and RGMWM are based on the inherent first-differencing of the series through the Haar wavelet filter so they can be used in this case. Analyzing the AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) in the bottom panels of Figure 3.5 one would identify an independent precipitation model for the this dataset. However, an AR(1) model was fitted to understand if the independent model was reasonable. Table 4.2 shows the estimated parameters for the AR(1) precipitation model. The latter model has different estimates between the classical estimators and the proposed robust one. In fact, the ML and GMWM estimates tend to agree with the independent model assumption where the Confidence Intervals (CI) for the autoregressive parameter are close to or include the value of zero, whereas the RGMWM detects a stronger dependence with the previous precipitation measurement and a smaller variance of the innovation process (with CI not overlapping those of the ML and GMWM). This could be due to the fact that the classical ACF and PACF are sensitive to outliers and may not detect this correlation structure (see Maronna et al., 2006).

If the ESM were to be used in this context, it would be greatly affected by a misspecified model for the precipitation since it would condition the model choice and relative parameter estimation in the following phases of the water cycle. In this example, the choice of an independent precipitation model would have lead to a domino-effect in terms of model misspecification and misestimation leading to possibly highly incorrect interpretations and conclusions.
3.4. Applications

![Graph](image.png)

Figure 3.5: (Top panel) Monthly precipitation series from 1907 to 1972 taken from Hipel and McLeod (1994). (Bottom panels) Estimated autocorrelation function (left) and estimated partial autocorrelation function (right) of the precipitation series.

Table 3.3: AR(1) estimates for the mean monthly precipitation data from 1907 to 1972 taken from Hipel and McLeod (1994). Estimated parameters with ML, GMWM and RGMWM estimators with $\hat{\phi}$ being the estimated autoregressive parameter and $\hat{\sigma}^2$ the innovation variance. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b).

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\phi}$</th>
<th>$\hat{\sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>$6.843 \cdot 10^{-2}$</td>
<td>$2.199 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[1.110 \cdot 10^{-3}, 1.212 \cdot 10^{-1}]$</td>
<td>$[2.003 \cdot 10^{-1}, 2.3929 \cdot 10^{-1}]$</td>
</tr>
<tr>
<td>GMWM</td>
<td>$5.577 \cdot 10^{-2}$</td>
<td>$2.179 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[-4.357 \cdot 10^{-3}, 1.153 \cdot 10^{-1}]$</td>
<td>$[1.985 \cdot 10^{-1}, 2.365 \cdot 10^{-1}]$</td>
</tr>
<tr>
<td>RGMWM</td>
<td>$4.049 \cdot 10^{-1}$</td>
<td>$1.066 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[3.345 \cdot 10^{-1}, 4.662 \cdot 10^{-1}]$</td>
<td>$[9.623 \cdot 10^{-2}, 1.175 \cdot 10^{-1}]$</td>
</tr>
</tbody>
</table>

3.4.4 Application to Cloud Data

The cloud data set consists in four regions selected from a temperature map of the southeast Pacific stratocumulus clouds near the Chilean coast made available by Geo-stationary Operational Environmental Satellites (GOES) Imagery (see Mondal and Percival, 2012a, for details on this data). These regions were selected by atmospheric scientists and have already been analysed in Mondal and Percival (2012b) and in Chapter 1 using the WV as a means to detect variance patterns according to the type of cloud formation. The study of these cloud formations and their related dynamics is of particular importance
Table 3.4: Exponential model estimates for the cloud image data in the third region using the GMWM and RGMWM estimators. \( \hat{\phi} \) is the estimated range parameter and \( \hat{\sigma}^2 \) the estimated sill. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b).

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\phi} )</th>
<th>( \hat{\sigma}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMWM</td>
<td>( 1.35 \cdot 10^0 )</td>
<td>( 8.443 \cdot 10^{-2} )</td>
</tr>
<tr>
<td>CI</td>
<td>( [1.128 \cdot 10^0, 1.575 \cdot 10^0] )</td>
<td>( [6.537 \cdot 10^{-2}, 1.035 \cdot 10^{-1}] )</td>
</tr>
<tr>
<td>RGMWM</td>
<td>( 1.5 \cdot 10^0 )</td>
<td>( 2.415 \cdot 10^{-2} )</td>
</tr>
<tr>
<td>CI</td>
<td>( [1.403 \cdot 10^0, 1.596 \cdot 10^0] )</td>
<td>( [8.35 \cdot 10^{-3}, 3.994 \cdot 10^{-2}] )</td>
</tr>
</tbody>
</table>

To understand the radiation patterns which this region is subject to, also considering the presence of specific industrial activities which release atmospheric aerosol and make this a complex region to study from a climatological point of view.

More specifically, Mondal and Percival (2012b) and Chapter 1 highlighted that the third and fourth regions delivered considerably different patterns when using the standard and robust estimators of WV. A possible reason for this difference can be identified in the fact that, as opposed to the first two regions, the third and the fourth depict inhomogeneous or varying scenarios since they represent respectively a formation of broken clouds and a forming Pocket of Open Cells (POC) which are clear areas in the clouds that are mainly characterized by low-aerosol air mass. Indeed, the first two regions represent a well-formed POC and a uniform formation of clouds respectively and, possibly due to this “stability”, the variance patterns did not appear to change much between the standard and robust analysis. Correctly analysing the variance patterns is essential to identify the type of cloud formation and consequently select the appropriate model to describe these climatological phenomena. In Mondal and Percival (2012a) they consider an Exponential covariance model to describe these regions. Let us denote this covariance model as \( \varphi(d) \), where \( d \) is the Euclidean distance between two variables \( X_k \) and \( X_{k'} \) (with \( k \) denoting the coordinates of a location), and therefore define the Exponential covariance model as

\[
\varphi(d) = \sigma^2 \exp \left( -\frac{d}{\phi} \right),
\]

where \( \phi \) is the range parameter and \( \sigma^2 \) is the sill. After testing for isotropy using a robust version of the method proposed in Thon et al. (2015), we cannot reject the hypothesis that the images are isotropic, so we can therefore consider the Exponential model as a valid candidate. Using the GMWM and RGMWM to estimate its parameters, the results of these two methods on the third and fourth regions are presented in Tables 3.4 and 3.5 along with their confidence intervals.

For both regions, the estimates of the range parameter \( \phi \) for the GMWM and RGMWM are not significantly different and substantially agree with the interpretation in Mondal and Percival (2012a) who, taking into account their reparametrization, suggest a value of \( \phi \) roughly between 1 and 2. However, a significant difference can be seen in the estimates of the sill parameter \( \sigma^2 \) which determines the unconditional variance of the random field. This parameter can make a considerable difference in practical situations where the estimated models are used as climatological models for simulation purposes in order to determine the evolution and interactions of cloud formations. An overestimation of the
sill parameter $\sigma^2$ can introduce an undesirable increase in the variability of the simulation studies, thereby adding uncertainty to the conclusions made based on these models.

### Table 3.5: Exponential model estimates for the cloud image data in the fourth region using the GMWM and RGMWM estimators. $\hat{\phi}$ is the estimated range parameter and $\hat{\sigma}^2$ the estimated sill. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b). 

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\phi}$</th>
<th>$\hat{\sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMWM</td>
<td>$1.453 \cdot 10^0$</td>
<td>$3.615 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[1.385 \cdot 10^0, 1.52 \cdot 10^0]$</td>
<td>$[2.914 \cdot 10^{-2}, 4.316 \cdot 10^{-2}]$</td>
</tr>
<tr>
<td>RGMWM</td>
<td>$1.456 \cdot 10^0$</td>
<td>$1.063 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[1.416 \cdot 10^0, 1.496 \cdot 10^0]$</td>
<td>$[4.225 \cdot 10^{-3}, 1.703 \cdot 10^{-2}]$</td>
</tr>
</tbody>
</table>

This chapter presented a new framework for the robust estimation of time series and spatial models called RGMWM which extends to various classes of models that are intrinsically stationary. This framework provides estimators which are easy-to-implement, computationally efficient and have suitable asymptotic properties. The simulation studies and the applied examples confirm that the robust estimators delivered via the proposed approach adequately bound the influence of outliers on the estimation procedure and compare satisfactorily to alternative estimators which, with a few exceptions, are numerically challenging and/or computationally intensive. This chapter hence provides a contribution in the direction of developing a general theoretical framework to robust inference for (latent) time series and spatial models as well as a method which is computationally efficient and straightforward to implement in practice thereby diminishing many of the challenges that researchers and practitioners are faced with when dealing with contamination in dependent data with possibly complex models and large sample sizes.

**3.5 Conclusion**

This chapter presented a new framework for the robust estimation of time series and spatial models called RGMWM which extends to various classes of models that are intrinsically stationary. This framework provides estimators which are easy-to-implement, computationally efficient and have suitable asymptotic properties. The simulation studies and the applied examples confirm that the robust estimators delivered via the proposed approach adequately bound the influence of outliers on the estimation procedure and compare satisfactorily to alternative estimators which, with a few exceptions, are numerically challenging and/or computationally intensive. This chapter hence provides a contribution in the direction of developing a general theoretical framework to robust inference for (latent) time series and spatial models as well as a method which is computationally efficient and straightforward to implement in practice thereby diminishing many of the challenges that researchers and practitioners are faced with when dealing with contamination in dependent data with possibly complex models and large sample sizes.
Chapter 4

The gmwm R package

The gmwm R package for inference on time series models is mainly based on the quantity called wavelet variance which is derived from a wavelet decomposition of a time series. This quantity provides a means to summarize and graphically represent the features of time series in order to identify possible models. Moreover, it is used as a moment condition for model estimation through the generalized method of wavelet moments. Based on the latter method, this package not only provides an alternative method to estimate classical ARMA models but also delivers a general framework for the robust estimation of many time series models as well as a quick and efficient estimation of many linear state-space models.

4.1 Introduction

Time series analysis in R can be carried out with a wide range of tools and packages. However, there are many practical limitations to the methods which are currently implemented when it comes to specific requirements such as, for example, a user-friendly estimation of certain state-space models, robust inference for time series models and estimation on large datasets. The new gmwm R package helps to overcome some of these limitations by implementing the Generalized Method of Wavelet Moments (GMWM), which was proposed by Guerrier et al. (2013b), along with its relative inference tools. This method uses a quantity called the Wavelet Variance (WV), which is the variance of the wavelet coefficients that are issued from a wavelet decomposition of a time series (see for example Percival and Walden, 2006). The WV is a widely used quantity in different fields such as geology and aerospace engineering since it helps to decompose and interpret the variance of a time series across different “scales” and represents a good statistic to summarize the “information” of time series which respect certain properties (e.g. intrinsically stationary). The GMWM uses this quantity as an auxiliary parameter in a minimum distance estimator setting allowing it to estimate a wide range of intrinsically second-order stationary models in a numerically stable and computationally efficient manner.

In this chapter, we will focus primarily on some specific features of the gmwm package, highlighting other important features in the process. More specifically, this package delivers two main advantages for univariate\(^1\) time series model estimation compared to currently available packages: (i) easy-to-use, computationally efficient and numerically stable estimation of linear state-space models and (ii) robust estimation and inference

\(^1\)The extension to multivariate time series is a current research topic.
for the whole range of time series models available for estimation in the package. Moreover, for these two aspects, a related model selection procedure is available based on the Wavelet Information Criterion (WIC) (see Guerrier et al., 2015). For this reason we will list the main tools available in R for these two particular goals whereas we refer the reader to Shumway and Stoffer (2013) and the R website\(^2\) for a more general overview of all available packages and functions for time series analysis.

Firstly, when dealing with parameter estimation of state-space models there is a relative abundance of options in R for which we provide a non-exhaustive summary. The \texttt{KFKSDS} package, for example, allows to treat state-space models via Kalman-filtering as does the \texttt{dlm} package for Gaussian linear state-space models. The \texttt{bsts} package uses Markov-Chain Monte-Carlo to simulate from the posterior distribution of Bayesian structural time series models which provides estimates for the coefficients representing the states in the desired model. The \texttt{dse} package represents a general framework for linear and multivariate time series estimation with general AutoRegressive Moving Average (ARMA) representations and functions allowing for the conversion of these models to state-space model representations. An even more general and computationally efficient framework is represented by \texttt{KFK} where a fast implementation of the Kalman filter is available to estimate state-space models also for large datasets while the \texttt{KFAS} package also makes forecasting tools available. The \texttt{dlm} package represents one of the main available packages when it comes to maximum-likelihood estimation of Gaussian linear state-space models which includes Bayesian analysis tools as well as time series smoothing options. Along these lines we also find the \texttt{MARSS} package which allows for the estimation of multivariate first-order autoregressive state-space models with Gaussian errors via the Expectation-Maximization algorithm, with the additional possibility of constraining the estimation procedure and adding covariates. Finally, the \texttt{pomp} package extends the estimation possibilities to non-Gaussian and non-linear state-space models by using the framework of partially observed Markov processes.

On the other hand, as far as robustness is concerned, the availability of R functions that allow robust estimation and inference for time series models is almost nonexistent for classic models such as ARMA, without mentioning state-space models. The only package which directly deals with robustness for time series is \texttt{robfilter} which makes available a series of robust filters to smooth the observed time series. Another package which includes robustness for time series analysis is \texttt{robcor} in which the function \texttt{robacf()} delivers a robust equivalent of the function \texttt{acf()} to compute the empirical autocovariance and autocorrelation function. However, none of these are able to deal directly with robust parametric estimation and inference for time series models. An interesting option that is available for this purpose is the \texttt{quantreg} package in which the function \texttt{dynrq()} allows to perform robust-type estimation of autoregressive models (being based on quantiles). A “naive” approach resembling the latter method would be to use the robust regression functions such as \texttt{rlm()} or \texttt{lmrob()}, in the \texttt{MASS} and \texttt{robustbase} packages respectively, to also robustly estimate autoregressive models. These autoregressive models can then eventually be used as auxiliary models to estimate more complex models via indirect inference (see Gourieroux et al., 1993) but this can come at the price of computational inefficiency and numerical convergence problems when the models are complex.

As a final note, as highlighted above there are different solutions which are either directly available or can be obtained using different tools in R. Nevertheless, even when solutions are available, these may not necessarily be numerically stable or computationally

\(^2\)https://cran.r-project.org/web/views/TimeSeries.html
efficient as stated above. For many applications, such as economics, the time series are roughly of length $T = 1,000$ or smaller and, if discarding the need for robust analysis, these problems can easily be treated using currently available tools. However, there are many other applications where the observed time series can easily be much longer (sometimes in the order of 500,000 or 1 million) and a comprehensive estimation and inference procedure for these cases is hardly feasible by using the available tools in R.

Considering the above, the \texttt{gmwm} package adds to the state-space model estimation tools by providing numerically stable and computationally efficient estimations also for large datasets while delivering the first readily available software for the robust estimation of a wide range of intrinsically second-order stationary (state-space) models based on the Robust GMWM (RGMWM) proposed in Chapter 3. All this is available along with a series of inference and model selection tools which allow to have a general purpose package for time series analysis. With this in mind, Section 4.2 briefly introduces the syntax used in the package by focusing on time series model simulation, thereby listing the time series models available for simulation and estimation purposes. Section 4.3 discusses the tools the package makes available for computing and representing the WV of a time series while in Section 4.4 we present the estimation and inference tools for different time series models from ARMA to state-space models where their corresponding robust framework is described. Finally, Section 4.5 briefly discusses the computational efficiency of the new package and Section 5.5 concludes by providing a list of upcoming and future features with which the package will be updated.

### 4.2 Time Series Models

In this section we briefly list, describe and provide the syntax for the models available for estimation and simulation purposes within the \texttt{gmwm} package. Under some minor constraints, all these models can then be combined into a specific class of linear state-space models. Having stated this, some of the basic models available in the \texttt{gmwm} package to simulate from (and to estimate) are the following:

- White Noise ($\text{WN}()$);
- Quantization Noise ($\text{QN}()$);
- Random Walk ($\text{RW}()$);
- Drift ($\text{DR}()$);
- AR(1): First-order autoregressive process ($\text{AR1}()$);
- ARIMA: Integrated Autoregressive Moving Average process ($\text{ARIMA}()$);
- SARIMA: Seasonal ARIMA process ($\text{SARIMA}()$).

The expressions in brackets in the above list represent the syntax used in the \texttt{gmwm} package to specify the model. The only model which may be less known is the quantization noise which is a process that is often used in engineering fields and can be described in layperson terms as being a good approximation of a rounding error. The brackets for the syntax of each model are left to specify parameter values for simulation purposes or to specify starting values for estimation purposes. The code below shows how these model
specifications can be used for simulations based on the built-in function \texttt{gen.gts()} which allows to generate samples from all these models.

```r
# Set seed for reproducibility
set.seed(1337)

# Number of observations
n = 1e4

# Generate a White Noise Process
wn = gen.gts(WN(sigma2 = 1), n)

# Generate a Quantization Noise
qn = gen.gts(QN(q2 = .5), n)

# Generate a Random Walk
rw = gen.gts(RW(gamma2 = .75), n)

# Generate a Drift
dr = gen.gts(DR(omega = 0.10), n)

# Generate an AR(1)
ar1 = gen.gts(AR1(phi = .9, sigma2 = .1), n)

# Generate an MA(1)
ma1 = gen.gts(MA1(theta = .3, sigma2 = .5), n)

# Generate an ARMA(1,1)
arma11 = gen.gts(ARMA11(phi = .9, theta = .2, sigma2 = 1), n)

# Generate an SARIMA(1,0,1)x(2,1,1)[12] process
sarima = gen.gts(SARIMA(ar = 0.3, i = 0, ma = -0.27, 
sar = c(-0.12, -0.2), si = 1, sma = -0.9, 
sigma2 = 1.5, s = 12), n)
```

The \texttt{gmwm} package therefore allows to easily simulate from a wide variety of models, including SARIMA models, but does not limit itself to these basic models. Indeed, under some restrictions highlighted in Section 4.4, these models can be combined in different ways to deliver many state-space (latent) models which can be represented by the sum of basic models. The construction of such linear state-space models is very simple with the \texttt{gmwm} package allowing it to be considerably user-friendly. In fact, to specify that a model is a combination of different models, all that is needed is to use the “+” symbol between them and, supposing that different \texttt{AR1()} processes are present in the state-space model, the syntax to insert “k” of these models in a state-space model is \texttt{k*AR1()}. So, for example, the sum of three \texttt{AR(1)} models, a random walk and a white noise process can be specified as: \texttt{3*AR1()+RW()+WN()}. A function specifically provided to generate and represent these models is the \texttt{gen.lts()} function which, while simulating from these models, also gives the option to plot a breakdown of the underlying processes by applying
the \texttt{plot()} function on the result of \texttt{gen.lts()} (see Section 4.4 for an example of this feature). It must be noted that this particular function is designed specifically to highlight how these state-space models are obtained and that the \texttt{gen.gts()} function should instead be used if the sole purpose is to simulate time series since it is computationally faster.

### 4.3 Modeling based on the Wavelet Variance

As highlighted in the introduction, the WV is a very useful quantity for the analysis of time series and provides the auxiliary parameter for the GMWM. Therefore in this section we discuss how this quantity is estimated in the package and underline how it can be used for modelling purposes, consequently inviting the reader to go directly to Section 4.4 if their interest lies mainly in model estimation and inference. To introduce the WV, which we denote as \( \nu_j^2 \), let \((Y_t)\) represent a time series and let \((W_{j,t})\) denote the wavelet coefficient process issued from the a Haar wavelet decomposition of \((Y_t)\) at the \(j^{th}\) scale (Percival and Walden, 2006). Then the WV at scale \( j \) can be defined as \( \text{Var}(W_{j,t}) \) (i.e. the variance of the wavelet coefficients) and an unbiased estimator for this quantity was proposed by Percival (1995) and is given by

\[
\hat{\nu}_j^2 = \frac{1}{M_j} \sum_{t=1}^{M_j} W_{j,t}^2,
\]

where \( M_j \) represents the number of wavelet coefficients at the considered scale. An M-estimator which generalizes this estimator was proposed in Chapter 1 that can be made robust by selecting a bounded estimating function which, for the \texttt{gmwm} package, consists in the Tukey biweight function (Beaton and Tukey, 1974). The function that is used for the estimation of these quantities is called \texttt{wvar()}, where the option \texttt{robust=TRUE} allows to make use of the robust estimator of WV. The output of this function is a vector of estimated WV of length \( J = \lfloor \log_2(T) \rfloor \), with \( T \) being the length of the observed time series and \( \lfloor x \rfloor \) representing the largest integer smaller than \( x \), along with the element-wise confidence intervals.

The WV is widely used for non-parametric analysis of time series, for example to interpret natural or economic phenomena across different time-scales (see Percival and Walden, 2006). In some engineering fields, such as navigation, a commonly used quantity is the Allan Variance (AV) which is directly related to the Haar WV (i.e. \( \text{AV} \equiv 2\text{WV} \)) and supports practitioners who want to identify certain behaviors in the observed time series. Indeed, as mentioned earlier, the WV well summarizes the information contained in the spectral density and, for example, a plot of the logarithmic transform of WV versus its scales can help to understand the kind of models that are generating the observed process since different models have different linear (or non-linear) behaviors when represented in this manner. Therefore, this visualization tool represents another helpful option to understand what kind of models could explain the time series, just like the basic AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) plots can help understand the order of certain ARMA models for example.

Considering the usefulness of these graphical representations, the \texttt{gmwm} package allows to easily generate these figures by applying the \texttt{plot()} function to the object issued from the \texttt{wvar()} function. An additional function which allows to graphically compare the WV issued from different time series or to compare the standard and robust WV on the same time series is \texttt{compare.wvar()}. This is a very useful tool if one is interested in
understanding if different time series can be explained by similar models or if the time series suffers from outliers or other forms of contamination. To provide some examples, the code below estimates the WV from some of the time series simulated in Section 4.2 and produces the plots in Figure 4.1.

```r
# Compute WV
wv.wn   = wvar(wn)
wv.qn   = wvar(qn)
wv.rw   = wvar(rw)
wv.dr   = wvar(dr)
wv.ar1  = wvar(ar1)
wv.sarima = wvar(sarima)

# Plot WV
compare.wvar(wv.wn, wv.qn, 
             split = F, auto.label.wvar = F, 
             legend.label = c("WN","QN"))
compare.wvar(wv.rw, wv.dr, 
             split = F, auto.label.wvar = F, 
             legend.label = c("RW","DR"))
plot(wv.ar1)
plot(wv.sarima)
```

A first detail which helps to interpret these graphs is that for stationary processes we generally have that the WV decreases at the larger scales while it increases for non-stationary ones. For example, a WV plot that decreases linearly with a certain slope indicates that the process could be white noise while a linearly increasing WV could indicate a random walk or a drift (depending on the slope of the WV). A plot of the WV which does not have a clear linear behavior could instead indicate the presence of a more complex process such as ARMA processes or a combination of different models. Indeed, models like white noise and quantization noise appear as negative linear functions of the scales of decomposition (which are denoted as \( \tau \) in the plots) with different slopes (i.e. steeper for the quantization noise) while drift and random walk models appear as positive linear functions where the drift has a steeper slope. The plots of the AR(1) and SARIMA processes show how these models don’t exactly have a linear behaviour in terms of WV. As can be observed, these plots by default give confidence intervals for the estimated WV which allow to understand if, for example, there is a significant difference between the standard and robust estimators of WV, thereby indicating that there could be contamination in the data. As an example, let us take the production index data made available, among others, in the astsa package and named prodn. Shumway and Stoffer (2013) suggest a SARIMA\((2,1,0)\times(0,1,3)_{12}\) as a good model for this data so, taking the appropriate differences, the output of the following commands given further on is shown in Figure 4.2.

---

3All the figures in this chapter are enhanced versions of the gmwm package plots.
4.3. Modeling based on the Wavelet Variance

Figure 4.1: Top-left: Estimated WV for white noise (orange line) and estimated WV for quantization noise (blue line) with confidence intervals (shaded areas). Top-right: Estimated WV for random walk (orange line) and estimated WV for drift (blue line) with confidence intervals (shaded areas). Bottom-left: Estimated WV for an AR(1) process with confidence intervals (shaded areas). Bottom-right: Estimated WV for an SARIMA(1, 0, 1) × (2, 1, 1)12 process with confidence intervals (shaded areas).
Chapter 4. The \texttt{gmwm} R package

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image}
\caption{Comparison between standard and robust WV for the \texttt{prodn} data.}
\end{figure}

\begin{verbatim}
# Take Non-seasonal and seasonal differences
prodn_diff = diff(diff(prodn), 12)

# Compute standard and robust WV
wv.prodn_diff = wvar(prodn_diff)
wv.prodn_diff.r = wvar(prodn_diff, robust = TRUE)

# Compare standard and robust WV
compare.wvar(wv.prodn_diff, wv.prodn_diff.r, split = FALSE)
\end{verbatim}

It can be seen how even though there are slight differences between the standard and robust WV estimates, the confidence intervals show that these differences do not appear to be significant and therefore a classical analysis appears to be appropriate for this data.

4.4 Model Estimation and Inference

The GMWM was initially developed for the estimation of so-called \textit{latent} or \textit{composite} processes in the context of sensor calibration. This estimator is defined as follows

$$\hat{\theta} = \arg\min_{\theta \in \Theta} (\hat{\nu} - \nu(\theta))^T \Omega (\hat{\nu} - \nu(\theta))$$  \hfill (4.4.1)

where $\theta$ represents the time series model parameter vector that we intend to estimate belonging to the compact set $\Theta$; $\hat{\nu} = [\nu^2_j]_{j=1,\ldots,J}$ represents the vector of estimated WV; $\nu(\theta) = [\nu^2_j(\theta)]_{j=1,\ldots,J}$ represents the WV implied by the model; $\Omega$ is a positive definite weighting matrix chosen in a suitable manner (see Guerrier et al., 2013b). The general estimation function which implements this estimator is called \texttt{gmwm()} for which different options can be specified such as, for example, the method to estimate the weighting matrix $\Omega$. However, the main arguments for this function are of course the observed time series as well as the model that we intend to estimate (see Section 4.2).
4.4. Model Estimation and Inference

4.4.1 The class of SARIMA models

The class of SARIMA models includes many of the basic models that can be estimated within the gmwm package, with a few exceptions such as quantization noise. This is possible based on the results in N. F. Zhang (2008) where analytic expressions for $\nu(\theta)$ can be obtained for all the models listed in Section 4.2. Given this, the package therefore delivers an additional tool to estimate SARIMA models in R, with a possible advantage residing in the fact that it can easily estimate these models even for considerably large sample sizes. To see how these models can be estimated with the gmwm() function, let us take the prodn data whose WV we analysed at the end of Section 4.3. As mentioned earlier, Shumway and Stoffer (2013) suggest a SARIMA($2,1,0) \times (0,1,3)_{12}$ model for this data so let us estimate this using also the standard arima() function for Maximum Likelihood Estimation (MLE) for comparison. The code for these two procedures can be found below.

# MLE
mle.fit = arima(prodn,  
    order = c(2,1,0),  
    seasonal = list(order = c(0,1,3), period = 12),  
    include.mean = FALSE)

# GMWM
gmwm.fit = gmwm(SARIMA(ar = 2, i = 1, ma = 0, sar = 0, si = 1, sua = 3),  
    data = prodn)

# View GMWM Fit
plot(gmwm.fit)

The SARIMA model is specified within the gmwm() function by using the syntax SARIMA() in which the orders are defined and its result is stored in an object called gmwm.fit. The results of the arima() and gmwm() estimation functions, along with their confidence intervals, are given in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>GMWM</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>$2.972 \cdot 10^{-1}$</td>
<td>[1.875 \cdot 10^{-1}, 3.913 \cdot 10^{-1}]</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>$1.760 \cdot 10^{-1}$</td>
<td>[6.236 \cdot 10^{-2}, 2.880 \cdot 10^{-1}]</td>
</tr>
<tr>
<td>$\Theta_1$</td>
<td>$-8.872 \cdot 10^{-1}$</td>
<td>[-1.045 \cdot 10^{0}, -6.207 \cdot 10^{-1}]</td>
</tr>
<tr>
<td>$\Theta_2$</td>
<td>$-2.764 \cdot 10^{-2}$</td>
<td>[-5.344 \cdot 10^{-1}, 6.637 \cdot 10^{-1}]</td>
</tr>
<tr>
<td>$\Theta_3$</td>
<td>$2.864 \cdot 10^{-1}$</td>
<td>[-3.386 \cdot 10^{-1}, 8.071 \cdot 10^{-1}]</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$1.154 \cdot 10^{0}$</td>
<td>[9.060 \cdot 10^{-1}, 1.238 \cdot 10^{0}]</td>
</tr>
</tbody>
</table>

Table 4.1: MLE and GMWM estimated parameters for the SARIMA model on the prodn data. $\phi_i$ represents the $i^{th}$ autoregressive parameter, $\Theta_i$ represents the $i^{th}$ seasonal moving-average and $\sigma^2$ represents the innovation variance.

There appear to be no significant differences between the two methods, although the last two seasonal moving-average parameters do not appear to be significantly different from zero for the GMWM. Once the model has been estimated and saved into an object (i.e. gmwm.fit in the above code), it is possible to make a visual assessment of how
well the model fits the observed time series by comparing the estimated WV with the
WV implied by the estimated parameters simply by applying the function plot() to the
estimated model. This can be seen in Figure 4.3 where the orange line, representing
the WV implied by the estimated SARIMA model, closely follows the WV estimated directly
on the prodn data (blue line) and lies within its confidence intervals (shaded area).

4.4.2 Linear State-Space Models

As mentioned in Section 4.2, the gmwm package is able to estimate all those linear state-
space models which can be represented as the sum of different underlying processes. These
models extend from a wide variety of linear state-space models to structural time series
models. An example of these models is given by the process \( (Y_t) \) defined as follows

\[
X_t = X_{t-1} + \delta + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)
\]

\[
Y_t = X_t + U_t, \quad U_t \sim \mathcal{N}(0, \sigma^2)
\]

where \( \delta \) represents a drift. This state-space model, also known as a local-linear trend
model, is often used in ecology, for example, to describe population dynamics with measure-
ment errors and basically corresponds to the sum of a white noise, a random walk
and a drift (an example of how this model is estimated within the package is provided
further on). The reason why the GMWM manages to estimate these kind of models in a
computationally efficient manner is because their WV is simply the result of the sum of
individual WV of the underlying models whose analytic expressions are known (see Sec-
tion 4.4.1) and the number of scales (i.e. the number of auxiliary parameters) is always
reasonable even for large sample sizes. Therefore many models can be summed together
to deliver different types of latent models. There are however some restrictions on certain
model specifications due to parameter identifiability issues as emphasized in Chapter 2.
For example, we can have a \( \text{WN}() \), a \( \text{QN}() \), a \( \text{RW}() \) and a \( \text{DR}() \) only once in a state-space
model whereas an \( \text{ARMA}() \) or an \( \text{AR1}() \) can possibly be included as many times as desired.
It must be specified however that it is not certain that a sum of ARMA models is identifi-

Figure 4.3: Estimated WV from the prodn data (blue line) with confidence intervals
(shaded area) and WV implied by estimated SARIMA model (orange line)
able while a sum of AR(1) models, which well approximates many processes, is generally identifiable.

Let us now give a simple example to illustrate how the \texttt{gmwm} package estimates these types of models and therefore consider the process \((Y_t)\) defined as follows

\[
X_t = \phi_1 X_{t-1} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \nu^2),
\]

\[
W_t = \phi_2 W_{t-1} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),
\]

\[
Y_t = X_t + W_t + Z_t \quad (4.4.2)
\]

where \(Z_t = \delta\) is a drift process. The process \((Y_t)\) is hence the result of the sum of two first-order autoregressive processes and a drift process. This is another example of a state-space model where the states are represented by the unobserved processes \((X_t), (W_t)\) and \((Z_t)\). For this model, let us moreover define the parameter values as follows: \(\phi_1 = 0.9, \nu^2 = 1, \phi_2 = 0.1, \sigma^2 = 4\) and \(\delta = 0.01\). The code below simulates from this model using the \texttt{gen.lts()} function mentioned in Section 4.2, consequently plotting the breakdown of this state-space model (left plot of Figure 4.4) and then estimating its parameters using the \texttt{gmwm()} function.

```r
# Length of time series
n = 1e3

# Define model
true.model = AR1(0.9,1) + AR1(0.1,4) + DR(0.01)

# Simulate and plot breakdown of time series
set.seed(1337)
sim.ts = gen.lts(true.model, n)
plot(sim.ts)

# Estimate model
model = 2*AR1() + DR()
fit = gmwm(model, sim.ts)

# Visualize WV of estimated model
plot(fit, process.decomp = TRUE)

# Model inference
inference = summary(fit, inference = TRUE, bs.gof = TRUE)
```

As highlighted Section 4.4.1, the object containing the result of the estimation procedure (i.e. \texttt{fit} in the example code above) can be used within the \texttt{plot()} function and, in particular, if specifying the option \texttt{"process.decomp = TRUE"} within this function, the user will obtain a plot similar to the one presented in the right plot of Figure 4.4. The blue line represents the estimated WV while the orange one represents the WV implied by the estimated model. The other lines are the result of specifying the option \texttt{"process.decomp = TRUE"} and show the WV implied by the estimated parameters for the individual models composing the overall sum. It can be seen how the individual WV help to “support” the overall implied WV in order to fit the estimated WV and, in this case, the fitted model appears to be a good one since it closely follows the estimated WV and lies within its
confidence intervals (shaded area). This is confirmed by the Goodness-of-Fit (GoF) test contained in the inference object generated by executing the command `summary(fit, inference = TRUE, bs.gof = TRUE)` whose output can be found below.

**Model Information:**

<table>
<thead>
<tr>
<th></th>
<th>Estimates CI Low</th>
<th>CI High</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.8808235</td>
<td>0.800135885</td>
<td>0.94028587</td>
</tr>
<tr>
<td>SIGMA2</td>
<td>0.7454827</td>
<td>0.216068102</td>
<td>1.46524184</td>
</tr>
<tr>
<td>AR1</td>
<td>0.2025861</td>
<td>0.075190570</td>
<td>0.30140290</td>
</tr>
<tr>
<td>SIGMA2</td>
<td>4.5995931</td>
<td>3.659867741</td>
<td>5.42981374</td>
</tr>
<tr>
<td>DR</td>
<td>0.0095902</td>
<td>0.005945415</td>
<td>0.01409644</td>
</tr>
</tbody>
</table>

**Objective Function:** 0.0012

**Bootstrapped Goodness of Fit:**

Test Statistic: 0

P-Value: 0.97 CI: (0.93, 1)

To replicate the results, use seed: 1337

It can be seen how the result of this procedure is a table containing the estimated parameters, their confidence intervals and, as already mentioned, the details on the GoF test which is based on the Sargan-Hansen test (also known as “J-test”, see Sargan, 1958). For the latter, it is possible to specify the option `bs.gof = TRUE` which means that the distribution of the test-statistic (i.e. the value of the GMWM objective function at the estimated parameter) is approximated by parametric bootstrap

4NB: The GoF test statistic and p-value in the output are rounded to two decimal places.

Let us now use the gmwm package on a real dataset containing measurements on the annual flow of the Nile river at Ashwan between 1871 and 1970 available in R within the datasets package. Petris and Petrone (2011) suggest two structural models for this dataset:

- **Local level model:** this model is given by the sum of a white noise model and a random walk;

- **Local linear trend:** this model is given by the sum of a local level model and a drift (mentioned as an example at the beginning of this section).

These models can both be estimated with the gmwm package as shown in the code further on. The objects `loc.level` and `loc.trend` contain the estimations of the local level and local linear trend models respectively. In addition to the estimation of these state-space (structural) models, the gmwm package allows to perform a GoF test, as seen earlier in the simulated example, as well as to select the “best” model based on the mentioned the WIC. The latter criterion tells us in very general terms how well the estimated model can predict the values of the WV issued from another realization of the same process and is defined as follows:

\[
\text{WIC} = \mathbb{E} \left[ \mathbb{E}_0 \left( \nu_0 - \nu(\hat{\theta}) \right)^T \Omega \left( \nu_0 - \nu(\hat{\theta}) \right) \right] \tag{4.4.3}
\]
where $E_0[\cdot]$ and $\hat{\nu}_0$ denote respectively the expectation and the estimated WV based on another realization of the same process. Its corresponding estimator is given by

$$
\hat{WIC} = \left( \hat{\nu} - \nu(\hat{\theta}) \right)^T \Omega \left( \hat{\nu} - \nu(\hat{\theta}) \right) + 2 \text{tr} \left\{ \text{cov} \left[ \hat{\nu}, \nu(\hat{\theta}) \right] \Omega^T \right\}
$$

(4.4.4)

where $\text{cov}(\cdot)$ denotes a consistent estimator of the covariance. The first term in the estimator corresponds to the value of the objective function of the GMWM at the estimated parameter $\hat{\theta}$ while the second term is sometimes referred to as “optimism”. Indeed, the first term is expected to decrease as parameters are added while the second is expected to increase in these cases, thereby providing some kind of penalty for overfitting. This estimator is computed within the function `rank.models()` where the user can provide a set of nested candidate models, as well as the dataset, to finally obtain a list of models ranked in ascending order according to the WIC. Moreover, it is possible to specify how to compute $\text{cov}(\cdot)$ in the second term, which is either through parametric bootstrap (see Guerrier et al., 2015) or through an analytic form (see X. Zhang and Guerrier, 2015). Since the aim is to minimize this quantity, the first model in the list can be considered as the one which best predicts the WV and therefore, generally speaking, the time series itself.

---

5The models for which the WIC can be computed through its analytical form are the models (and combination thereof): WN(), QN(), RW(), DR(), AR1(), MA1() and ARMA11().
# Retrieve Nile dataset
nile = datasets::Nile

# Estimate the models
loc.level = gmwm(WN() + RW(), nile)
loc.trend = gmwm(WN() + RW() + DR(), nile)

# Goodness-of-fit of the models
> summary(loc.level, inference = TRUE, bs.gof = TRUE)

Model Information:

<table>
<thead>
<tr>
<th>Estimates</th>
<th>CI Low</th>
<th>CI High</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN</td>
<td>13611.69</td>
<td>9133.942</td>
<td>17598.004</td>
</tr>
<tr>
<td>RW</td>
<td>2095.55</td>
<td>167.214</td>
<td>3787.207</td>
</tr>
</tbody>
</table>

Objective Function: 0.0288

Bootstrapped Goodness of Fit:

Test Statistic: 0.03
P-Value: 0.24 CI: (0.16, 0.33)

To replicate the results, use seed: 1337

> summary(loc.trend, inference = TRUE, bs.gof = TRUE)

Model Information:

<table>
<thead>
<tr>
<th>Estimates</th>
<th>CI Low</th>
<th>CI High</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN</td>
<td>13613.98e+04</td>
<td>9135.6649579</td>
<td>1.759593e+04</td>
</tr>
<tr>
<td>RW</td>
<td>2.092888e+03</td>
<td>183.5483823</td>
<td>3.767013e+03</td>
</tr>
<tr>
<td>DR</td>
<td>4.928475e-01</td>
<td>0.4897141</td>
<td>4.946456e-01</td>
</tr>
</tbody>
</table>

Objective Function: 0.0289

Bootstrapped Goodness of Fit:

Test Statistic: 0.03
P-Value: 0.26 CI: (0.18, 0.35)

To replicate the results, use seed: 1337

# Select model
WIC = rank.models(RW() + WN(), RW() + WN() + DR(), data = nile)
> WIC

The model ranking is given as:

<table>
<thead>
<tr>
<th>Obj Fun</th>
<th>Optimism</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. RW WN</td>
<td>0.0288</td>
<td>0.9289</td>
</tr>
<tr>
<td>2. RW WN DR</td>
<td>0.0289</td>
<td>0.9314</td>
</tr>
</tbody>
</table>

plot(WIC)
4.4. Model Estimation and Inference

Figure 4.5: Left: Plot of Annual Nile river flow from 1871 - 1970. Right: WV implied by the selected local level model (orange line) and estimated WV (blue line) with confidence intervals (shaded area).

As can be observed, the two models appear to be close in terms of the WIC which is given in the last column (the first two columns are the first and second term of (4.4.4) respectively). However, following the rule of picking the model with the lowest WIC, this would be the local-level model and, if applying the function plot() to the object issued from the rank.models() function (i.e. plot(WIC)), it is possible to assess how well the selected model fits the observed WV. In this case Figure 4.5 shows that the WV implied by the selected model fits the observed WV well at the first scales and less so at the last, remaining however strictly within the confidence intervals.

4.4.3 Robust Inference for Time Series Models

There has been a large amount of research dedicated to providing sound statistical methods to deliver robust estimation and inference for time series models (see Maronna et al., 2006). However, as mentioned in the introduction, the implementation of these methods has been scarce (if not nonexistent) and researchers and practitioners have not been able to make use of these approaches which are often complicated to implement and numerically unstable. The gmwm package offers a first stable solution to this gap in the available software by simply replacing the estimated WV $\hat{\nu}$ in (4.4.1) with the robust M-estimator proposed in Chapter 1 which has well defined asymptotic properties as well as better finite sample performance compared to existing estimators. To perform a robust estimation of time series model parameters within the package, all that is needed is to specify the option robust = TRUE within the gmwm() function.

As highlighted at the end of Section 4.3, a first step to understanding if a robust analysis is needed, for example, is to compare the standard estimated WV with its robust counterpart and check if their confidence intervals overlap. If there are problems of robustness, these should be more evident at the first scales since the variability at the last scales often does not allow to detect a significant difference between the standard and robust WV. To illustrate this procedure, let us take another example where we simulate from an AR1()+WN() model and then add noise coming from a normal random variable with variance equal to 100 to 1% of the observations. The code below shows how this is done and then computes the standard and robust WV on the resulting time series.
# Specify model
ttrue.model = AR1(\(\phi = .99\), \(\sigma^2 = .01\)) + WN(\(\sigma^2 = 1\))

# Generate time series
set.seed(213)
n = 1e3
sim.ts = gen.gts(true.model, n)

# Contaminate time series
cont.eps = 0.01
cont.num = sample(1:n,round(n*cont.eps))
sim.ts[cont.num] = sim.ts[cont.num] + rnorm(round(n*cont.eps),0,sqrt(100))

# Compute standard and robust WV
wv.classic = wvar(sim.ts)
wv.robust = wvar(sim.ts, robust = TRUE)

# Plot the robust WV
plot(sim.ts)
compare.wvar(wv.classic, wv.robust, split = FALSE)

# Run robust estimation
rob.fit = gmwm(AR1()+WN(), sim.ts, robust = TRUE)

# Run inference
summary(rob.fit, inference = TRUE, bs.gof = TRUE)

Model Information:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>CI Low</th>
<th>CI High</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.98502549</td>
<td>0.96729698</td>
<td>0.99601635</td>
</tr>
<tr>
<td>SIGMA2</td>
<td>0.01433612</td>
<td>0.00629577</td>
<td>0.02184096</td>
</tr>
<tr>
<td>WN</td>
<td>0.95763144</td>
<td>0.86667884</td>
<td>1.03607348</td>
</tr>
</tbody>
</table>

Objective Function: 0.0285

Bootstrapped Goodness of Fit:

Test Statistic: 0.03
P-Value: 0.65 CI: (0.55, 0.74)

To replicate the results, use seed: 1337

The estimates of classic and robust WV, contained respectively in the objects \texttt{wv.classic} and \texttt{wv.robust}, can be compared using the function \texttt{compare.wvar()} seen earlier, using the option \texttt{split = FALSE} to see the two estimates superposed. Figure 4.6 shows the result of this plotting function along with the plot of the simulated time series, highlighting how there appears to be contamination, confirmed by that fact that the two WV differ
significantly across the first scales. Based on this analysis, a robust estimation procedure would appear necessary and for this purpose, as mentioned earlier, we use the function \( \text{gmwm}() \) with the option `robust = TRUE`. It must be emphasized that, when estimating the WV or model parameters robustly, it is possible to specify the desired level of efficiency of the robust estimator with respect to the standard WV and GMWM estimators. The default value is \( \text{eff} = 0.6 \), meaning that we want a low level of efficiency to obtain more robust estimates. However, the user can specify other values, usually between 0.5 and 1, the last value simply implying that the estimator corresponds to the standard estimator of WV (i.e. non-robust). It can be seen how the estimated parameters from the robust fit are close to the true simulation values despite the contamination in the time series.

Let us now consider a real data example which was already analyzed in Chapter 3 and is given by the monthly precipitation series from 1907 to 1972 used in Hipel and McLeod (1994) which was obtained from DataMarket\(^6\) and made available within the `datapkg` package\(^7\). In hydrology the study of water cycles is an extremely relevant topic and different models are used to easily interpret these cycles, such as the Environmental System Model of a watershed. The AR(1) model is among the candidates for the precipitation phase so let us estimate this model for this data (hereinafter `hydro`) using the code below.

```r
# Check if there are robustness issues
wv = wvar(hydro)
rob.wv = wvar(hydro, robust = TRUE)
compare.wvar(wv, rob.wv, split=FALSE)

# Compare robust and non-robust fits
mle.fit = arima(hydro, order=c(1,0,0))
gmwm.fit = gmwm(AR1(), hydro)
rob.fit = gmwm(AR1(), hydro, robust=TRUE)
```

The plot comparing the classic and robust WV (shown in Figure 4.7) indicates that there could be some contamination in the observations so we can compare the estimates.

---

\(^6\)https://datamarket.com/data/set/22w1/mean-monthly-precipitation-1907-1972

\(^7\)The `datapkg` package can be installed using `gmwm::install_datapkg()` and the source code is available at https://github.com/smac-group/datapkg
of the MLE, GMWM and RGMWM to understand the influence of this contamination on the estimations. These are given in Table 4.2 where $\phi$ represents the AR(1) parameter and $\sigma^2$ represents the innovation variance. A significant difference can be observed between the non-robust estimators (i.e. MLE and GMWM) and the RGMWM suggesting that, even if the model is not the correct one, a robust estimation procedure is preferable for this dataset.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\phi}$</th>
<th>$\hat{\sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>$6.843 \cdot 10^{-2}$</td>
<td>$2.199 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[1.110 \cdot 10^{-3}, 1.212 \cdot 10^{-1}]$</td>
<td>$[2.003 \cdot 10^{-1}, 2.3929 \cdot 10^{-1}]$</td>
</tr>
<tr>
<td>GMWM</td>
<td>$5.577 \cdot 10^{-2}$</td>
<td>$2.179 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[-4.357 \cdot 10^{-3}, 1.153 \cdot 10^{-1}]$</td>
<td>$[1.985 \cdot 10^{-1}, 2.365 \cdot 10^{-1}]$</td>
</tr>
<tr>
<td>RGMWM</td>
<td>$4.049 \cdot 10^{-1}$</td>
<td>$1.066 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>CI</td>
<td>$[3.345 \cdot 10^{-1}, 4.662 \cdot 10^{-1}]$</td>
<td>$[9.623 \cdot 10^{-2}, 1.175 \cdot 10^{-1}]$</td>
</tr>
</tbody>
</table>

Table 4.2: Estimated parameters for an AR1() model on the hydro dataset.

### 4.5 Computational Efficiency

The gmwm package has been developed and made available within the R programming language across Windows, Mac (OS X), and Linux with support for Solaris to arrive soon. Even though the framework’s main residency is within R, the primary functions have been entirely implemented within C++ to obtain a high level of computational efficiency and to enable the ease of porting key functions to alternative computational frameworks. Moreover, the implemented functions use the Armadillo C++ Matrix Library API, which has a syntax that is very similar to Octave and R. As a result, the ability to extend the C++ library for users in both the engineering and statistical domains has a lower threshold.
Part of the computational efficiency afforded to the package is due to the ability to interface between R and C++ with ease. The creation of this interface is through the considerable use of both Rcpp (Eddelbuettel et al., 2011) and RcppArmadillo (Eddelbuettel and C. Sanderson, 2014) to blend R data types (SEXP) with the GMWM C++ library. Particular attention has been placed on the use of this interface to avoid the trap of repeated calls from R into C++ thereby leading to excessive object copies during data type conversions and processing slowdowns. To reduce overall dependency on R and to ensure computational efficiency, some functions written in R were rewritten in C++ using Armadillo. The implementations are available in both the gmwm package and in the “R to Armadillo” repository\(^8\). The rewritten implementations typically yield an increase in computational speed between 1.5 and 4 times faster and also serve as well documented examples of porting R code into C++ using Rcpp.

Having discussed the software framework, in this section we briefly underline the computationally efficiency of the gmwm package in estimating the WV and the models mentioned in Sections 4.4.1 and 4.4.2. To provide a first example of the computational speed of the functions in the gmwm package, let us compare the speed at which the WV is computed using the wave.variance() function within the waveslim package and the new wvar() function in the gmwm package, including the robust option. Table 4.3 collects the results of one hundred benchmark tests using the rbenchmark package.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Package</th>
<th>100</th>
<th>1,000</th>
<th>10,000</th>
<th>100,000</th>
<th>1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>waveslim</td>
<td>0.0004</td>
<td>0.0013</td>
<td>0.0086</td>
<td>0.0823</td>
<td>0.7432</td>
<td></td>
</tr>
<tr>
<td>gmwm</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0011</td>
<td>0.0223</td>
<td>0.3575</td>
<td></td>
</tr>
<tr>
<td>gmwm (robust)</td>
<td>0.0069</td>
<td>0.0125</td>
<td>0.0530</td>
<td>0.6531</td>
<td>9.6700</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Benchmarks of the WV computation in seconds.

It can be observed how the new function wvar() considerably increases the speed at which the WV is computed compared to the already existing wave.variance() function, going at least twice as faster than the latter which is of considerable importance when carrying out analyses on large samples. On the other hand, the robust estimation is visibly slower, as would be expected, preserving however reasonable computational times even in large samples.

Let us now consider the computational times when estimating two models:

- **ARMA(3,1)**: autoregressive moving average model;
- **2*AR1()+WN()**: a latent model made by the sum of two AR(1) processes and a white noise.

We estimate the parameters of these models 10 times with the following estimators:

- **MLE**;
- **GMWM**;
- **Robust GMWM (RGMWM)** presented in Section 4.4.3;

---

\(^8\)R to Armadillo source is available at: [https://github.com/coatless/r-to-armadillo](https://github.com/coatless/r-to-armadillo)

For the ARMA(3,1) model, only the robust estimators were compared (i.e. RGMWM and KUNSCH) while all of them were compared for the latent model. Therefore each GMWM estimator is compared with an alternative that is available (or easily implemented) in R. In all cases, the optimization procedures used the true parameter values as starting values. Figure 4.8 shows the median computation times for these estimators on the two models for sample sizes going from 100 to 10 million, with a roof for computation placed at 6 hours.

It can be seen how the GMWM and RGMWM estimators have extremely low computational times compared to the other estimators in the two cases. Indeed, the computational times for the KUNSCH estimator are beyond the chosen maximum time limit for sample sizes larger than 100,000 while the RGMWM remains under 1 hour even for samples of size 10 million. Similar conclusions can be made for the latent model where the GMWM estimators are clearly faster than the other two methods in all cases. The RGMWM is slower than the GMWM as would be expected but it is able to estimate the latent model in a little over two minutes for sample sizes of 10 million. These results show that not only the gmwm package is able to consistently estimate a wide range of time series models, but it is also able to do so robustly and in an extremely computationally efficient way even for very large sample sizes.

4.6 Conclusion

The implementation of the GMWM framework within the gmwm R package represents a first step in the direction of a platform which delivers a wide set of computationally efficient tools for (robust) time series model estimation and inference. Aside from the many advantages highlighted in this chapter, the main contributions that this package delivers are the ease with which linear state-space models can be estimated and the tools for the robust estimation of a wide range of time series models. These two aspects considerably
reduce the barriers to which users were subject to due to the unavailability, complexity and/or high computational demand of existing statistical software. In addition, the nature of this estimator and the idea of latent model structures allows it to be extended to different settings.
Chapter 5

A Computationally Efficient Platform for Inertial Sensor Calibration

The calibration of (low-cost) inertial sensors has become increasingly important over the past years since their use has grown exponentially in many applications going from unmanned aerial vehicle navigation to 3D-animation. However, this calibration procedure is often quite problematic since the signals issued from these sensors have a complex spectral structure and the methods available to estimate the parameters of these models are either unstable, computationally intensive and/or statistically inconsistent. This chapter presents a new software platform for inertial sensor calibration based on the Generalized Method of Wavelet Moments which provides a computationally efficient, flexible, user-friendly and statistically sound tool to estimate and select from a wide range of complex models. In addition, all this is possible also in a robust framework allowing to perform sensor calibration when the data is affected by outliers. The software is developed within the open-source statistical software \texttt{R} and is based on \texttt{C++} language allowing it to achieve high computational performance.

5.1 Introduction

An inertial measurement unit (IMU) is a device whose composition is typically a triad of accelerometers and gyroscopes, which provide a measurement of the specific force of acceleration as well as the rate of angular movement. With the advances in microtechnology, the premise of a low-cost miniaturized IMU is possible through the construction of a Micro-Electro-Mechanical System (MEMS) device, which is a compact and light version of an IMU (El-Sheimy and Niu, 2007). As a result of low cost, light weight, and lower energy consumption, the number of applications privy to inertial sensors have increased tremendously, leading to a considerable amount of attention being afforded to device performance by researchers. These developments have thus enabled inertial sensors to be embedded in countless consumer-facing goods that range from smart phones to exercise equipment as autonomous automobiles, while also finding further use in military applications such as unmanned aerial vehicles (Woodman, 2007).

Despite the considerable amount of applications that MEMS sensors are used for, their small size and low cost compared to the higher grade IMU comes at the price of fairly large and variable measurement errors. As a result, there is a considerable amount of research that solely focuses on modelling and compensating for these errors by attempting to characterize their complex noise structure. To do so, there are different methods which
have been regularly used to date, albeit being affected by different drawbacks. The main and most statistically sound of the latter approaches is represented by the Maximum Likelihood (ML) which has been widely studied and used for the task of inertial sensor calibration. In Stebler et al. (2011), the ML was implemented through the use of the Expectation-Maximization (EM) algorithm employed alongside a Kalman filter to attempt the parameter estimation of a State-Space Model (SSM). Although this procedure works well in simple SSM settings, it tends to diverge in only moderately more complex scenarios which are often more realistic for the characterization of inertial sensor error signals. A similar approach was used in Zaho et al. (2011) and Nikolic et al. (2016) to improve the results by implementing, respectively, a nonlinear adaptive Kalman Filter (KF) and a log-sampling scheme without however overcoming the mentioned convergence and numerical instability issues. Moreover, the ML can be time consuming to implement and may require considerable effort to estimate a new model (see Stebler et al., 2011). To mainly tackle the latter obstacle, another existing method is the “graphical” Allan Variance Linear Regression (AVLR) which makes use of the slope in the linear regions of a log-log plot of the Allan variance (AV) to estimate the parameters of the underlying processes. For this method, Woodman (2007) and El-Sheimy et al. (2008) provide a thorough overview of the approach describing how to distinguish the different underlying processes via graphical representations of the AV. Nevertheless, although still being widely used, Guerrier et al. (2016a) formally proved that this method is inefficient and inconsistent in the vast majority of cases. Considering this, a different use of the AV was presented in Vaccaro and Zaki (2012) where the parameters of a process composed by a white noise and a random walk were simultaneously estimated in a method of moments fashion.

Given the limitations of existing techniques, a new estimation method was proposed by Guerrier et al. (2013b) called the Generalized Method of Wavelet Moments (GMWM). Based on the idea of Generalized Method of Moments (GMM) estimators, the GMWM makes use of the relation between the Wavelet Variance (WV) and the parameters of a process, estimating the latter by minimizing the distance between the empirical and model-based WV (see Section 5.2). It must be noted that this method could also be based on the AV since the Haar WV is simply twice the AV. Having said this, in a way this estimator can be viewed as a generalized version of Vaccaro and Zaki (2012) with additional benefits. Among these advantages we can highlight its good statistical properties, its ease of implementation and its computational efficiency. Moreover, the nature of this estimator has allowed to deliver additional statistical tools which are of particular interest for the purpose of inertial sensor calibration. For instance, the development of a model selection criterion to effectively rank models while taking into account their complexity. In addition, there exists a robust estimation version of the GMWM that is applicable when dealing with data that has suffers from contamination (e.g. outliers or influential points). Additional research is being developed to extend this estimation framework to multivariate and non-stationary (dynamic) settings which are also of interest for sensor calibration.

Considering these results and properties, this chapter introduces the software platform which allows practitioners and researchers to have wide access to all the tools linked to the GMWM, going from the calculation of the WV to the estimation of parameters and selection of simple and complex error models also in the presence of outliers. In particular, this chapter gives details on how these approaches have been implemented through specific algorithms in order for them to be fully tailored to the features which characterize IMU measurement error signals. The platform in which these procedures are
delivered can be found in the newly developed “gmwm” package within the R open source statistical software and makes use of the C++ language to ensure higher computational efficiency (see Appendix D.1 for some details). As described in Section 5.3, developing an R package not only allows anyone to access the methodologies and functions of the GMWM but also provides a platform which is efficient and can easily be brought up to date with the most recent developments.

The chapter is organized as follows: in Section 5.2 we give a brief overview of the GMWM and its related existing tools while in Section 5.3 we introduce the software platform by presenting its main features in the relative subsections going from model identification to model selection. Finally, Section 5.4 concludes by giving an overview of the upcoming developments of the GMWM which will be included in future software updates to deliver additional important tools for inertial sensor calibration.

5.2 The Generalized Method of Wavelet Moments

In this section, we provide a brief description of the GMWM estimator along with an overview of the main tools that have been derived and are implemented within the package (for a more thorough description the methods and their properties refer to Guerrier et al., 2015, 2013b, and Chapter 3). As mentioned in the previous section, the GMWM relies on the WV, which is the variance of the wavelet coefficients \( W_{j,t} \), issued from a wavelet decomposition of a signal (see, for example, Percival and Walden, 2006, for an overview). In the context of the GMWM, this decomposition is currently based solely on the Haar wavelet and therefore all the results in this chapter rely on this type of filter. Furthermore, the platform is able to provide both the Discrete Wavelet Transform (DWT) and the Maximum Overlap Discrete Wavelet Transform (MODWT) WV estimators, with the latter being more efficient. For this reason, the MODWT WV estimator \( \hat{\nu}^2_j \) is the default method which, based on the proposal of Percival (1995), is defined as

\[
\hat{\nu}^2_j = \frac{1}{T-L_j+1} \sum_{t=L_j}^{T} W_{j,t}^2.
\]

(5.2.1)

where \( T \) represents the length of the signal, \( j = 1, \ldots, J \) (with \( J = \lfloor \log_2(T) \rfloor \), where \( \lfloor \cdot \rfloor \) denotes the largest integer not greater than the corresponding argument) represents the scale of decomposition and \( L_j \) is the length of the (Haar) wavelet filter at scale \( j \).

Having presented the WV estimator, we can now briefly describe the WV implied by a stochastic process \( F_\theta \) where \( \theta \in \Theta \subseteq \mathbb{R}^p \) represents the parameter vector which defines the model. Denoting \( S_{F_\theta}(f) \) as being the Power Spectral Density (PSD) of the process \( (Y_t) \), this parameter vector is mapped to the WV through the following relationship

\[
\nu^2_j(\theta) = \int_{-1/2}^{1/2} S_{W_j}(f) df = \int_{-1/2}^{1/2} |H_j(f)|^2 S_{F_\theta}(f) df.
\]

(5.2.2)

where \( H_j \) is the transfer function of the wavelet filters at scale \( j \) and \( |\cdot| \) denotes the modulus operator. We can therefore see that the WV is a function of the model parameters thereby providing a theoretical quantity which can be matched with its estimated counterpart. The GMWM is based exactly on the latter idea by attempting to inverse the relation between the WV and the parameters to obtain \( \theta(\hat{\nu}) \), where \( \hat{\nu} = [\nu^2_j]_{j=1,\ldots,J} \). To do so, the GMWM estimator \( \hat{\theta} \) is defined as

\[
\hat{\theta} = \arg\min_{\theta \in \Theta} \langle \hat{\nu} - \nu(\theta) \rangle^T \Omega (\hat{\nu} - \nu(\theta)),
\]

(5.2.3)
which therefore is the result of a generalized least squares minimization where $\Omega$ represents a positive definite weighting matrix chosen in a suitable manner (see e.g. Guerrier et al., 2013b). Conditioned on the positive definiteness of this matrix and denoting the covariance matrix of $\hat{\nu}$ as $V$, the main role of $\Omega$ is to make the GMWM estimator as efficient as possible and it was shown that $\Omega = V^{-1}$ is the matrix which allows the GMWM estimator to achieve maximum asymptotic efficiency.

In Guerrier et al. (2013b) the large sample properties of the GMWM were studied showing that it is consistent and normally distributed and these results are reinforced by the results in Chapter 2 showing identifiability for a wide range of time series models. A robust version of the GMWM was also given in Chapter 3 allowing to limit the influence of contamination in the observed signal on the estimation process while preserving suitable asymptotic properties. Using these properties, additional useful results were derived for inference purposes allowing to obtain, for example, confidence intervals for the parameters, goodness-of-fit of the estimated models and model selection criteria.

While the parameter confidence intervals are simply based on the asymptotic normality of the GMWM, the goodness-of-fit test is based on the test for over-identifying restrictions for GMM-type estimators (also known as J-test). The test statistic is given by

$$ T \left( \hat{\nu} - \nu(\hat{\theta}) \right)^T \Omega^* \left( \hat{\nu} - \nu(\hat{\theta}) \right) \xrightarrow{D} \chi^2_{J-p} $$(5.2.4)

where the $\Omega^*$ denotes an appropriate weighting matrix whose form can be found in Hansen (1982). The test statistic, as can be seen, follows a $\chi^2$ distribution under the null hypothesis $H_0$ which generally states that the difference between the empirical and the theoretical WV is zero. Hence, if the null hypothesis is not rejected, it implies that there is not enough evidence to state that the estimated model does not fit the signal well. Given this setting, another useful development of the GMWM is the proposal of a selection criterion in Guerrier et al. (2015) called the Wavelet Information Criterion (WIC) whose estimator is defined as follows

$$ \text{WIC} = E \left[ \left( \hat{\nu} - \nu(\hat{\theta}) \right)^T \Omega \left( \hat{\nu} - \nu(\hat{\theta}) \right) \right] $$

+ $2 \text{tr} \left[ \text{cov} \left[ \hat{\nu}, \Omega \nu(\hat{\theta}) \right]^T \right]$ (5.2.5)

where $\text{tr} [\cdot]$ denotes the trace operator. This estimator is an unbiased estimator of the “prediction” error made by using the estimated parameters from one signal to predict the WV on another signal from the same data-generating process. Term $A$ could be called “apparent loss” and is a measure of how well the model fits the observed signal. The latter typically diminishes as the model complexity increases (e.g. adding more underlying models and parameters to the composite process of interest) while term $B$ does the opposite and therefore grows as the model complexity increases. The latter term could be called “optimism” and acts as a complexity-based penalty. There are different manners to compute this term such as parametric bootstrap (see Efron, 2004; Guerrier et al., 2015) or using its asymptotic approximation given in X. Zhang and Guerrier (2015). Considering these terms, the goal would be to select the model with the smallest WIC value, meaning that we are selecting the model with the smallest estimated prediction error for the WV.

To summarize, the GMWM provides a set of extremely useful tools for the task of modelling the purely stochastic errors of inertial sensors for calibration purposes. Based
5.3 Implementation of the GMWM

The new software platform is developed and made available within the R statistical software which is an open source implementation made available by the R Foundation Team (2014). In this framework, the new platform is implemented within the R package called “gmwm” whose main functions have been implemented based on C++ language to ensure a high computational efficiency. This is possible by making use of Rcpp (Eddelbuettel et al., 2011) and RcppArmadillo (Eddelbuettel and C. Sanderson, 2014) which provide a seamless means to link R to C++ interfaces and the C++ matrix algebra library known as Armadillo (C. Sanderson, 2010). Therefore, having been streamlined in C++, all methods written within the package are highly efficient and are considerably faster than the few existing counterparts implemented in R (see Appendix D.1).

Using this flexible and efficient framework, the gmwm package contains all the features that are available within its predecessor (see Stebler et al., 2012), while adding a considerable amount of new features and functions. The main features of the new platform can be summarized as follows:

1. **Signal decomposition** across dyadic scales $\tau_j = 2^j$, $j = 1, ..., J$:
   - Supported wavelet filters: Daubechies (d), Fejer-Korovkin (fk), Battle-Lemarie (bl), Least Asymmetric (la), and Minimum Bandwidth (mb)
   - Discrete Wavelet Transform (DWT);
   - Maximum Overlap DWT (MODWT).

2. Computation of **summary statistics**:
   - Allan Variance (AV): computed under the traditional definition or the modified definition as described in Riley (2008) (denoted as $\sigma_g$);
   - Hadamard Variance (HV): a modification of the AV as described in Riley (2008);
   - Wavelet Variance (WV): computed based on the different filters and transforms (with an option for it to be computed robustly in case of outliers in the signal).

3. **Identifying the Models**:
   - WV plots: the log-log WV plots are used in the same way as the AV plots where the slopes of the linear parts and the non-linear parts indicate the presence of certain processes in the model which underlies the observed signal (see e.g. El-Sheimy et al., 2008).

4. **Estimating the Models**:
   - Parameter estimation: under a supplied model, identified from the WV plot, the “gmwm()” function contains many options to estimate the model parameters;
• **Statistical Inference:** either asymptotic or bootstrapped inference is available to create parameter confidence intervals and model goodness-of-fit tests.

5. Model Selection:

• **Graphical WV fit:** Having estimated the model, the WV implied by the latter can be compared to the observed WV to understand how well it describes the error signal;

• **Selection procedure:** under a specific set of models (or a general model) a criterion is computed to obtain the “best” model(s) for the signal as described in Guerrier et al. (2013a).

The first step to take in order to make use of these features is, of course, to load the calibration measurements onto the platform using either the function `read.imu()` (specifically tailored to loading certain IMU measurements) or the general R function `read.table()`. In the context of this chapter, in order to better describe these features within the following sections, we will make use of real-life IMU data which is made available within a package called “imudata”\(^1\). Each dataset comes from a different sensor (e.g. MTi-G Micro-Electro-Mechanical System (MEMS), NavChip, etc.) and has varying lengths of measurement. For the purposes of this chapter, we focus on examples using the NavChip and MTi-G datasets that contain respectively around 3,100,000 and 900,000 static observations with six columns representing the measurements for each accelerometer and gyroscope axis. The MTi-G MEMS data were previously analyzed in Guerrier et al. (2013b) (we will refer to this dataset as `mtig`), while the NavChip data is a recent acquire which was sampled at 250Hz for roughly 3.5 hours (we will refer to this dataset as `navchip`).

5.3.1 Identifying the Models

Once the data is available in the R session (see Appendix D.2), we can start the procedure of modelling the stochastic error issued from an inertial sensor which, first of all, consists in understanding what kind of model could best describe the observed signal. This is not necessarily an easy task but a simple and useful tool which is often employed for this purpose is the log-log plot of the AV versus its scales (see El-Sheimy et al., 2008). Indeed, according to the form of the AV, a researcher can understand what kind of underlying processes are contributing to the overall error model where a linear or non-linear behaviour in certain regions of the plot can be linked to specific processes. The WV based on the Haar wavelet decomposition is simply a rescaled version of the AV (i.e. \( \sigma_y = 2^\nu \)) and therefore the `gmwm` package makes use of the same visualization technique (i.e. a log-log plot of the WV versus its corresponding scales). In order to begin the visualization process, one must first obtain the WV of the process(es) by using the “`wvar()`” function. The results of this calculation can then be plotted by using the function “`plot()`” to obtain the plot in Figure 5.3.1, which depicts the WV for each signal separately. By adding the parameter of “`split = FALSE`” to the “`plot()`” function, one is able to have the WV superimposed according to whether the sensor is either a gyroscope or an accelerometer as in Figure 5.3.1. It is also possible to compute the robust WV (see Chapter 1) by adding the “`robust = TRUE`” parameter to the “`wvar()`” function. Within the same function, it

\( ^1 \)“imudata” R package is available at: [https://github.com/smac-group/datarepo](https://github.com/smac-group/datarepo) or by using “`gmwm::install_imudata()`” in R
5.3. Implementation of the GMWM

is also possible to specify the degree of efficiency of the robust estimator with respect to
the standard one where an efficiency close to 0.5 ensures a high level of robustness and
viceversa for a value close to 1. Computing the WV from a robust perspective is helpful to
check whether there appears to be contamination (e.g. outliers or extreme observations)
in the captured signal. Indeed, to understand if modeling within a robust paradigm is
required, one should compare the standard WV with the robust version by using the
function “compare.wvar()” to produce Figure 5.2. If the robust WV is visually different
from the standard WV, then a robust analysis of the signals is preferable.

5.3.2 Estimating the models

The new platform, as seen in the previous section, makes available some flexible plotting
tools which are already used (in different forms) to identify the models characterizing
inertial sensor stochastic errors. However, the parameters of these models are obviously
not known and need to be estimated. As described in Section 5.2, the GMWM allows for
the estimation of these parameters in an efficient and consistent manner and the function
which implements this method is the gmwm() function. There are multiple arguments
to this function which provide the users with a flexible range of options to tailor the
estimation to their needs. In the case of IMUs, there exists a function with preset values
ideal to model IMU error signals called gmwm.imu(). Both of these functions rely on users
supplying an error model which can be specified using a combination of all or a subset of
the following processes:

- GM($\beta, \sigma_{GM}^2$): a Gauss-Markov process;
- AR1($\phi, \sigma^2$): a first-order autoregressive process;
- WN($\nu^2$): a white noise process;
- QN($Q^2$): quantization noise process;
- RW($\gamma^2$): a random walk process;
- DR($\omega$): a drift process;
• **ARMA**(p, q): an autoregressive-moving average process.

It must be emphasized that a first-order autoregressive process (**AR1()**) is simply a reparametrization of the Gauss-Markov process (**GM()**). Denoting the frequency as \( f \), the parameterization of \( \phi \) to \( \beta \) is given by

\[
\phi = \exp (-\beta \Delta t)
\]

where \( \Delta t = \frac{1}{f} \). Similarly, the \( \sigma^2 \) is related to \( \sigma_{GM}^2 \) in the following manner

\[
\sigma^2 = \sigma_{GM}^2 \left( 1 - \exp (-2\beta \Delta t) \right).
\]

With this in mind, if the frequency \( f \) of the data is equivalent to 1, then the **GM()** process will be equivalent to the **AR1()**. This is the default frequency assumed by the software unless otherwise specified during the estimation procedure.

The latent processes underlying the error signal, whose parameters need to be estimated, can be specified by simply adding the different processes mentioned above via the “+” operator. However, there are some limits as to how many times a process can be included in a model. In particular, only the **GM()** or **AR1()** models can be included more than once (say \( k \) times) by specifying, for example, \( k*GM() \) while the other processes can only be included once within the same model\(^2\). Considering these conditions, once the model is specified, the software will perform a grid search to obtain appropriate starting values for the optimization procedure in (5.2.3) (see Appendix D.3 for details). However, one can also supply starting values for the GMWM optimization by writing the exact values within the bracket (e.g. “**AR1**(phi=0.9, sigma2=0.1)+**WN**(sigma2 = 1)”).

To provide an example of model estimation using the above features, let us consider the **navchip** data and take the second column (Y-axis gyroscope). In order to describe the

---

\(^2\)A sum of **ARMA**(p, q) models can in principle be estimated but the conditions for this to be certain are not clear.
latter we consider a reasonably complex model made by the sum of three Gauss-Markov processes in addition to a white noise, a quantization noise and random walk process which we estimate with the code below.

```r
> gyro.y = gmwm.imu(3*GM()+WN()+QN()+RW(), navchip[,2])
```

As for the `wvar()` function, also in this case it is possible to opt for a robust model estimation by simply adding the “`robust = TRUE`” parameter. The estimates for both these models are presented in Table 5.1 alongside their standard deviation. For the most part, the estimates between the classical and robust GMWM seem to agree since there does not appear to be a significant difference between them. The only exception is represented by the third Gauss-Markov process where \( \beta_3 \) and \( \sigma^2_{GM,3} \) differ between the two estimations. These differences can be also noticed in Figure 5.2 where the standard and robust WV appear to slightly differ (keeping in mind the logarithmic scale). It therefore appears that a robust analysis might be more appropriate.

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Classical</th>
<th>Robust (eff = 0.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>( 2.50 \cdot 10^{-1} ) (1.30 \cdot 10^{-11})</td>
<td>( 3.11 \cdot 10^{-1} ) (1.93 \cdot 10^{-11})</td>
</tr>
<tr>
<td>( \sigma^2_{GM,1} )</td>
<td>( 7.08 \cdot 10^{-9} ) (5.92 \cdot 10^{-13})</td>
<td>( 6.57 \cdot 10^{-9} ) (6.58 \cdot 10^{-13})</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>( 6.28 \cdot 10^{-3} ) (1.83 \cdot 10^{-13})</td>
<td>( 7.26 \cdot 10^{-3} ) (2.74 \cdot 10^{-13})</td>
</tr>
<tr>
<td>( \sigma^2_{GM,2} )</td>
<td>( 1.28 \cdot 10^{-8} ) (1.10 \cdot 10^{-13})</td>
<td>( 1.12 \cdot 10^{-8} ) (1.19 \cdot 10^{-13})</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>( 8.48 \cdot 10^{0} ) (1.52 \cdot 10^{-14})</td>
<td>( 1.25 \cdot 10^{1} ) (3.32 \cdot 10^{-14})</td>
</tr>
<tr>
<td>( \sigma^2_{GM,3} )</td>
<td>( 6.48 \cdot 10^{-9} ) (1.44 \cdot 10^{-11})</td>
<td>( 1.33 \cdot 10^{-8} ) (2.73 \cdot 10^{-11})</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( 6.94 \cdot 10^{-7} ) (2.01 \cdot 10^{-9})</td>
<td>( 5.30 \cdot 10^{-7} ) (2.52 \cdot 10^{-9})</td>
</tr>
<tr>
<td>( Q^2 )</td>
<td>( 2.29 \cdot 10^{-6} ) (8.79 \cdot 10^{-9})</td>
<td>( 2.74 \cdot 10^{-6} ) (1.24 \cdot 10^{-8})</td>
</tr>
<tr>
<td>( \gamma^2 )</td>
<td>( 3.90 \cdot 10^{-14} ) (4.50 \cdot 10^{-14})</td>
<td>( 5.44 \cdot 10^{-14} ) (5.10 \cdot 10^{-14})</td>
</tr>
</tbody>
</table>

Table 5.1: Standard vs robust parameter estimates (with corresponding standard deviations) for the NavChip Y-axis gyroscope. The robust estimator is based on an efficiency of 0.6 compared to the standard estimator.

Once the model is estimated, the software provides the ability to graphically validate it. Indeed, a first way to understand if the model fits the observed error signal well is to compare the relative observed WV and the WV implied by the estimated model as shown in the left plot of Figure 5.3.2. This plot can be produced simply by applying the `plot()` function to the object containing the estimated model. Additionally, one can seek to improve the fit by graphically observing individual processes which contribute to the model by using the parameter “`process.decomp = TRUE`”, which gives the right plot in Figure 5.3.2.

Aside from this graphical validation tool, the software provides statistical measures to assess the significance of the estimated parameters and of the model as a whole. Supposing that the model has been saved in an object called “`gyro.y`”, it is possible to obtain parameter confidence intervals (CIs) and a goodness-of-fit test (see equation (5.2.4)) by using the function “`summary(gyro.y, inference = TRUE)`”. The CIs provide a measure of how accurate the parameters estimations are thereby giving information as to whether or not parameters associated with a specific underlying process are significantly different from zero (e.g. \( H_0 : \theta = 0 \), with \( \theta \) being an element of the parameter vector \( \theta \)). Thus, this gives a reasonable indication of whether a process could be kept in the model or not. However, with this being said, it is known that confidence intervals are conditional on the
Figure 5.3: WV implied by the estimated $3\times\text{AR1()} + \text{WN()} + \text{QN()} + \text{RW()}$ model compared to the empirical WV of the NavChip Y-axis gyroscope.

The tools provided by the new platform and described in this section are extremely important to understand which models can be considered as good candidates to best describe the observed signal. Therefore, the following step that is needed is to select the “best” model(s) for which the related tools are described in the next section.

5.3.3 Selecting the Model

The main goal of model selection is to identify the most appropriate model, among those that have been estimated, to describe and predict future stochastic error measurements. Using the WIC, the two main options available to rank the models according to this criterion are the following:

1. **Manual**: the function `rank.models()` allows to enter a set of nested candidate models from which the user would like to select the best.

2. **Automatic**: the function `auto.imu()` allows to define an overall general model, say $\mathcal{M}_K$, in which all $K$ candidate models are nested (e.g. $\mathcal{M}_k \subseteq \mathcal{M}_K$, for $k = 1, \ldots, K$). An example of this approach applied to the `mtig` data is given in Figure 5.4 where the implied WV of the selected models is compared to the observed WV for each accelerometer and gyroscope.

Considering these two options, for both of them the platform also allows to choose the approach to estimate term $B$ in (5.2.5) as mentioned in Section 5.2. The first option is given by the parametric bootstrap where $H$ simulations are performed on the $k^{th}$ candidate model $F_{\theta_k}$ to obtain estimates of $\hat{\nu}(h)$ and $\nu(\hat{\theta}_k(h))$ that are then used to estimate the “optimism” term. This option has been shown to possess good properties in various simulation studies and can be used by specifying the parameter “bootstrap = TRUE”. However, the latter is very demanding from a computational standpoint while the asymptotic option, due to its closed analytical form, is less computationally intensive and has some optimal theoretical properties as shown in X. Zhang and Guerrier (2015). Indeed, in terms of computational time, the asymptotic option is roughly $K$ times faster than the
5.4 Upcoming features

The platform responds to the current needs for inertial sensor calibration while adding useful features which were not available before. However, the idea of modelling latent processes which is behind the GMWM allows it to be particularly suitable for extensions to more complicated settings. Different developments are indeed underway, or close to completion, concerning issues which are extremely relevant for inertial sensor calibration. The following sections briefly describe these extensions which will soon be implemented within the new software platform.

5.4.1 Multivariate GMWM

Currently sensor calibration is carried out separately for each accelerometer and gyroscope that build an IMU. This procedure is valid if one supposes that the error signals of each of

Figure 5.4: Results of running auto.imu() on the “mtig” data. Dashed line is the implied WV while the solid line is the empirical WV.
these components are independent from one another which may not always be a realistic assumption to make. Using the latent process structure, the idea is to explain the possible dependence of gyroscopes and accelerometers through a common model which is shared among them. To provide an example, we could suppose that the error signals of the triad of gyroscopes \((Y_{t}^{(x)}, Y_{t}^{(y)}, Y_{t}^{(z)})\) can be expressed as follows:

\[
Y_{t}^{(x)} = V_{t} + W_{t}^{(x)} \\
Y_{t}^{(y)} = V_{t} + W_{t}^{(y)} \\
Y_{t}^{(z)} = V_{t} + W_{t}^{(z)}
\]

where \(V_{t}\) is the process which is shared between the gyroscopes and \(W_{t}^{(i)}\) is the specific process that characterizes the error signal of the \(i^{th}\)-axis gyroscope. The idea of expressing the model in this manner defines a new method for the modelling of multivariate time series that is inspired by dynamic factor analysis. In this manner, we don’t only explain the dependence between the sensor components but we also reduce the number of parameters to estimate and possibly obtain better fits. This feature will be added to the proposed platform once the theoretical properties of the method are confirmed.

### 5.4.2 Calibration under dynamics

The correlation of IMU errors with system dynamics is a topic of great importance that may increase with the spreading use of low-cost inertial sensors. Recent studies confirmed that system dynamics may have an important influence on the characteristics of the error signal issued from inertial sensors (see Stebler et al., 2014a). The possibility of adapting noise models according to the dynamics that the system is subject to should therefore be considered in the navigation filters to enhance the performance of stochastic modeling. The proposed methodology is based on a non-stationary adaptation of the GMWM framework. Let us consider the following example of a non-stationary AR1 process:

\[
Y_{t} = \rho_{t}Y_{t-1} + \epsilon_{t}, \quad \epsilon_{t} \sim iid (0, \sigma_{\epsilon}^{2})
\]

where \(\rho_{t}\) is the autoregressive parameter that now depends on the time \(t\) through a function \(m(\cdot) : \mathbb{R} \mapsto (-1, 1)\). Indeed, we define \(\rho_{t} = m(Z_{t-1})\beta\) where \(Z_{t-1}\) denotes a vector of dynamics and \(\beta\) is the parameter that explains the impact of these dynamics on the parameters of the error signal model. This approach can be extended to other parameters of typical error models considered for inertial sensor calibration. The properties of the WV estimator \(\hat{v}\) and the implied WV \(v(\theta)\) considering such circumstances are being developed and will be later included in the proposed platform.

### 5.4.3 Additional features

Other features are envisaged for the future updates of the package that are based on recently developed methods and improvements on existing results. For example, in Guerrier et al. (2016b) it is emphasized how adding further process moments to the WV vector, such as the mean or autocovariances, can improve the finite sample performance of the GMWM estimator. This estimator is called the Augmented GMWM (GMWM+) and will be implemented as an additional option in the `gmwm.imu()` function with the possibility to specify the requirement on additional moments of the WV vector. Furthermore, a test
will be implemented to determine whether two error signals are generated from the same model or not. Finally, additional features are planned such as the possibility to employ other kinds of wavelet filters and use the robust WV as a basis for fault detection and isolation.

5.5 Conclusion

This chapter presented the main features and function descriptions of the new software platform for inertial sensor calibration based on the GMWM. This software is contained in an open-source R package called “gmwm” which provides all the tools for the modelling of the complex stochastic error signals that characterize IMUs, although this does not exclude the possibility of using it to estimate complex stochastic error models from other engineering applications. As a statistically sound alternative to the AVLR and as a computationally efficient and numerically stable alternative to the ML, the new platform based on C++ language allows to visualize data using the WV as the main summary statistic in order to identify the possible models which can then be quickly and consistently estimated. With graphical tools to assess how well the models fit the observed signals, the package also provides functions to (automatically) select the best model(s) based on their estimated prediction error. These approaches have been implemented in a computationally efficient manner and this chapter has also provided details on how these methods have been adapted to the specific and complex characteristics of IMU error signals through tailored algorithms. All this allows to identify the model and estimate its parameters which can then be inserted into a navigation filter (usually an extended Kalman filter) to improve the navigation accuracy of the sensors.
Conclusion

This thesis has been mainly developed around the study and extension of the GMWM framework towards the field of robustness and the parametric estimation of random fields, in particular spatial models. To do so, a new (robust) M-estimator of the WV for multi-dimensional random fields was proposed and its joint asymptotic properties were studied in order for it to be used within the GMWM. Moreover, a specific focus was given to the class of dependent data models that we call “latent” models and the properties of common estimators, such as the MLE and GMM, were studied when estimating these models based on the identifiability results in Chapter 2. These results allow to clarify the conditions for consistency of estimators for many commonly used models such as state-space and structural time series models. Based on these findings, the third chapter proposes the RGMWM as a robust, flexible and computationally efficient method to estimate time series and spatial models while the final two chapters present the implementation of the GMWM framework within the \texttt{gmwm R} package where the last chapter specifically focuses on its usefulness for inertial sensor calibration. The relative importance of this package is testified by the number of downloads in \texttt{RStudio} since its release on February 10\textsuperscript{th}, 2016, which is just below two thousand by June 17\textsuperscript{th}, 2016 (without considering the downloads in \texttt{R} and the direct downloads of the package from the GitHub repository).

The new WV estimator provides the basis for a robust version of WV-based tests such as unit root, Portmanteau or isotropy tests. Various research is underway to study improvements and extensions of these tests given the asymptotic and robustness results presented in Chapter 1. However, these results are particularly important for the development of the GMWM framework. Indeed, the main advantage of the GMWM over alternative methods such as GMM or MD estimators lies in the fact that the WV provides an auxiliary parameter of relatively small dimensions even for extremely large sample sizes in which all the information of the sample is summarized. Moreover, an explicit analytic expression is not hard to derive for this auxiliary parameter. Therefore, at the price of a slight loss in efficiency, the GMWM greatly reduces the problem of moment or auxiliary parameter selection and can eventually be improved by choosing an appropriate weighting matrix as well as adding few other moments to the WV vector. Both of these topics were considered before and are now currently under investigation with encouraging preliminary results.

The flexibility of the GMWM framework highlighted in this thesis allows it to adapt to a variety of different settings. Firstly, the concept of latent models based on which the GMWM was initially proposed can provide an alternative framework to the modelling of multivariate time series with a particular interest for mixed-effect models. Indeed, latent models can be shared among time series therefore explaining the dependence between them and this idea has already given promising results when applied to examples coming from the domain of ecology. Moreover, the possibility of adding covariates to explain the WV can allow the GMWM to adapt to non-stationary settings with time-varying
parameters. Finally, the study of the statistical properties of the WIC model selection criterion can deliver another sound and flexible tool for the selection of standard and latent time series and spatial models.

To conclude, the nature of research wants it to lead to projects that were not envisaged at the start but became interesting as a result of the main investigations. For this reason a variety of projects have stemmed from this work among which there is the study and extension of the model selection algorithm mentioned in the introduction and a new non-parametric bootstrap for dependent data for moderate sample sizes.
Appendices
Appendix A

Appendices for Chapter 1

A.1 Choice of tuning constant

Given that Theorem 1.3.1 provides an expression for the variance of the WV estimator, here we provide a brief discussion on the choice of the tuning constant $c$. The definition of this value is based on the desired level of efficiency compared to the classical estimator and varies according to the chosen $\psi$-function. However, an explicit and intuitive rule for the choice of this constant is available only when considering the process $(W_{k,j})$ as Gaussian. In the latter case, the estimator in (1.3.1) is the result of a minimization under the standard Gaussian assumption (i.e. zero mean and unit variance) and because of this we can obtain expressions for the variance of both the classical estimator and the robust estimator, which we denote as $\sigma_j^2$ and $\tilde{\sigma}_j^2(c)$ respectively. In this setting we see that these expressions depend solely on $c$ and therefore, for a general scale $j$ and defining $\alpha \in [0, 1]$ as the desired level of efficiency, a rule to select the tuning constant $c$, given a specific $\psi$-function, is to find the solution in $c$ to the expression

$$\frac{\sigma_j^2}{\tilde{\sigma}_j^2(c)} - \alpha = 0.$$ 

For example, choosing $\alpha = 0.95$ delivers a tuning constant $c \approx 7.88$ when using the Tukey $\psi$-function and $c \approx 2.38$ when using the Huber $\psi$-function (respectively $c \approx 4.97$ and $c \approx 1.22$ for $\alpha = 0.6$). The choice of the efficiency level is subjective and can be supported by a sensitivity analysis comparing the classical and the robust estimates starting from a low efficiency level (e.g. 0.5).

A.2 Bias Study

We consider the setting where the observations of the random field generate wavelet coefficients issued from a Gaussian process. If one supposes that the Gaussian distribution is at most a fair approximation of the data generating model, then the latter should be considered as belonging to a neighbourhood of the postulated model, say $F_{\epsilon}$, also called the contaminated distribution. Under $F_{\epsilon}$, the proposed estimator is hence biased, albeit that the bias is bounded as a result of the bounded $\psi$-functions. The size of the bias depends on the tuning constant $c$ and the form of the $\psi$-function. Hence, by fixing the desired efficiency of the resulting estimator via an appropriate tuning constant $c$, the choice of the $\psi$-function could be seen as a bias-minimization problem.
This bias-minimization problem was considered by Huber (1981) within the “minimax” approach. Hampel et al. (1986) consider the (general) case of a contaminated distribution $F_\epsilon = (1 - \epsilon)F_\theta + \epsilon H$, $0 \leq \epsilon < 1$. In the present setting, we consider $F_\theta$ to be the standard Gaussian model. Let us also write the estimator defined through the function $\psi(\cdot)$ as a functional of the underlying distribution, i.e. $T(F_\epsilon)$ or $T(F_\theta)$, and we suppose it to be Fisher consistent (i.e. $E_{F_\theta}[\psi(r_{k,j};c)] = 0$, with $r_{k,j}$ being a standard Gaussian variable which represents the standardised wavelet coefficients in the context of this chapter. Using a Von Mises expansion of $T(F_\epsilon)$ around $T(F_\theta)$, the (approximate) asymptotic bias of the estimator is expressed as

$$T(F_\epsilon) - T(F_\theta) \approx \epsilon \int \text{IF}(r, \psi, F_\theta)dH(r) \quad (B-1)$$

where $\text{IF}(r, \psi, F_\theta)$ is the IF of the estimator and $r = r_{k,j}$ for $j = 1, \ldots, J$.

The bias of $T(\cdot)$ under a contaminated distribution $F_\epsilon$ is hence directly proportional to its IF under the contaminating distribution $H$. Therefore, given a desired level of efficiency (i.e. having selected the tuning constant $c$), it is possible to select a $\psi$-function over another one by comparing their (approximate) bias.

The IF of an $M$-estimator is given by

$$\text{IF}(r, \psi, F_\theta) = M^{-1}(\psi, F_\theta)\psi(r, \theta) \quad (B-2)$$

where $M(\psi, F_\theta)$ for Fisher consistent $M$-estimators is defined as

$$M(\psi, F_\theta) = \int \psi(r, \theta)s^T(r, \theta)dF_\theta(r)$$

with $s(r, \theta)$ being the score function ($s(r, \theta) = r^2$ in our case). Since $M(\psi, F_\theta)$ depends on the underlying $\psi$-function and is constant under a postulated model $F_\theta$, we will denote it generically as $M_\psi$. Using (B-1) yields the following approximate bias for the proposed estimator

$$B_\psi(H, F_\theta) = \epsilon M_\psi^{-1} \int \psi(r, \theta)dH(r). \quad (B-3)$$

If $H = F_\theta$, the IF has value 0 and, consequently, so does the bias. As (B-3) shows, given a contaminating distribution $H$, the bias of the proposed estimator ultimately depends on the chosen $\psi$-function. The choice of the $\psi$-function can therefore be made based on the minimization of a risk-function which takes into account expression (B-3). A possible risk-function could simply be

$$tr \left(B_\psi(H, F_\theta)B_\psi^T(H, F_\theta)\right). \quad (B-4)$$

To compare the bias-performance of the Huber and Tukey biweight $\psi$-functions (indexed with $[Hub]$ and $[Bi]$ respectively), we computed (B-4) for $H$ being the dirac distribution with pointmass at $\delta \in [0, 10]$. The tuning constants for the two $\psi$-functions were chosen to guarantee 95% asymptotic efficiency at the normal model, yielding $c_{[Hub]} \cong 2.38$ and $c_{[Bi]} \cong 7.88$.

As the top part of Figure A.1 highlights, the risk function of the $\psi_{[Bi]}$-function peaks and descends becoming constant around $c_{[Bi]} \cong 7.88$, whereas the $\psi_{[Hub]}$-function grows and remains constant after $c_{[Hub]} \cong 2.38$. Having approximately the same behavior until around $\delta = 5$ (with the $\psi_{[Bi]}$-function’s risk being greater over a small interval), the risk
Figure A.1: Top: theoretical risk function for a series of Dirac points. Bottom: theoretical risk function for a Gaussian contamination with different scale parameters.
of the \( \psi_{[B]} \)-function is constantly smaller, indicating that the latter function appears to have an overall better performance in terms of risk.

In the bottom of Figure A.1 the risk function of the \( \psi \)-functions are computed with \( H \) being another zero-mean Gaussian distribution with \( \sigma^2 \in (0, 30] \). It can again be seen how the risk of the two functions initially behave in a similar way whereas the \( \psi_{[B]} \)-function’s risk becomes constantly smaller compared to the one of the \( \psi_{[H, b]} \)-function after a given magnitude of contamination. In addition, we can see how both functions deliver unbiased estimators when the contaminating distribution corresponds to the true distribution \( F_\theta \) (i.e. when the contaminating variance equals 1).
A.3 Proofs

A.3.1 WV Influence Function

Below is the proof of Theorem 1.2.1 which follows standard procedures for the derivation of the influence function for M-estimators.

Proof. Let us define \( e = [\epsilon_d]_{d=1,\ldots,D} \) as the vector of contamination levels for each dimension \( d \) and, under the assumption of Theorem 1.2.1, let \( \epsilon = \prod_{d=1}^{D} \epsilon_d \). Moreover, let \((X_k)\) be a random field generated by a model in a neighbourhood of \( F_\theta \), i.e. \( F_\epsilon = (1-\epsilon)F_\theta + \epsilon \Delta_z \), with small \( \epsilon > 0 \). \( F_\epsilon \) is the standard contamination model with \( \Delta_z \) the Dirac function at \( z = X_k \), since the maximal bias on a statistic (estimator) is obtained when the contaminating distribution is a Dirac (see Hampel et al., 1986).

Very generally, let \((x_i), i \in I \subset \mathbb{N}_+\), represent an observed sequence of a random variable \((X_i)\) following a model \( F_\theta \). An M-estimator is defined as the solution for \( \lambda \) of the following equation

\[
\sum_{i \in I} \psi(x_i, \lambda) = 0 \quad \text{(C-1)}
\]

where \( \lambda \) can be \( \theta \) or a function of it (i.e. \( \lambda(\theta) \)) and \( \psi(\cdot) \) is a function which can be unbounded or bounded. Considering \( F_\epsilon \) as the data generating model, then \( \lambda(F_\epsilon) \) (i.e. the estimator written as a functional of \( F_\epsilon \)) is implicitly defined in

\[
\mathbb{E}_{F_\epsilon} \left[ \psi \left( x_i, \lambda(F_\epsilon) \right) \right] = 0. \quad \text{(C-2)}
\]

In our case we consider the wavelet coefficients that can also be written as functionals of \( F_\epsilon \), namely \( W_{k,j}(F_\epsilon) \), and seek to estimate the WV considering \( F_\epsilon \), i.e.

\[
\mathbb{E}_{F_\epsilon} \left[ \psi \left( W_{k,j}(F_\epsilon), \hat{\psi}_j^2(F_\epsilon) \right) \right] = 0. \quad \text{(C-3)}
\]

Under the assumption of Theorem 1.2.1 and using the time series adaptation of the IF from Kunsch (1984), the IF of \( \hat{\psi}_j^2 \) is obtained by taking the Gâteaux derivative of (C-3) with respect to \( \epsilon \) when \( \epsilon \to 0^+ \), i.e.

\[
\frac{\partial}{\partial \epsilon} \left[ (1-\epsilon)\mathbb{E}_{F_\theta} \left[ \psi \left( W_{k,j}(F_\epsilon), \hat{\psi}_j^2(F_\epsilon) \right) \right] \right]_{\epsilon=0^+} + \frac{\partial}{\partial \epsilon} \left[ \epsilon \psi \left( W_{k,j}(F_\epsilon), \hat{\psi}_j^2(F_\epsilon) \right) \right]_{\epsilon=0^+}
\]

\[
= - \mathbb{E}_{F_\theta} \left[ \psi \left( W_{k,j}(F_\theta), \nu_j^2 \right) \right] + \mathbb{E}_{F_\theta} \left[ \frac{\partial}{\partial W_{k,j}^2} \psi \left( W_{k,j}^2, \nu_j^2 \right) \right]_{W_{k,j}^2=W_{k,j}^2(F_\theta)} \cdot \frac{\partial}{\partial \epsilon} \left[ W_{k,j}^2(F_\epsilon) \right]_{\epsilon=0^+}
\]

\[
+ \mathbb{E}_{F_\theta} \left[ \frac{\partial}{\partial \hat{\psi}_j^2} \psi \left( W_{k,j}(F_\epsilon), \nu_j^2 \right) \right] \cdot \frac{\partial}{\partial \epsilon} \left[ \hat{\psi}_j^2(F_\epsilon) \right]_{\epsilon=0^+} + \psi \left( W_{k,j}(z), \nu_j^2 \right)
\]

\[
= - \mathbb{E}_{F_\theta} \left[ \psi \left( W_{k,j}(F_\theta), \nu_j^2 \right) \right] + \mathbb{E}_{F_\theta} \left[ \frac{\partial}{\partial W_{k,j}^2} \psi \left( W_{k,j}^2, \nu_j^2 \right) \right]_{W_{k,j}^2=W_{k,j}^2(F_\theta)} \cdot \text{IF}(z, W_{k,j}^2, F_\theta)
\]

\[
+ \mathbb{E}_{F_\theta} \left[ \frac{\partial}{\partial \hat{\psi}_j^2} \psi \left( W_{k,j}(F_\theta), \nu_j^2 \right) \right] \cdot \text{IF}(z, \hat{\psi}_j^2, F_\theta) + \psi \left( W_{k,j}^2(z, F_\theta), \nu_j^2 \right) \equiv K
\]
By the chain rule we have that \( \text{IF}(\mathbf{z}, \hat{\nu}^2, F_\theta) \propto \text{IF}(\mathbf{z}, W_{k,j}^2, F_\theta) \),
therefore
\[
K \propto -E_{F_\theta} \left[ \psi \left( W_{k,j}^2(F_\theta), \nu^2_j \right) \right] +
E_{F_\theta} \left[ \frac{\partial}{\partial W_{k,j}^2} \psi \left( W_{k,j}^2, \nu^2_j \right) \right]_{W_{k,j}^2 = W_{k,j}^2(F_\theta)} \text{IF}(\mathbf{z}, \hat{\nu}^2_j, F_\theta) +
E_{F_\theta} \left[ \frac{\partial}{\partial \nu_j^2} \psi \left( W_{k,j}^2(F_\theta), \nu_j^2 \right) \right]_{W_{k,j}^2 = W_{k,j}^2(F_\theta)} \text{IF}(\mathbf{z}, \hat{\nu}^2_j, F_\theta) + \psi \left( W_{k,j}^2(\mathbf{z}, F_\theta), \nu^2_j \right) = 0.
\]

For consistent estimators of the WV, we have that \( E_{\nu} \left[ \psi(W_{k,j}^2(F_\theta), \nu^2_j) \right] = 0 \). We finally get
\[
\text{IF}(\mathbf{z}, \hat{\nu}^2_j, F_\theta) \propto -D^{-1} \psi \left( W_{k,j}^2(\mathbf{z}), \theta \right)
\]
with
\[
D = E_{F_\theta} \left[ \frac{\partial}{\partial W_{k,j}^2} \psi \left( W_{k,j}^2, \nu_j^2 \right) \right]_{W_{k,j}^2 = W_{k,j}^2(F_\theta)} + E_{F_\theta} \left[ \frac{\partial}{\partial \nu_j^2} \psi \left( W_{k,j}^2(F_\theta), \nu_j^2 \right) \right].
\]

Since \( D \) does not depend on the contamination mass \( \mathbf{z} \), the IF of the estimator of the WV is bounded if \( \psi(\cdot) \) is bounded, thus concluding the proof.

### A.3.2 WV Identifiability

In this appendix we discuss the identifiability of the WV \( \nu^2_j \) when using the Huber and Tukey biweight \( \psi \)-functions. We first define the Tukey biweight function with redescending weights (see Beaton and Tukey, 1974). The biweight \( \psi \)-function delivers the following weights \( \omega(\cdot) \)
\[
\omega_{[Bi]}(r_{k,j}; \nu^2_j, c) = \begin{cases} 
\left( \frac{r_{k,j}}{c} \right)^2 - 1 & \text{if } |r_{k,j}| \leq c \\
0 & \text{if } |r_{k,j}| > c
\end{cases}
\]
and, if one supposes the normality for the wavelet coefficients, then the correction term \( a(\nu^2_j) \) is
\[
a_{[Bi]}(c) = E_{\Phi} \left[ \omega_{[Bi]}^2(r_{k,j}; \nu^2_j, c) r_{k,j}^2 \right] = \frac{1}{c^8} \mu^6_c - \frac{4}{c^6} \mu^4_c + \frac{6}{c^4} \mu^2_c - \frac{4}{c^2} \mu^2_c + \mu^2_c
\]
with \( \mu^i_c \) being the \( i \)-th truncated moment under the standard normal distribution between \(-c \) and \( c \).

On the other hand, Huber’s \( \psi \)-function has well known properties and has easily tractable derivatives when developing its asymptotic properties. Its weights are given by
\[
\omega_{[Hub]}(r_{k,j}; \nu^2_j, c) = \min \left( 1; \frac{c}{r_{k,j}} \right)
\]
when using these functions, it is necessary to understand if they deliver functions which enable to identify the unknown parameter \( \nu^2_j \) and state that
\[
E \left[ \psi \left( W_{k,j}, \nu^2 \right) \right] = 0
\]
if and only if \( \kappa^2 = \nu_j^2 \) (i.e. there is a unique solution for \( \nu_j^2 \)). This condition is often called “global identifiability” and it is an essential condition to prove consistency and asymptotic normality of the estimator which is often assumed for simplicity (as, for example, in Mondal and Percival, 2012b). To verify global identifiability, let us set the condition below:

\[(C4)' \quad \text{The process } (W_{k,j}) \text{ is a zero-mean Gaussian process with autocovariance sequence } (\varphi_W(h)) \text{ such that } |\varphi_W(h)| = \mathcal{O}(\rho^k), 0 < \rho < 1.\]

Condition \((C4)'\) is always verified regarding the mean constraint since all stationary models deliver zero-mean wavelet coefficients \((W_{k,j})\) with finite \(\text{WV} \nu_j^2\) and many non-stationary models with stationary backward differences can also respect this condition. This is the case, in the time series setting for example, for all stationary ARMA and various state-space models. However, the assumption of a Gaussian model for \((W_{k,j})\) issued from the previously mentioned models is a relatively strong one but (apart from the case where \((X_k)\) is itself Gaussian) it is a frequently assumed condition for the wavelet coefficients and, according to the type of process, could be a reasonable approximation due to the averaging nature of the filter. Considering these observations and defining \(\gamma \equiv c(\kappa^2/\nu_j^2)^{1/2}\), we have the following theorem.

**Theorem A.3.1.** Under Condition \((C4)'\) we have that \(\nu_j^2\) is identifiable using the Huber weight function and is identifiable using the Tukey biweight function for \(\gamma > 3.5\).

The condition of \(\gamma > 3.5\) is very mild. Indeed, the result of Theorem A.3.1 implies that the equation \(\mathbb{E}[\psi(W_{k,j}, \kappa^2)] = 0\) has the unique solution \(\kappa^2 = \nu_j^2\) if \(\kappa\) belongs to the set \(\{x \in \mathbb{R} | \frac{\kappa^2}{2c} < x < \infty\}\) for the Tukey biweight function. In other words, the parameter \(\nu_j^2\) is identifiable if \(c > 3.5\) so that it belongs to the previously defined set. Using the results of Theorem 1.3.2, this condition is very reasonable as it is satisfied for any efficiency larger than approximately 2.5%, an efficiency which is already too low to make any sense in practice. The proofs of Theorem A.3.1 can be found below.

**Proof.** Let us start with Huber weight function and, for this, let

\[
X = \begin{cases} 
    r_{k,j}^2 & \text{for } |r_{k,j}| \leq c \\
    c^2 & \text{for } |r_{k,j}| > c
\end{cases}
\]

with \(r_{k,j} = W_{k,j}/\kappa\), \(\kappa^2 \in \{x \in \mathbb{R} | 0 < x < \infty\}\) and let us consider the function \(\mathbb{E}[\psi(W_{k,j}, \kappa^2)]\). For Huber weights we define \(q(r_{k,j}, c) \equiv \mathbb{E}[X - a_\psi(c)]\) where \(a_\psi(c)\) is a constant for a given \(c\). For global identifiability we need to prove that \(q(r_{k,j}, c)\) has a unique solution in \(\nu_j^2\) and to do so we prove that its derivative is a strictly monotone function in \(\kappa^2\). Indeed, we have by definition that \(\mathbb{E}[\psi(W_{k,j}, \kappa^2)] = 0\) if \(\kappa^2 = \nu_j^2\) and if the derivative of \(q(r_{k,j}, c)\) is strictly monotone then the solution is unique. Let us denote \(\mathbb{P}[X]\) as the probability of
Appendix A. Appendices for Chapter 1

If we prove that the term A in (C-6) is strictly positive or negative, we prove that the derivative is too. By rewriting A we have

\[ 2 \gamma \phi(\gamma) + 1 - 2 \Phi(\gamma) = 2 \gamma \phi(\gamma) + 2 \Phi(0) - 2 \Phi(\gamma) = 2(\gamma \phi(\gamma) + \Phi(0) - \Phi(\gamma)) \]

and we prove that this is quantity is strictly negative since \( \gamma \phi(\gamma) < \Phi(\gamma) - \Phi(0) \) given that \( \gamma > 0 \).

Now, let us prove the identifiability for the Tukey biweight function. Therefore, in the same manner let

\[ X = \begin{cases} \left( \frac{(r_{k,j})^2}{c} - 1 \right)^2 r_{k,j}^2 & \text{for } |r_{k,j}| \leq c \\ 0 & \text{for } |r_{k,j}| > c \end{cases} \]

and let \( \kappa \) belong to the set \( \{ x \in \mathbb{R} \mid c^* < x < \infty \} \) where \( c^* \) denotes a positive constant such that \( c^* < \nu_j \). Let us again follow the same procedure and notations as used for
the proof of global identifiability of the Huber weights. With $a_\psi(c)$ being this time the correction term for the Tukey biweight function, in this case we have

$$
E[X - a_\psi(c)] = E[X] - a_\psi(c) = E[X \mid r_{k,j} \leq c] P[r_{k,j} \leq c] - a_\psi(c)
$$

$$
= \frac{\nu_1^{10}}{\kappa^{10} c^8} \left\{ E \left[ \frac{r_{k,j}^{10}}{\nu_1^{10}} \mid |\alpha| \leq \gamma \right] P[|\alpha| \leq \gamma] \right\}
- \frac{4\nu_2^8}{\kappa^{8} c^6} \left\{ E \left[ \frac{r_{k,j}^8}{\nu_2^8} \mid |\alpha| \leq \gamma \right] P[|\alpha| \leq \gamma] \right\}
+ \frac{6\nu_3^6}{\kappa^{6} c^4} \left\{ E \left[ \frac{r_{k,j}^6}{\nu_3^6} \mid |\alpha| \leq \gamma \right] P[|\alpha| \leq \gamma] \right\}
- \frac{4\nu_4^4}{\kappa^{4} c^2} \left\{ E \left[ \frac{r_{k,j}^4}{\nu_4^4} \mid |\alpha| \leq \gamma \right] P[|\alpha| \leq \gamma] \right\}
+ \frac{\nu_5^2}{\kappa} \left\{ E \left[ \frac{r_{k,j}^2}{\nu_5^2} \mid |\alpha| \leq \gamma \right] P[|\alpha| \leq \gamma] \right\} - a_\psi(c)
$$

$$
= c^2 \left[ \frac{1}{\gamma^{10}} \left( 1890 \Phi(\gamma) - 2(159 + 315\gamma^2 + 63\gamma^4 + 9\gamma^6 + \gamma^8) \phi(\gamma) - 945 \right) \frac{\mu_1^{10}}{\mu_1} 
- \frac{4}{\gamma^8} \left( 210 \Phi(\gamma) - 2(105 + 35\gamma^2 + 7\gamma^4 + \gamma^6) \phi(\gamma) - 105 \right) \frac{\mu_2^8}{\mu_2} 
+ \frac{6}{\gamma^6} \left( 30 \Phi(\gamma) - 2(15 + 5\gamma^2 + \gamma^4) \phi(\gamma) - 15 \right) \frac{\mu_3^6}{\mu_3} 
+ \frac{4}{\gamma^4} \left( 6 \Phi(\gamma) - 2(3 + \gamma^2) \phi(\gamma) - 3 \right) \frac{\mu_4^4}{\mu_4} + \frac{1}{\gamma^2} \left( 2 \Phi(\gamma) - 2 \phi(\gamma) - 1 \right) \frac{\mu_5^2}{\mu_5} \right] - a_\psi(c).
$$

Next, we define $g(\gamma) \equiv E[X - a_\psi(c)]$ and we know that $g(\gamma)$ has a unique solution in $\gamma$ if the expression in square brackets in $g(\gamma)$ has a unique solution in $\gamma$. Hence, by taking the derivative we obtain

$$
\frac{\partial}{\partial \gamma} g(\gamma) = -\frac{\mu_1^{10}}{\gamma^{11}} + \frac{32}{\gamma^2} \mu_8 + \frac{1}{\gamma^{10}} \left( 1890 \phi(\gamma) - \left( 1890 + 1890\gamma^2 + 630\gamma^4 + 126\gamma^6 + 18\gamma^8 \right) \phi(\gamma) - \gamma \phi(\gamma) \mu_1^{10} \right)
- \frac{4}{\gamma^8} \left( 210 \phi(\gamma) - \left( 210 + 210\gamma^2 + 70\gamma^4 + 14\gamma^6 \right) \phi(\gamma) - \gamma \phi(\gamma) \mu_2^8 \right)
- \frac{36}{\gamma^6} \mu_6 + \frac{6}{\gamma^6} \left( 30 \phi(\gamma) - \left( 30 + 30\gamma^2 + 10\gamma^4 \right) \phi(\gamma) - \gamma \phi(\gamma) \mu_3^6 \right)
- \frac{16}{\gamma^4} \mu_4 - \frac{4}{\gamma^4} \left( 6 \phi(\gamma) - \left( 6 + 6\gamma^2 \right) \phi(\gamma) - \gamma \phi(\gamma) \mu_4^4 \right)
- \frac{2}{\gamma^2} \mu_2 + \frac{1}{\gamma^2} \left( 2 \phi(\gamma) - \left( 2 \phi(\gamma) - 2 \gamma^2 \phi(\gamma) \right) \right)
$$

whose value is strictly negative for $\gamma > 3.5$ thereby defining the value $c^* = \tau_{\nu_1} / 2c$. □
A.3.3 Scale-wise Asymptotic Properties of $\hat{\nu}_j^2$

In this appendix we discuss the asymptotic normality of the proposed M-estimator for each wavelet decomposition scale $j$. Before proving Theorem 1.3.1, we need the result of the following lemma if we consider using the Huber $\psi$-function.

**Lemma A.3.1.** The function $\psi(W_{k,j},\nu_j^2)$ using Huber weights is Bouligand-differentiable as follows

$$\psi'(W_{k,j},\nu_j^2) = \begin{cases} -\frac{W_{k,j}^2}{\nu_j^2} & \text{if } |r_{k,j}| \leq c \\ 0 & \text{if } |r_{k,j}| > c \end{cases}$$

The proof of this lemma is given below.

**Proof.** Let us define $r_0 \equiv W_{k,j}/\sqrt{\nu_0^2}$ and $r \equiv W_{k,j}/\sqrt{\nu^2}$ where $\nu^2 = \nu_0^2 + h$. By the definition in Scholtes (2012), a function $f(\cdot)$ is Bouligand differentiable (B-differentiable) at point $x_0$ if it is directionally differentiable at this point and there exists a function $f'(\cdot)$ such that $f(x_0 + h) = f(x_0) + f'(x_0)h + o(h)$. Using the approach of Christmann and Van Messem (2008), we first show that the function $\psi(W_{k,j},\nu_j^2)$ is first degree B-differentiable using Huber weights. Below are the computations of the B-derivatives for the five cases of the Huber weight function:

1. Setting $r_0 = c$ we have:
   - If $h \geq 0$ ($r \leq c$):
     $$\psi'(W_{k,j},\nu_j^2)(h) + o(h) = \psi(W_{k,j},\nu_0^2 + h) - \psi(W_{k,j},\nu_0^2) = r^2 - a_\psi(c) - r_0^2 + a_\psi(c)$$
     $$= \frac{W_{k,j}^2}{\nu_0^2} - \frac{W_{k,j}^2}{\nu_0^2} \frac{h}{\nu_0^2 + h}$$
     $$= \frac{W_{k,j}^2}{\nu_0^2} \left( \frac{h}{\nu_0^2} - \frac{h^2}{\nu_0^2(\nu_0^2 + h)} \right)$$
     $$= -\frac{W_{k,j}^2}{\nu_0^2} h + \frac{W_{k,j}^2 h^2}{\nu_0^2(\nu_0^2 + h)} \equiv \Delta$$

   - If $h < 0$ ($r > c$):
     $$\psi'(W_{k,j},\nu_j^2)(h) + o(h) = \psi(W_{k,j},\nu_0^2 + h) - \psi(W_{k,j},\nu_0^2) = c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = c^2 - c^2 = 0$$

2. Setting $r_0 = -c$ we have:
   - If $h < 0$ ($r < -c$):
     $$\psi'(W_{k,j},\nu_j^2)(h) + o(h) = \psi(W_{k,j},\nu_0^2 + h) - \psi(W_{k,j},\nu_0^2) = c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = 0$$
• If \( h \geq 0 \) (\( r \geq -c \)):

\[
\psi'(W_{k,j}, \nu_0^2)(h) + o(h) = \psi(W_{k,j}, \nu_0^2 + h) - \psi(W_{k,j}, \nu_0^2) = r^2 - a(\psi)(c) - r_0^2 + a(\psi)(c) = \ldots = \Delta
\]

3. Setting \( r_0 > c \) we have:

\[
\psi'(W_{k,j}, \nu_0^2)(h) + o(h) = \psi(W_{k,j}, \nu_0^2 + h) - \psi(W_{k,j}, \nu_0^2) = c^2 - a(\psi)(c) - r_0^2 + a(\psi)(c) = 0
\]

4. Setting \( r_0 < -c \) we have:

\[
\psi'(W_{k,j}, \nu_0^2)(h) + o(h) = \psi(W_{k,j}, \nu_0^2 + h) - \psi(W_{k,j}, \nu_0^2) = c^2 - a(\psi)(c) - r_0^2 + a(\psi)(c) = 0
\]

5. Setting \(-c < r_0 < c\) we have:

\[
\psi'(W_{k,j}, \nu_0^2)(h) + o(h) = \psi(W_{k,j}, \nu_0^2 + h) - \psi(W_{k,j}, \nu_0^2) = r^2 - a(\psi)(c) - r_0^2 + a(\psi)(c) = \ldots = \Delta
\]

We therefore have that the first B-derivative of the function \( \psi(W_{k,j}, \nu_j^2) \) is given by

\[
\psi'(W_{k,j}, \nu_j^2) = \begin{cases} 
-\frac{W_{k,j}^2}{\nu_j^2} & \text{if } |r_{k,j}| \leq c \\
0 & \text{if } |r_{k,j}| > c
\end{cases}
\]

The approach used in this proof can be used to obtain expressions for the B-derivatives of other piecewise differentiable weight functions (see Scholtes, 2012). It can be seen how it extends the classic derivative for \( |r_0| < c \) also to the points \( \nu_0^2 \) such that \( |r_0| = c \). However, the Frechet differentiability of this function has also been discussed in Clarke (1986).

\[\square\]

Technical Lemma A.3.1 is useful for the results on asymptotic normality of the proposed estimator to hold in case the choice of the \( \psi \)-function corresponds to the Huber \( \psi \)-function, proving that this function respects Condition (C6) which is required by Theorem 1.3.1. The proof of this theorem, which is valid for a general \( \psi \)-function that respects Condition (C6), is given below.

**Proof.** Since \( (W_{k,j}) \) is a strictly stationary ergodic process, then so is \( \left( \psi(W_{k,j}, \nu_j^2) \right) \) which is a bounded and time-invariant function of it. Hence, \( \left( \psi(W_{k,j}, \nu_j^2) \right) \) satisfies the Uniform Weak Law of Large Numbers (UWLLN) under Theorem 4.1 of Wooldridge (1994). Condition (C2) implies that there is a unique minimum for the function \( q(W_{k,j}, \nu_j^2) = -\int \psi(W_{k,j}, \nu_j^2)d\nu_j^2 \) and, with condition (C1), the weak consistency of \( M \)-estimators (Theorem 4.3, Wooldridge, 1994) yields

\[
\nu_j^2 \xrightarrow{p} \nu_j^2.
\]
Given condition (C6), let us denote \( \psi'(W_{k,j}, \nu_j^2) = \partial_\nu \psi(W_{k,j}, \nu_j^2) \) and apply a Maclaurin expansion of \( \sum_{k_j \in K_j} \psi(W_{k,j}, \nu_j^2) \) around \( \nu_j^2 \) obtaining

\[
\sum_{k_j \in K_j} \psi(W_{k,j}, \nu_j^2) = \sum_{k_j \in K_j} \psi(W_{k,j}, \nu_j^2) + \sum_{k_j \in K_j} \psi'(W_{k,j}, \nu_j^2)(\hat{\nu}_j^2 - \nu_j^2) = 0
\]

where

\[
\|\nu_j^2 - \nu_j^2\| \leq \|\hat{\nu}_j^2 - \nu_j^2\|. \tag{C-7}
\]

Multiplying by \( \sqrt{M_j} \) and rewriting yields

\[
\sqrt{M_j}(\hat{\nu}_j^2 - \nu_j^2) = \left[ -\frac{1}{M_j} \sum_{k_j \in K_j} \psi'(W_{k,j}, \nu_j^2) \right]^{-1} \left[ \frac{1}{M_j} \sum_{k_j \in K_j} \psi(W_{k,j}, \nu_j^2) \right].
\]

Let us start from term \( A_j \). We can rewrite this term as

\[
-\frac{1}{M_j} \sum_{k_j \in K_j} \psi'(W_{k,j}, \nu_j^2) = -\frac{1}{M_j} \sum_{k_j \in K_j} \psi'(W_{k,j}, \nu_j^2) - \frac{1}{M_j} \sum_{k_j \in K_j} \left[ \psi'(W_{k,j}, \nu_j^2) - \psi'(W_{k,j}, \nu_j^2) \right].
\]

Since \( \psi'(W_{k,j}, \nu_j^2) \) is a time-invariant function of \( (W_{k,j}) \), it is also a stationary and ergodic process (see Wooldridge, 1994). Let us define \( m_j = \mathbb{E}[\psi'(W_{k,j}, \nu_j^2)] \), then by Birkhoff’s ergodic theorem we know that

\[
\frac{1}{M_j} \sum_{k_j \in K_j} \psi'(W_{k,j}, \nu_j^2) \xrightarrow{a.s.} \mathbb{E}[\psi'(W_{k,j}, \nu_j^2)] = m_j.
\]

As for term \( C_j \), it is also a stationary and ergodic process. Since \( \hat{\nu}_j^2 \) is a consistent estimator of \( \nu_j^2 \), by (C-7) so is \( \nu_j^2 \) which yields \( \mathbb{E}[C_j] = 0 \) for \( M_j \rightarrow \infty \).

Hence, by again using Birkhoff’s ergodic theorem, we have that

\[
\frac{1}{M_j} \sum_{k_j \in K_j} \left[ \psi'(W_{k,j}, \nu_j^2) - \psi'(W_{k,j}, \nu_j^2) \right] \xrightarrow{a.s.} 0
\]

which finally yields

\[ A_j \xrightarrow{a.s.} m_j. \]

Let us now focus on term \( B_j \) and let us define \( S_{M_j} = \sum_{k_j \in K_j} \psi(W_{k,j}, \nu_j^2) \), with \( \left( \psi(W_{k,j}, \nu_j^2) \right) \) being a stationary ergodic process with \( \mathbb{E}[S_{M_j}] = 0 \) by definition and \( \sigma_{M_j}^2 = \text{Var}[S_{M_j}] \rightarrow \infty \). Then, following Theorem 3 of Denker (1986), we have

\[
\frac{S_{M_j}}{\sigma_{M_j}} \xrightarrow{M_j \rightarrow \infty} \mathcal{N}(0,1).
\]
if $S_{M_j}^2/\sigma_{M_j}^2$ is a uniformly integrable sequence. To show the latter, we use the criterion of Billingsley (2009) for uniformly integrable sequences based on which we need to prove that there exists a $\delta > 0$ such that

$$\sup_{M_j} E \left[ \frac{S_{M_j}^2}{\sigma_{M_j}^2} \right]^{1+\delta} < \infty.$$  

Let us take $\delta = 1$ so that we have

$$\sup_{M_j} E \left[ \frac{S_{M_j}^4}{\sigma_{M_j}^4} \right] < \infty. \quad (C-8)$$

Let us define $Z_{k_j} = \psi(W_{k_j,j}, \nu_j^2)$ and let $Var[Z_{k_j}] = S_{\psi}(0) < \infty$ (i.e. the power spectral density of $(\psi(W_{k_j,j}, \nu_j^2))$ at zero frequency is finite given that $\psi(\cdot)$ is a function of bounded variation). Let us consider the indices $i, j, k, l = 1, \ldots, M_j$, representing four distinct points on the $D$-dimensional wavelet decomposition lattice. We know that the fourth and lower order moments of $Z_{k_j}$ are finite based on condition (C5) since they are bounded functions, hence $E[Z_{k_j}^4]$ is also finite and is bounded by a quantity we denote as $T$. By the Cauchy-Schwarz inequality we have that also $E[Z_{k_j} Z_{k_j'}], E[Z_{k_j}^2 Z_{k_j'}^2], E[Z_{k_j} Z_{k_j'} Z_{k_j'} Z_{k_j}']$ and $E[Z_{k_j} Z_{k_j'} Z_{k_j'} Z_{k_j}']$ are bounded by $T$. Considering the above bound $T$ and defining $a_{k_j}$ as the exponent for the $k_j$th term in the expansion ($a_{k_j} = 0, \ldots, 4, a_{k_j} \in \mathbb{N}$), using the multinomial theorem we have

$$E[S_{M_j}^4] = E\left[ \left( \sum_{k_j \in K_j} Z_{k_j} \right)^4 \right] = E\left[ \sum_{a_1 + \ldots + a_{M_j} = 4} 4! \prod_{a_i} Z_{k_j}^{a_i} \right] \leq T \sum_{a_1 + \ldots + a_{M_j} = 4} 4! \frac{1}{O(M_j^2)}.$$  

Now let us define $C(M_j) = \frac{2}{M_j(M_j-1)} \sum_{k_j \in K_j} \sum_{k_j' < k_j} \rho_{k_j,k_j'}$, with $\rho_{k_j,k_j'}$ being the correlation between $Z_{k_j}$ and $Z_{k_j'}$ and $0 \leq |C(M_j)| \leq 1, \forall M_j$. We then have

$$Var[S_{M_j}]^2 = Var\left[ \sum_{k_j \in K_j} Z_{k_j} \right]^2 = \left( \sum_i Var[Z_i] + 2 \sum_i \sum_{j < i} \rho_{i,j} Var[Z_i] \right)^2 = \left( \sum_i Var[Z_i] + 2 \sum_i \sum_{j < i} C(M_j) Var[Z_i] \right)^2 = (M_j Var[Z_i] + 2M_j(M_j - 1)C(M_j) Var[Z_i])^2 = M_j^2 Var[Z_i]^2 + 4 M_j^2 (M_j - 1) Var[Z_i]^2 C(M_j) + \underbrace{O(M_j^4)}_{O(M_j^2)}.$$
Appendices for Chapter 1

Considering Condition (C-8) and the order of the expressions in $T$, we therefore have

$$\sup_{M_j} \frac{M_j^3 24T}{M_j^2 \text{Var}[Z_{k_j}]^2 + 4M_j^3 \text{Var}[Z_{k_j}]^2 C(M_j) + 4M_j^4 C(M_j)^2 \text{Var}[Z_{k_j}]^2} < \infty.$$  

Hence we have that $\frac{S_{S_{M_j}}^2}{S_{M_j}}$ is uniformly integrable, thereby giving us

$$B_j = \frac{1}{\sqrt{M_j h(M_j)}} \sum_{k_j \in K_j} \psi(W_{k_j, j}, \nu_j^2) \xrightarrow{D_{M_j \to \infty}} N(0, S_{\psi}(0)).$$

where $h(\cdot)$ is a slowly varying function. Using Condition (C4) to fulfill the assumptions of Corollary 5.1 in Hall and Heyde (2014) for term $B_j$ and applying Slutsky’s theorem on both $A_j$ and $B_j$ we finally obtain

$$\sqrt{M_j} (\hat{\nu}_j^2 - \nu_j^2) \xrightarrow{D_{M_j \to \infty}} N\left(0, \frac{S_{\psi}(0)}{m_j^2}\right). \quad (C-9)$$

\[\square\]

A.3.4 Proof of Condition (C7) for Equal Number of Observations per Dimension

Proof. If $K_d = K, \forall d$, then we have that $J_d = J, \forall d, a_d = a, \forall d$ and Condition (C7) therefore becomes

$$\frac{\left(\lfloor \log_2(K) \rfloor - a\right)^D}{\sqrt{(K - L_J + 1)^D}}$$

By the definition of $J$, we know that $L_J = K/2$ and that

$$a = \log_2 \left(\frac{2K(L_1 - 1)}{K + 2L_1 - 4}\right) = \log_2 \left(\frac{2(L_1 - 1)}{1 + \frac{2L_1 - 4}{K}}\right) > 0$$

since $K \geq 2L_1$ and $L_1 \geq 2$ necessarily. We therefore have that Condition (C7) is bounded from above as follows

$$\frac{\left(\lfloor \log_2(K) \rfloor - a\right)^D}{\sqrt{(K - L_J + 1)^D}} \propto \frac{\left(\log_2(K) - a\right)^D}{\sqrt{(K - \frac{K}{2} + 1)^D}} < \frac{2^{D/2}\left(\log_2(K)\right)^D}{K^{D/2}}.$$  

Applying Hopital’s rule $D$ times to the last expression gives us

$$\lim_{K \to \infty} \frac{(D - 1)!}{K^{D/2} D^{(D-1)} \ln(2)^D 2^{D/2}} = 0$$

which implies that Condition (C7) is satisfied when the number of observations in each dimension are equal.

\[\square\]
A.3.5 Mean Square Consistency of $\hat{\nu}$

Below we give the proof of Corollary 1.3.1.

**Proof.** Using the results of Proof A.3.3 we have that

$$\sum_{j} M_j (\hat{\nu}_j^2 - \nu_j^2) \overset{D}{\to} \mathcal{N} \left( 0, \frac{S_\psi(0)}{m_j^2} \right)$$

which by Condition (C4) implies

$$(\hat{\nu}_j^2 - \nu_j^2) = O_p \left( \frac{1}{\sqrt{M_j}} \right).$$

Then for the estimator $\hat{\nu}$ we have

$$\|\hat{\nu} - \nu\| = \sqrt{\sum_{j=1}^{J} (\hat{\nu}_j^2 - \nu_j^2)^2} \leq \sqrt{J \max_j (\hat{\nu}_j^2 - \nu_j^2)^2} = O_p \left( \frac{1}{M_{js}} \right)$$

which by Condition (C7) implies

$$\|\hat{\nu} - \nu\| \overset{p}{\to} 0$$

thus concluding the proof. \qed

A.3.6 Joint Asymptotic Normality of $\hat{\nu}$

Let us now give two technical corollaries which are needed to prove Lemma A.3.2 further on. This lemma is then necessary to obtain the joint asymptotic normality of $\hat{\nu}$ and, consequently, of $\hat{\nu}$.

**Corollary A.3.1.** Let $\hat{\nu}_j^2$ be the implicit solution in $\nu_j^2$ of

$$\sum_{k_j \in K_j} \psi(W_{k_j,\nu_j^2}) = 0. \quad \text{(C-10)}$$

Then, under the conditions of Theorem 1.3.1, we have that

$$\sqrt{M_{js}}(\hat{\nu}_j^2 - \nu_j^2) \overset{D}{\to} \mathcal{N} \left( 0, \bar{\sigma}^2 \right)$$

where $\bar{\sigma}^2 < \infty$ is a known variance, which implies

$$(\hat{\nu}_j^2 - \nu_j^2) = O_p \left( \frac{1}{\sqrt{M_{js}}} \right).$$

**Corollary A.3.2.** Let

$$D_j = \sum_{k_j \in K_j} (\psi'(W_{k_j,\nu_j^2}) - m_j) \quad \text{(C-11)}$$

with $m_j = E[\psi'(W_{k_j,\nu_j^2})]$ and defining $D_{js} = 1/M_{js}D_j$. Then, under the conditions of Theorem 1.3.1 and Condition (C8), we have that

$$\frac{D_{js}}{\text{Var}[D_{js}]} \overset{D}{\to} \mathcal{N} \left( 0, 1 \right)$$

which implies

$$(D_{js} - m_j) = O_p \left( \frac{1}{\sqrt{M_{js}}} \right).$$
Proof. The proof simply uses the results in Proof A.3.3. Indeed, \( \hat{\nu}^2_j \) is another estimator which is different from \( \tilde{\nu}^2_j \) only because it is based on fewer observations (i.e. \( M_{js} \leq M_j, \forall j \)) and therefore converges in distribution at a rate of \( O_p(1/\sqrt{M_j}) \) \( \geq O_p(1/\sqrt{M_{js}}) \). As for term \( D_{js} \), this is a stationary and ergodic process since it is a time-invariant function of \( (W_{k_j,j}) \) and we can therefore define \( Z_{k_j} = (\psi'(W_{k_j,j}, \hat{\nu}^2_j) - m_j) \) as in Proof A.3.3. Under the same conditions and arguments, we therefore have that \( D^2_j/\text{Var}[D_j]^2 \) is a uniformly integrable sequence and we consequently have that \( (D_{js} - m_j) \) is of order \( O_p\left(1/\sqrt{M_{js}}\right)\). 

Using these corollaries, we have the following lemma.

**Lemma A.3.2.** Under the conditions of Corollaries A.3.1 and A.3.2, consider term

\[
A_{js} = \frac{1}{M_{js}} \sum_{k_j \in K_{js}} \psi'(W_{k_j,j}, \nu^2_j)
\]

and \( m_j = \mathbb{E}[\psi'(W_{k_j,j}, \nu^2_j)] \). Then we have that

\[
A^{-1}_{js} - m_j^{-1} = O_p\left(\frac{1}{\sqrt{M_{js}}}\right).
\]

**Proof.** Assuming the conditions for Corollaries A.3.1 and A.3.2 hold and using the definition of \( \hat{\nu}^2_j \) given in (C-10), let us apply a Maclaurin expansion of \( \sum_{k_j \in K_j} \psi(W_{k_j,j}, \hat{\nu}^2_j) \) around \( \nu^2_j \)

\[
\sum_{k_j \in K_j} \psi(W_{k_j,j}, \hat{\nu}^2_j) = \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu^2_j) + \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) (\hat{\nu}^2_j - \nu^2_j).
\]

In addition, let us also apply a Taylor expansion of \( \sum_{k_j \in K_j} \psi(W_{k_j,j}, \hat{\nu}^2_j) \) around \( \nu^2_j \)

\[
\sum_{k_j \in K_j} \psi(W_{k_j,j}, \hat{\nu}^2_j) = \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu^2_j) + \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) (\hat{\nu}^2_j - \nu^2_j) + \sum_{k_j \in K_j} \gamma_{k_j,j}
\]

where \( \gamma_{k_j,j} \equiv O\left((\hat{\nu}^2_j - \nu^2_j)^2\right) = O_p\left(1/M_{js}\right) \) based on Corollary A.3.1.

These expansions imply that

\[
\sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) (\hat{\nu}^2_j - \nu^2_j) = \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) (\hat{\nu}^2_j - \nu^2_j) + \sum_{k_j \in K_j} \gamma_{k_j,j}
\]

which, by rearranging, delivers

\[
\sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) = \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) - \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) = \sum_{k_j \in K_j} \gamma_{k,j} / \nu^2_j - \nu^2_j.
\]

Since we have that \( \gamma_{k,j}/(\nu^2_j - \nu^2_j) = O_p\left(1/\sqrt{M_{js}}\right) \), dividing by \( M_{js} \) on both sides yields

\[
\frac{1}{M_{js}} \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) - \frac{1}{M_{js}} \sum_{k_j \in K_j} \psi'(W_{k_j,j}, \nu^2_j) = O_p\left(\frac{1}{\sqrt{M_{js}}}\right).
\]
Using the results of Corollaries A.3.1 and A.3.2, we have

\[ A_{js} - m_j = \frac{A_{js} - D_{js}}{O_p(\sqrt{M_j})} + \frac{D_{js} - m_j}{O_p(\sqrt{M_j})} = O_p\left(\frac{1}{\sqrt{M_j}}\right) \]

which implies

\[ A_{js}^{-1} - m_j^{-1} = O_p\left(\frac{1}{\sqrt{M_j}}\right). \]

Given the above lemma, below we give the proof of Theorem 1.3.2.

**Proof.** Let us define \( U_j \equiv \tilde{\nu}_j^2 - \nu_j^2 \), where \( \tilde{\nu}_j^2 \) is the solution for \( \nu_j^2 \) in

\[ \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) = 0. \]

From (C-8) we have that

\[ \sqrt{M_j} \left( \tilde{\nu}_j^2 - \nu_j^2 \right) = \sqrt{M_j} U_j = A_{js}^{-1} \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) \]

where \( A_{js} \) is defined as in Lemma A.3.2. Now let \( \sqrt{M_j} Q_j \equiv m_j^{-1} \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) \), let \( a = [a_j]_{j \in J} \) be a vector of constants with at least one element different from zero and finally let \( Q = [Q_j]_{j \in J} \). Then we have

\[ \sqrt{M_j} a^T Q = \sum_{j \in J} a_j m_j^{-1} \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) \]

\[ = \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} \sum_{j \in J} a_j m_j^{-1} \psi(W_{k_j,j}, \nu_j^2) = \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} R_{k_j} \]

where \( R_{k_j} \equiv \sum_{j \in J} a_j m_j^{-1} \psi(W_{k_j,j}, \nu_j^2) \). Being a time-invariant function of \( (W_{k_j,j}) \), we know that also \( (R_{k_j}) \) is a stationary and ergodic process with \( \mathbb{E} \left[ \sum_{k_j \in K_j} R_{k_j} \right] = 0 \) and \( \sigma^2_R \equiv \text{Var} \left[ \sum_{k_j \in K_j} R_{k_j} \right] \overset{M_j \to \infty}{\to} \infty \). By using the same argument as for \( Z_k \) in Appendix A.3.3, we have that \( (\sum_{k_j \in K_j} R_{k_j}^2 / \sigma^2_R) \) is a uniformly integrable sequence. Based on these results and again following Theorem 3 of Denker (1986) we have that

\[ \sqrt{M_j} a^T Q \overset{D}{\to} \mathcal{N} \left( 0, \sigma^2_Q \right). \]

Now let us take \( B_j = \frac{1}{\sqrt{M_j}} \sum_{k_j \in K_j} \psi(W_{k_j,j}, \nu_j^2) \) based on term B in (C-8) and consider

\[ \sqrt{M_j} \sum_{j \in J} a_j (U_j - Q_j) = \sum_{j \in J} a_j \left( \sqrt{M_j} U_j - \sqrt{M_j} Q_j \right) \]

\[ = \sum_{j \in J} a_j \left( A_{js}^{-1} - m_j^{-1} \right) B_j. \]
Since from Lemma A.3.2 we have that $A_{js}^{-1} - m_j^{-1} = \mathcal{O}_p(1/\sqrt{M_j})$ and $B_j$ converges to a Gaussian distribution, we have that

$$\sum_{j \in J} a_j \left( A_{js}^{-1} - m_j^{-1} \right) B_j = \sum_{j \in J} \mathcal{O}_p \left( \frac{1}{\sqrt{M_j}} \right) = \mathcal{O}_p \left( \frac{J}{\sqrt{M_j}} \right)$$

where $J/\sqrt{M_j} \to 0$ by Condition (C7). Therefore we have that $\sqrt{M_j} \sum_{j \in J} a_j (U_j - Q_j) = \mathcal{O}_p(1/\sqrt{M_j})$ and, by finally defining $U = [U_j]_{j \in J}$, we have

$$\sqrt{M_j} a^T U = \sqrt{M_j} \sum_{j \in J} a_j (Q_j + (U_j - Q_j))$$

$$= \sqrt{M_j} a^T Q + \mathcal{O}_p \left( \frac{J}{\sqrt{M_j}} \right) \xrightarrow{\mathcal{D}_{M_j \to \infty}} \mathcal{N}(0, \sigma_Q^2)$$

following Slutsky’s theorem. Given these results and defining $\|s\| = 1$, by the Cramer-Wold theorem we have

$$\sqrt{M_j} s^T \Sigma^{-1/2} (\hat{\nu} - \nu) \xrightarrow{\mathcal{D}_{M_j \to \infty}} \mathcal{N}(0, 1)$$

where, based on (C-9), it can be shown that $\Sigma = M^{-T} S_{\theta}(0) M^{-1}$ with $S_{\theta}(0)$ being the power spectral density of $\Psi(W_k, \nu)$ at zero-frequency and $M = \mathbb{E} \left[ -\frac{\partial}{\partial \nu} \Psi(W_k, \nu) \right]$, thus concluding the proof.

### A.3.7 Joint Asymptotic Normality of $\hat{\nu}$ for $D = 1$

In this appendix we give Theorem A.3.2 and Corollary A.3.3 which are useful results to prove joint asymptotic properties of M-estimators when applied to different sample sizes which are related to each other (as is the case for wavelet decomposition or, for example, for autocovariance sequences at different lags). These results are useful to obtain Theorem 1.3.3 whose proof can be found at the end of this section. For this reason, let us state the following setting:

**(S5)** Let $M$ represent the number of vectorized coefficients $(W_{k,j})$ respecting conditions of Theorem 1.3.1 issued from a given filtering of a random field $(X_k)$ at scale $j$, where the number of observations in the random field $N$ is such that $N > M$. Moreover, let $M^* < M$ and let $\hat{T}$ and $\bar{T}$ represent the same consistent and asymptotically normally distributed M-estimator, for the true parameter value $\theta_0$, computed on $M$ and $M^*$ observations respectively. Moreover, let the convergence of these estimators to a normal distribution be of order $\mathcal{O}_p(1/\sqrt{M})$ and $\mathcal{O}_p(1/\sqrt{M^*})$ for $\hat{T}$ and $\bar{T}$ respectively. Finally, let us define $\Delta = M - M^* > 0$ and assume $M > 2\Delta$, $\forall \Delta$, as well as $\Delta \xrightarrow{M \to \infty} \infty$.

**Theorem A.3.2.** Under Setting (S5) we have that

$$(\hat{T} - \bar{T}) = \mathcal{O}_p \left( \frac{\sqrt{\Delta}}{M} \right).$$
Proof. Let \( \hat{T} \) and \( \bar{T} \) respectively be the implicit solutions, for a parameter \( \theta \), of
\[
\sum_{i=1}^{M} \psi(W_{i,j}, \theta) = 0 \quad \text{and} \quad \sum_{i=1}^{M} \psi(W_{i,j}, \theta) = 0.
\]

Let us now make a Maclaurin expansion of \( \sum_{i=1}^{M} \psi(W_{i,j}, \hat{T}) \) around \( \bar{T} \) as follows
\[
\sum_{i=1}^{M} \psi(W_{i,j}, \hat{T}) = \sum_{i=1}^{M} \psi(W_{i,j}, \bar{T}) + \sum_{i=1}^{M} \psi'(W_{i,j}, T^*)(\hat{T} - \bar{T}) = 0
\]
with \( \|T^* - \bar{T}\| \leq \|\hat{T} - \bar{T}\| \). Rearranging we have
\[
(\hat{T} - \bar{T}) = \left[ -\sum_{i=1}^{M} \psi'(W_{i,j}, T^*) \right]^{-1} \sum_{i=1}^{M} \psi(W_{i,j}, \bar{T})
\]
where
\[
\sum_{i=1}^{M} \psi(W_{i,j}, \bar{T}) = \sum_{i=1}^{M} \psi(W_{i,j}, \hat{T}) + \sum_{i=M-M^*+1}^{M} \psi(W_{i,j}, \bar{T})
\]
which leaves us with
\[
(\hat{T} - \bar{T}) = \frac{\sqrt{\Delta}}{M} \left[ \frac{1}{M} \sum_{i=1}^{M} \psi'(W_{i,j}, T^*) \right]^{-1} \frac{1}{\sqrt{\Delta}} \sum_{i=M-M^*+1}^{M} \psi(W_{i,j}, \bar{T}).
\]

Let us focus on term \( B_j \) and, with \( b = E[\psi(W_{i,j}, \hat{T})] \), let us re-express it as
\[
B_j = \frac{1}{\sqrt{\Delta}} \sum_{i=M-M^*+1}^{M} \left( \psi(W_{i,j}, \bar{T}) - b \right) + \sqrt{\Delta} b.
\]
By using the same conditions and arguments as in Proof A.3.3, we have that
\[
\frac{1}{\sqrt{\Delta}} \sum_{i=M-M^*+1}^{M} \left( \psi(W_{i,j}, \bar{T}) - b \right) \xrightarrow{D} \mathcal{N}(0, \sigma\Delta^2)
\]
for some \( \sigma\Delta^2 < \infty \) thereby implying that \( 1/\sqrt{\Delta} \sum_{i=M-M^*+1}^{M} \left( \psi(W_{i,j}, \bar{T}) - b \right) = O_p(1) \).

Therefore, by Slutsky’s theorem, \( B_j \) is normally distributed with mean \( \sqrt{\Delta} b \). By taking a Taylor expansion of \( b \) around \( \theta_0 \) and using Condition (C6) it is straightforward to verify that \( b \to 0 \) at the rate \( O_p(1/\sqrt{M^*}) \) based on the fact that \( (\hat{T} - \theta_0) = O_p(1/\sqrt{M^*}) \).

This implies that
\[
B_j = \frac{1}{\sqrt{\Delta}} \sum_{i=M-M^*+1}^{M} \left( \psi(W_{i,j}, \bar{T}) - b \right) + \sqrt{\Delta} b \xrightarrow{O_p(1)} O_p\left( \frac{\sqrt{\Delta}}{\sqrt{M^*}} \right).
\]
Using the results in Lemma A.3.2, we have that $A_g = m_g + \mathcal{O}_p(1/\sqrt{\pi})$ where $m_g$ is a constant. Bringing everything together we have

$$
\hat{T} - \bar{T} = \sqrt{\Delta M} \left[ -\frac{1}{M} \sum_{i=1}^{M} \psi'(W_{i,j}, T^*) \right]^{-1} \frac{1}{\sqrt{\Delta}} \sum_{i=M-M^*+1}^{M} \psi(W_{i,j}, \bar{T}).
$$

This implies that when $M > 2\Delta$, we have that

$$
(\hat{T} - \bar{T}) = \mathcal{O}_p\left( \frac{\sqrt{\Delta}}{M} \right).
$$

**Corollary A.3.3.** Let $\Delta = M - M^* > 0$ be fixed and let $\hat{T}$ and $\bar{T}$ be $M$-estimators as defined in Theorem A.3.2. Then we have that

$$
(\hat{T} - \bar{T}) = \mathcal{O}_p\left( \frac{1}{M} \right).
$$

*Proof.* The proof follows directly from that of Theorem A.3.2.

We can now give the proof of Theorem 1.3.3.

*Proof.* Let

$$
\sqrt{M_j s^T \Sigma^{-1/2}} (\hat{\nu} - \nu) = \sqrt{M_j s^T \Sigma^{-1/2}} (\hat{\nu} - \bar{\nu}) + \sqrt{M_j s^T \Sigma^{-1/2}} (\bar{\nu} - \nu)
$$

where $\sqrt{M_j s^T \Sigma^{-1/2}} (\bar{\nu} - \nu) \xrightarrow{D} \text{N}(0, 1)$ by Theorem 1.3.2 and

$$
\sqrt{M_j s^T \Sigma^{-1/2}} (\hat{\nu} - \bar{\nu}) = \sqrt{M_j} \sum_{j=1}^{J} a_j (\hat{\nu}_j^2 - \bar{\nu}_j^2)
$$

where $a_j$ is an element of the constant $1 \times J$ vector $s^T \Sigma^{-1/2}$. If we take the order of this term, by Theorem A.3.2 we have that

$$
\sqrt{M_j s^T \Sigma^{-1/2}} (\hat{\nu} - \bar{\nu}) = \mathcal{O}_p\left( \sqrt{M_j} \sum_{j=1}^{J} \frac{\sqrt{2^j - 2^j}}{M_j} \right)
$$

with $M_j = N - 2^j + 1$. Let us first focus on the term in the sum of this order to obtain

$$
\frac{\sqrt{2^j - 2^j}}{N - 2^j + 1} = \frac{\sqrt{2^j - 2^j}}{N} + \frac{\sqrt{2^j - 2^j(2^j - 1)}}{N M_j} < 2 \frac{\sqrt{2^j - 2^j}}{N}
$$

since $2^{j-1}/M_j < 1$. This leaves us with

$$
\sqrt{M_j} \sum_{j=1}^{J} \frac{\sqrt{2^j - 2^j}}{M_j} < 2 \sqrt{\frac{M_j}{N}} \sum_{j=1}^{J} \sqrt{2^j - 2^j}
$$
and, by using $2^J - 2^j = 2^j(2^{J-j} - 1) < 2^j 2^J$, we have

$$2 \sqrt{\frac{M_j}{N}} \sum_{j=1}^{J} \sqrt{2^J - 2^j} < 2 \sqrt{\frac{M_j}{N}} 2^{j/2} \sum_{j=1}^{J} \sqrt{2^j} = (2^{J/2} - 1)(\sqrt{2} + 2).$$

The order of the latter term is $O_p(2^{J+1} \sqrt{M_j}/N)$ and since $\sqrt{M_j}/N < 1$, we finally have

$$2^{J+1} \sqrt{\frac{M_j}{N}} < \frac{2^{J+1}}{\sqrt{N}} = 2 \frac{N^\alpha}{\sqrt{N}}.$$

We conclude that

$$\sqrt{M_j} s^T \Sigma^{-1/2} (\hat{\nu} - \bar{\nu}) = O_p \left( N^{\alpha - 1/2} \right)$$

which goes to zero since $0 < \alpha < 1/2$. By finally applying Slutsky’s theorem, we obtain

$$\sqrt{M_j} s^T \Sigma^{-1/2} (\hat{\nu} - \nu) \overset{D}{\rightarrow} N \left( 0, 1 \right).$$
Appendix B

Appendices for Chapter 2

B.1 Proofs

B.1.1 Proof of Theorem 2.2.1

Let us consider Model 1 and, to start, let us just consider a sum of independent (T6) processes, which we denote simply as \( Y_t = \sum_{i=1}^{K} X_t^{(i)} \), for which a part of the ACVF sequence \( (\varphi_\theta(h))_{h=0}^{\infty} \) with positive lags is given by

\[
\varphi_\theta(h) = \sum_{i=1}^{K} \rho_i^h \frac{\upsilon_i^2}{1 - \rho_i^2},
\]

where \( \theta = [\rho_1 \upsilon_1^2 \cdots \rho_K \upsilon_K^2] \) represents the parameter vector containing the parameters of the \( K \) processes. The derivatives with respect to \( \rho_i \) and \( \upsilon_i^2 \) (i.e. the parameters of the \( i \)th (T6) process) are respectively

\[
\gamma_1(h) \equiv \frac{\partial}{\partial \rho_i} \varphi_\theta(h) = \frac{h \rho_i^{h-1}(1 - \rho_i^2) + 2 \rho_i^{h+1}}{(1 - \rho_i^2)^2} \upsilon_i^2
\]

\[
= \frac{\rho_i^h (h + (2 - h) \rho_i^2)}{\rho_i(1 - \rho_i^2)^2} \upsilon_i^2
\]

\[
\gamma_2(h) \equiv \frac{\partial}{\partial \upsilon_i^2} \varphi_\theta(h) = \frac{\rho_i^h}{(1 - \rho_i^2)}
\]

which exist based on the parameter values defined for process (T6). These thereby deliver a Jacobian matrix \( A \) whose first 2\( K \) rows are

\[
A = \begin{pmatrix}
\gamma_1(0) & \gamma_2(0) & \cdots & \gamma_1(K) & \gamma_2(K) \\
\gamma_1(1) & \gamma_2(1) & \cdots & \gamma_1(K) & \gamma_2(K) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\gamma_1(H) & \gamma_2(H) & \cdots & \gamma_1(K) & \gamma_2(K)
\end{pmatrix}
\]

where \( H = 2K - 1 \). To simplify notation, let us define \( \gamma_1(h) = [\gamma_1(h)]_{h=0,...,H} \) and \( \gamma_2(h) = [\gamma_2(h)]_{h=0,...,H} \) as being the columns of the matrix \( A \) and \( |A| \) as being the determinant of \( A \). Taking the determinant of the matrix \( A \), we perform some column permutations and operations. First, for more clarity in the proof, we permute the columns to obtain the following matrix determinant

\[
|A| = (-1)^{K-1}|\gamma_1(h) \cdots \gamma_1(K) \gamma_2(h) \cdots \gamma_2(K)|.
\]
Next we multiply each column by a different constant which leaves us with the modified columns
\[
\tilde{\gamma}^i_1(h) \equiv \gamma^i_1(h) \frac{\rho_i(1 - \rho_i^2)^2}{v_i^2} = \rho_i^h(h + (2 - h)\rho_i^2) = \rho_i^h + (2 - h)\rho_i^{h+2}
\]
\[
\tilde{\gamma}^i_2(h) \equiv \gamma^i_2(h)(1 - \rho_i^2) = \rho_i^h
\]
thereby leaving us with the determinant
\[
|A| = (-1)^{K-1} \prod_{i=1}^{K} \left( \frac{v_i^2}{\rho_i(1 - \rho_i^2)^2} \right) |\tilde{\gamma}^1_1(h) \cdots \tilde{\gamma}^K_1(h) \tilde{\gamma}^1_2(h) \cdots \tilde{\gamma}^K_2(h)|
\]
where \( c \neq 0 \). Now let us express \( \tilde{\gamma}^i_1(h) \) as \( \delta^i_1(h) + \delta^i_2(h) \) where \( \delta^i_1(h) = [\rho_i^h]_{h=0,\ldots,H} \) and \( \delta^i_2(h) = [(2 - h)\rho_i^{h+2}]_{h=0,\ldots,H} \). Furthermore, let us define the set \( S = \{a_1, \ldots, a_K\} \) with \( s \in S \) being an element of this set and \( s(i) = a_i \), being the \( i \)th element of \( s \) (note that the cardinality of \( S \), and therefore \( s \), is \( K \)). When using the “determinant as a sum of determinants” rule, we can split each column \( \tilde{\gamma}^i_1(h) \) which, starting with \( \tilde{\gamma}^i_1(h) \), creates a sum of the final nodes of a binary tree as follows
\[
|A| = c(\delta^1_1(h) \tilde{\gamma}^1_1(h) \cdots \tilde{\gamma}^K_1(h) \tilde{\gamma}^1_2(h) \cdots \tilde{\gamma}^K_2(h))
\]
\[
+ (\delta^1_2(h) \tilde{\gamma}^1_2(h) \cdots \tilde{\gamma}^K_2(h) \tilde{\gamma}^1_1(h) \cdots \tilde{\gamma}^K_1(h))
\]
\[
+ (\delta^2_1(h) \tilde{\gamma}^2_1(h) \cdots \tilde{\gamma}^K_1(h) \tilde{\gamma}^2_2(h) \cdots \tilde{\gamma}^K_2(h))
\]
\[
+ (\delta^2_2(h) \tilde{\gamma}^2_2(h) \cdots \tilde{\gamma}^K_2(h) \tilde{\gamma}^2_1(h) \cdots \tilde{\gamma}^K_1(h))
\]
\[
= \ldots \]
\[
= c \sum_{s \in S} |\delta^1_{s(1)}(h) \cdots \delta^K_{s(K)}(h) \tilde{\gamma}^1_2(h) \cdots \tilde{\gamma}^K_2(h)|.
\]
Notice that each column \( \delta^i_{s(i)}(h) \) with \( s(i) = 2 \) (i.e. \( \delta^i_2(h) \)) can be re-expressed as \( \delta^i_2(h) = [(2 - h)\rho_i^{h+2}]_{h=0,\ldots,H} \) and therefore, denoting \( \tilde{\delta}^i_2(h) = [(2 - h)\rho_i^h]_{h=0,\ldots,H} \), we have that the determinant becomes
\[
|A| = c \sum_{s \in S} (\prod_{i=1}^{K} (1_{s(i)=2} - 2\rho_i^2 + 1_{s(i)=1})) |\tilde{\delta}^1_{s(1)}(h) \cdots \tilde{\delta}^K_{s(K)}(h) \tilde{\gamma}^1_1(h) \cdots \tilde{\gamma}^K_1(h)|.
\]
where \( \tilde{\delta}^i_{s(i)}(h) = 1_{s(i)=2} \tilde{\delta}^i_2(h) + 1_{s(i)=1} \delta^i_1(h) \) and \( 1_{s(i)=z} \) represents the indicator function which takes the value 1 if \( s(i) = z \) and 0 otherwise. Following the same procedure as before, let us express \( \tilde{\delta}^i_1(h) = \delta^i_1(h) + \delta^i_2(h) \), where \( \delta^i_3(h) = [2\rho_i^h]_{h=0,\ldots,H} = 2\tilde{\delta}^i_2(h) \) and \( \delta^i_4(h) = [-\rho_i^h]_{h=0,\ldots,H} = -\delta^i_1(h) \). If we split the columns \( \delta^i_2(h) \) into the sub-determinants (in the same way as for \( \tilde{\gamma}^i_1(h) \)) we have a binary tree for which each node has two children of which the one which includes the column \( \delta^i_2(h) \) will be null since this column is a linear function of the column \( \tilde{\gamma}^i_2(h) \). This implies that the only term which remains in each split is \( \delta^i_1(h) \) which is simply the negative of \( \delta^i_1(h) \) and there will be as many \( \delta^i_1(h) \) columns as the original \( \delta^i_2(h) \) columns. Given this structure, we have that the determinant becomes
\[
|A| = c \sum_{s \in S} c_s (-1)^{\sum_{i=1}^{K} (1_{s(i)=2} - 2\rho_i^2 + 1_{s(i)=1})} |\delta^1_1(h) \cdots \delta^K_1(h) \tilde{\gamma}^1_1(h) \cdots \tilde{\gamma}^K_1(h)|.
\]
We therefore have that the determinant $\lambda$ is the determinant of the following matrix

$$
B = \begin{pmatrix}
\rho_1^0 & \rho_2^0 & \cdots & \rho_K^0 & \rho_1^0 & \rho_2^0 & \cdots & \rho_K^0 \\
\rho_1^1 & \rho_2^1 & \cdots & \rho_K^1 & \rho_1^1 & \rho_2^1 & \cdots & \rho_K^1 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\rho_1^H & \rho_2^H & \cdots & \rho_K^H & \rho_1^H & \rho_2^H & \cdots & \rho_K^H 
\end{pmatrix}
$$

which is clearly full column rank. Indeed, the last $K$ columns are those of a Vandermonde matrix with distinct elements $\rho_i \neq \rho_j, \forall i \neq j$, implying that the columns are linearly independent, while the first $K$ columns are a Vandermonde matrix whose rows are multiplied by distinct constants thereby implying that also these columns are linearly independent. Moreover, the first $K$ columns can be seen as functions of the last $K$ columns which cannot however be expressed as a linear combination of the others. Therefore there is no column that can be expressed as a linear combination of the others implying that $\lambda \neq 0$.

As for the term $c_{ss}$, it can be interpreted in a geometric manner since it represents the $K$-dimensional volume of a hyperrectangle with sides $(1 - \rho_i^2)$ given by

$$
c_{ss} = \prod_{i=1}^{K}(1 - \rho_i^2),
$$

where $0 < \prod_{i=1}^{K}(1 - \rho_i^2) < 1$. If we permute the columns so as to respect the order of the parameters in the original vector $\theta$, this finally delivers the determinant

$$
|A| = \prod_{i=1}^{K} \left( \frac{\upsilon_i^2}{\rho_i(1 - \rho_i^2)^2} \right) \lambda \neq 0,
$$

thereby implying that the Jacobian matrix $A$ is of full column rank and, consequently, that the ACVF $\varphi_{\theta}(h)$ of a sum of $K$ (T6) processes is injective. Notice that if we were to add processes (T1) and (T2) to the above procedure, the only detail that would change is the matrix $B$. Indeed, considering the derivatives of their respective ACVFs, if they were inserted in the original matrix $A$ these columns would remain unaffected by the operations thereby delivering the matrix

$$
\tilde{B} = \begin{pmatrix}
\rho_1^0 & \rho_2^0 & \cdots & \rho_K^0 & 1 & 2 \\
\rho_1^1 & \rho_2^1 & \cdots & \rho_K^1 & 0 & -1 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\rho_1^H & \rho_2^H & \cdots & \rho_K^H & 0 & 0 
\end{pmatrix}
$$

which is also clearly full rank with determinant $\tilde{\lambda} \neq 0$. This implies that the ACVF of a latent process made by the sum of $K$ (T6) processes, a (T1) process and a (T2) process is also injective through the parameter vector $\theta$. 

### B.1.2 Proof of injection through an ARMA(2,1)

Let us denote the parameters of an ARMA(2,1) process as $\tilde{\theta} = [\tilde{\rho}_1 \tilde{\rho}_2 \tilde{\upsilon} \tilde{\upsilon}^2]^T$. Given the reparametrization of the sum of two (T6) processes to an ARMA(2,1) process given
in Hamilton (1994) (equation 4.7.26, Section 4.7), we have that the parameters of an ARMA(2,1) process are given by

$$\tilde{\theta} = \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\theta} \\ \hat{\upsilon}_1^2 \end{bmatrix} = \begin{bmatrix} \rho_1 + \rho_2 \\ -\rho_1 \rho_2 / (\upsilon_1^2 + \upsilon_2^2) \\ \rho_1 \rho_2 v_1^2 / (\upsilon_1^2 + \upsilon_2^2) \\ \upsilon_1^2 + \upsilon_2^2 \end{bmatrix} = g(\theta)$$

where \(\theta = [\rho_1 \upsilon_1^2 \rho_2 \upsilon_2^2]^T\) is the vector of parameters for the sum of two (T6) processes. By taking the Jacobian \(A = \partial/\partial \theta g(\theta)\) and using the reasoning given in (2.2.1), we have that the determinant of this matrix is given by

$$|A| = \frac{(\rho_1 - \rho_2)^2}{\upsilon_1^2 + \upsilon_2^2}$$

which is always positive thereby implying that the matrix \(A\) is of full rank and that there is a unique mapping from the parameters of a sum of two (T6) processes to those of an ARMA(2,1) process.

B.1.3 Proof of Theorem 2.2.2

This proof follows the procedure carried out to prove Theorem 2.2.1. Having said this, let us consider Model 3 and let

$$\gamma_{i}^{c}(d) = \sigma_i^2 \exp \left(-\left(\frac{d}{\phi_i}\right)^c\right)$$  \hspace{1cm} (C-1)

be the covariance function of the \(i\)th spatial model with parameters \(\phi_i\) and \(\sigma_i^2\) and where \(d\) is a distance between two point \(k\) and \(k'\), while \(c \in \mathbb{N}^+\) is a known constant. Special cases of this spatial model are the Exponential \((c = 1)\) and Gaussian \((c = 2)\) models. Defining \(\omega_i = \exp(-1/\phi_i)\), the derivatives with respect to \(\phi_i\) and \(\sigma_i^2\) are respectively

$$\alpha_i(d) \equiv \frac{\partial}{\partial \phi_i^2} \gamma_{i}^{c}(d) = \sigma_i^2 \omega_i^c C \frac{d^{c-1}}{c^2} = \sigma_i^2 \omega_i^c \frac{\delta_i^{c+1} \omega_i^c}{\Delta_i^{c+1}}$$

$$\beta_i(d) \equiv \frac{\partial}{\partial \sigma_i^2} \gamma_{i}^{c}(d) = \omega_i^c \omega_i^c.$$ 

With \(\theta = [\theta_1 \cdots \theta_K]^T = [\phi_1 \sigma_1^2 \cdots \phi_K \sigma_K^2]^T\), these derivatives thereby deliver a Jacobian matrix \(A\) whose first \(2K\) rows are

$$A = \begin{pmatrix} \alpha_1(d_0) & \beta_1(d_0) & \cdots & \alpha_K(d_0) & \beta_K(d_0) \\ \alpha_1(d_1) & \beta_1(d_1) & \cdots & \alpha_K(d_1) & \beta_K(d_1) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_1(d_H) & \beta_1(d_H) & \cdots & \alpha_K(d_H) & \beta_K(d_H) \end{pmatrix}$$

where \(H = 2K - 1\) and \(d_i < d_j, \forall i < j\) with \(d_0 = 0\). Let us define the columns of \(A\) as \(\alpha_i(d) = [\alpha_i(d)]_{d=d_0,\ldots,d_H}\) and \(\beta_i(d) = [\beta_i(d)]_{d=d_0,\ldots,d_H}\) respectively such that the determinant of matrix \(A\) is denoted as

$$A = |\alpha_1(d) \beta_1(d) \cdots \alpha_K(d) \beta_K(d)|.$$
Rearranging the columns we obtain

\[ |A| = (-1)^K |\alpha_1(d) \cdots \alpha_K(d) \beta_1(d) \cdots \beta_K(d)|. \]

We know that each column \( \alpha_i(d) \) is multiplied by a constant \( \Delta_i \) such that the matrix determinant becomes

\[ |A| = (-1)^K \prod_{i=1}^{K} \Delta_i \]

where \( \lambda \) therefore represents the determinant of the matrix

\[ B_c(d_0^H) = \begin{pmatrix}
\omega_0^c d_0^c & \cdots & \omega_K^c d_0^c & \omega_0^c & \cdots & \omega_K^c \\
\omega_1^c d_1^c & \cdots & \omega_K^c d_1^c & \omega_1^c & \cdots & \omega_K^c \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\omega_1^c d_H^c & \cdots & \omega_K^c d_H^c & \omega_1^c & \cdots & \omega_K^c \\
\end{pmatrix} \]

which is clearly full column rank by the same argument used for matrix \( B \) in Appendix B.1.1. Therefore there is no column that can be expressed as a linear combination of the others implying that \( \lambda \neq 0 \). Based on this, we know that there is a unique mapping of the parameter vector \( \theta \) to the covariance function \( \gamma_\theta(d) \).

**B.1.4 Proof of Corollary 2.3.2**

For a sum of \( K \) \((T6)\) processes the SDF is given by

\[ S_\theta(f) = \sum_{j=1}^{K} \frac{v_j^2}{1 - 2\rho_j \cos(2\pi f) + \rho_j^2}. \]

If we consider Condition \((C10)\), with \( \rho_j \) and \( v_j^2 \) denoting the elements of the true parameter vector \( \theta_0 \), we have

\[ \sum_{j=1}^{K} \frac{v_j^2}{1 - 2\rho_j \cos(4\pi f) + \rho_j^2} = \sum_{j=1}^{K} \frac{\tilde{v}_j^2}{2(1 - 2\rho_j \cos(2\pi f) + \rho_j^2)} \]

where \( \tilde{\rho}_j \) and \( \tilde{v}_j^2 \) denote the elements of a parameter vector \( \theta_1 \neq \theta_0 \). This condition is clearly not respected since the left side of the equation cannot be expressed as the right side. Indeed, we would need to obtain \( 2(1 - 2\rho_j \cos(2\pi f) + \rho_j^2) \) from the expression \( (1 - 2\rho_j \cos(4\pi f) + \rho_j^2) \) which is not possible. If we add processes \((T1)\) and \((T2)\) to this condition and denote the spectral density of the \( i^{th} \) \((T6)\) process as \( \alpha_i \) we have

\[ \sigma^2 + \frac{4Q^2}{\tau} \sin^2(\pi 2f \tau) + \sum_{i=1}^{K} \alpha_i - \bar{\sigma}^2 - \frac{4\bar{Q}^2}{\tau} \sin^2(\pi 2f \tau) - \sum_{i=1}^{K} \bar{\alpha}_i \]

\[ = \frac{\sigma^2}{2} + \frac{2Q^2}{\tau} \sin^2(\pi f \tau) + \frac{1}{2} \sum_{i=1}^{K} \alpha_i - \frac{\bar{\sigma}^2}{2} - \frac{2\bar{Q}^2}{\tau} \sin^2(\pi f \tau) - \frac{1}{2} \sum_{i=1}^{K} \bar{\alpha}_i. \]
Also in this case, it is straightforward to see that Condition (C10) is not respected either since the spectral density of process (T1) does not depend on $f$ and is constant while that of process (T2) cannot be re-expressed in this form.

## B.1.5 Unique mapping from the covariance function to the spectral density

### Lemma B.1.1.

There is a unique mapping between the covariance function $\varphi_\theta(h)$ and the spectral density $S_\theta(f)$ for a process made by the sum of a (T1) process, a (T2) and $K$ (T6) processes, $\forall K < \infty$.

For there to be a unique mapping from the ACVF to the spectral density $S_\theta(f)$ we need the sequence of autocovariances $(\varphi_\theta(h))_{h=-\infty}^{\infty}$ to be absolutely summable (see, for example, Proposition 6.1 in Hamilton, 1994). Given the parameter values for $\rho_k$, each of the $K$ (T6) processes can be expressed as a causal linear process $Y_t^{(k)} = \sum_{j=0}^{\infty} \psi_j^{(k)} \epsilon_{t-j}$ where $\psi_j^{(k)}$ are fixed coefficients which respect $\sum_{j=0}^{\infty} |\psi_j^{(k)}| < \infty$ and $\epsilon_{t-j} \sim \mathcal{N}(0, \kappa_k^2)$. Their sum gives

$$\sum_{k=1}^{K} \sum_{j=0}^{\infty} \psi_j^{(k)} \epsilon_{t-j} = \sum_{k=1}^{K} \sum_{j=0}^{\infty} \psi_j^{(k)} \epsilon_{t-j}$$

and, given that the processes are independent by condition (C1), we have that

$$\sum_{k=1}^{K} \psi_j^{(k)} (\epsilon_{t-j}) \sim \mathcal{N} \left( 0, \sum_{k=1}^{K} \psi_j^{(k)} \kappa_k^2 \right).$$

We have that $\delta^2$ can be re-expressed as

$$\delta^2 = \left( \psi_j^{(1)} \right)^2 \kappa_1^2 + \kappa_1^2 \sum_{k=2}^{K} \frac{\kappa_k^2}{\kappa_1^2} (\psi_j^{(k)})^2$$

$$= \left( \psi_j^{(1)} \right)^2 + \sum_{k=2}^{K} \frac{\kappa_k^2}{\kappa_1^2} (\psi_j^{(k)})^2 \kappa_1^2.$$

We therefore have the equivalence of processes

$$\sum_{k=1}^{K} \psi_j^{(k)} (\epsilon_{t-j}) \iff \tilde{\psi}_j u_{t-j} \sim \mathcal{N}(0, \kappa_1^2)$$

which implies that (C-2) can be written as $\sum_{j=0}^{\infty} \tilde{\psi}_j u_{t-j}$. If we define $z = \begin{bmatrix} \psi_j^{(1)} & \psi_j^{(2)} & \cdots & \psi_j^{(K)} \end{bmatrix}$, this gives us

$$\sum_{j=0}^{\infty} |\tilde{\psi}_j| = \sum_{j=0}^{\infty} \sqrt{ (\psi_j^{(1)})^2 + \sum_{k=2}^{K} \frac{\kappa_k^2}{\kappa_1^2} (\psi_j^{(k)})^2 } = \sum_{j=0}^{\infty} \|z\|_2 

\leq \sum_{j=0}^{\infty} \|z\|_1 = \sum_{j=0}^{\infty} |\psi_j^{(1)}| + \frac{\kappa_2}{\kappa_1} \sum_{j=0}^{\infty} |\psi_j^{(2)}| + \cdots + \frac{\kappa_K}{\kappa_1} \sum_{j=0}^{\infty} |\psi_j^{(K)}| < \infty.$$
where \( \|z\|_2 \) and \( \|z\|_1 \) represent the \( L^2 \) norm and \( L^1 \) norm of \( z \) respectively. This result implies that a sum of \( K \) \((T6)\) processes have absolutely summable linear coefficients \((\tilde{\psi}_j)_{j=0}^{\infty}\) which implies that its autocovariances \((\varphi_\theta)(h))_{h=-\infty}^{+\infty}\) are also absolutely summable. If we add the processes \((T1)\) and \((T2)\) to this reasoning, we have that their ACVF are given by

\[
\varphi_{\sigma^2}(h) = \begin{cases} 
\sigma^2 & \text{if } h = 0 \\
0 & \text{if } h > 0 
\end{cases}
\]

and

\[
\varphi_{Q^2}(h) = \begin{cases} 
2Q^2 & \text{if } h = 0 \\
-Q^2 & \text{if } h = 1 \\
0 & \text{if } h > 1 
\end{cases}
\]

where \( \varphi_{\sigma^2}(h) \) and \( \varphi_{Q^2}(h) \) represent the ACVF of the \((T1)\) and \((T2)\) process respectively.

If added to the ACVF of the sum of \( K \) \((T6)\) processes described earlier, it is straightforward to see that the resulting ACVF sequence is absolutely summable as well, thus concluding the proof.

**Lemma B.1.2.** There is a unique mapping between the covariance function \( \varphi_\theta(h) \) and the spectral density \( S_\theta(f) \) for a process made by the sum of a \((T5)\) process and \( K \) \((T6)\) processes, \( \forall K < \infty \).

The proof uses the same arguments as the proof of Lemma B.1.1.

**B.1.6 Proof of Lemma 2.3.2**

The proof is straightforward for both classes of processes. Considering the process \( Y_t = \sum_{i=1}^{4} X_i^{(i)} \), if we take the first four consecutive WV scales (i.e. \( \nu(\theta_1) = [\nu_1^2(\theta_1), \ldots, \nu_4^2(\theta_1)] \)), we have that the determinant of the relative Jacobian matrix is \( |A_{4,4}| = \frac{2205\omega}{4096} \) which implies that it is of full rank. Considering the other process \( Y_t = \sum_{i=3}^{5} X_i^{(i)} \), with the same approach as the proof of the first process, if we take the first four consecutive WV scales (i.e. \( \nu(\theta_1) = [\nu_2^2(\theta_1), \ldots, \nu_4^2(\theta_1)] \)), we have that the determinant of the relative Jacobian matrix is \( |A_{4,4}| = -\frac{2205\omega\rho}{256} \) which implies that it is of full rank.
Appendix C

Appendices for Chapter 3

C.1 Short literature review

A detailed discussion on robust estimation and inference methods for time series models can be found in Maronna et al. (2006), Chapter 8. Most of the literature in this domain has dealt with standard time series models such as autoregressive and/or moving average models. Künsch (1984) proposes optimal robust estimators of the parameters of autoregressive processes by studying the properties of their influence function (see also Martin and Yohai, 1986). Denby and Martin (1979) develop a generalized M-estimator for the parameter of a first-order autoregressive process whereas Bustos and Yohai (1986), Allende and Heiler (1992) and de Luna and M. G. Genton (2001) extend the research to include moving average models using generalized M-estimation theory and indirect inference (see e.g. Gourieroux et al., 1993). Bianco et al., 1996 propose a class of robust estimators for regression models with ARIMA errors based on $\tau$-estimators of scale (Yohai and Zamar, 1988). Ronchetti and Trojani (2001) develop a robust version of the generalized method of moments (proposed by Hansen, 1982) for estimating the parameters of time series models in economics, and Ortelli and Trojani (2005) further develop a robust efficient method of moments. Mancini et al. (2005) propose optimal bias-robust estimators for a class of conditional location and scale time series models while La Vecchia and Trojani (2010) develop conditionally unbiased optimal robust estimators for general diffusion processes, for which approximation methods for computing integrals are needed. Cizek (2008) studies the properties of a two-step least weighted squares robust time-series regression estimator and Agostinelli and Bisaglia (2010) propose a weighted maximum likelihood estimator for ARFIMA processes, for which Molinares et al. (2009) propose an alternative estimator under additive outliers. Sarnaglia et al. (2010) suggest a robust estimation procedure for the parameters of the periodic AR model as an extension of the robust scale and covariance functions given in, respectively, Rousseeuw and Croux (1993) and Ma and M. Genton (2000).

Another means to obtain robust estimators for the parameters of a time series model when it can be written as a state-space model is by means of robust (Kalman) filtering. Robustification of the Kalman filter was originated with Masreliez and Martin (1977) and Cipra (1992) who propose robust modifications of exponential smoothing (see also Cipra and Hanzak, 2011 and Croux et al., 2010 for a multivariate version). For a robust version of the Holt-Winters smoother, see Gelper et al., 2010. Muler et al. (2009) develop a class of robust estimates for ARMA models that are closely related to robust filtering. Robust filtering can also possibly provide a way to robustly estimate the WV (although stronger
assumptions on the underlying model would have to be made). However, in this case, the only attempt to studying the robustness properties of wavelet filtering has been made in the identically and independently distributed (iid) case where Renaud (2002) develops, among others, the IF of the Haar-based wavelet coefficients and concludes that the IF depends on the location of the contaminated data with respect to the dyadic grid and can be infinite. As in the case of the wavelet coefficients, many classical filtering methods are unbounded and for this reason several robust local filters have been proposed so far since the median filter proposal from Tukey (1977): Bruce et al. (1994) pre-process the estimation of the wavelet coefficients via a “fast and robust smooth/cleaner”; Krim and Schick (1999) derive a robust estimator of the wavelet coefficients based on minimax description length; Härdle and Gasser (1984) develop a locally weighted smoothing using M-estimation and Fried et al. (2007) propose a non-parametric, weighted repeated median filter. Sardy et al. (2001) propose a robust wavelet-based estimator using a robust loss-penalized function, for which appropriately choosing the smoothing parameter is an important robustness issue as revealed, for example, by Cantoni and Ronchetti (2001).

The literature specifically on robust estimation of spatial models is less abundant. An overview of some robust methods which are not necessarily directly linked to robust spatial model estimation can be found in Anselin (2013). In O. F. Christensen et al. (2006) a robust Markov-Chain approach is discussed for spatial generalized linear mixed models whereas a few methods have been proposed to robustly estimate the variogram (or semi-variogram) of a spatial process and some examples can be found in Cressie (1985) or M. G. Genton (1998). Ideally, these can be used as robust auxiliary parameters in a minimum-distance estimator approach. Literature dealing with statistical robustness in the estimation of higher-dimensional models is even more scarce, if not nonexistent, to the best of our knowledge.
C.2 Proofs

C.2.1 RGMWM Asymptotic Properties

The results of the proofs in this appendix largely follow the conditions and steps for standard extremum estimators (see Newey and McFadden, 1994). However, since we consider cases where we let \( J \to \infty \) (i.e. the number of moment conditions goes to infinity with the sample size), we nevertheless illustrate the steps and details which allow these standard conditions to hold also in the RGMWM setting. In the following proofs \( \| \cdot \| \) indicates the Frobenius norm if the object is a matrix.

Proof. Given conditions (C1) to (C3), let \( Q(\theta) = (\nu(\theta_0) - \nu(\theta))^T \Omega (\nu(\theta_0) - \nu(\theta)) = \| \nu(\theta_0) - \nu(\theta) \|^2_\Omega \) and \( Q_N(\theta) = (\hat{\nu} - \nu(\theta))^T \hat{\Omega} (\hat{\nu} - \nu(\theta)) = \| \hat{\nu} - \nu(\theta) \|^2_\hat{\Omega} \) and define \( \Omega^* = \hat{\Omega} - \Omega \). By Theorem 2.1. of Newey and McFadden (1994) we want to prove that \( Q_N(\theta) \) converges uniformly in probability to \( Q(\theta) \). By the triangular inequality and Cauchy-Schwartz we have that

\[
|Q_N(\theta) - Q(\theta)| \leq \left\| \| \nu(\theta_0) - \nu(\theta) \|^2_\Omega \right\|_{a_1} + \| \nu(\theta_0) - \nu(\theta) \|^2_\Omega \right\|_{a_2} + \left\| 2(\nu(\theta_0) - \nu(\theta))^T \hat{\Omega} (\hat{\nu} - \nu(\theta)) \right\|_{a_3}.
\]

Considering term \( a_1 \) and using the same inequalities, we have

\[
a_1 \leq \| \nu(\theta_0) - \nu(\theta) \|^2 \| \Omega^* \| = \sum_{j=1}^{J} (\nu_j(\theta_0) - \nu_j(\theta))^2 \left\| \sum_{i=1}^{J} \sum_{j=1}^{J} (\omega_{i,j} - \omega_{i,j})^2 \right\|.
\]

By (C1) and (C6) we have that \( (\nu_j(\theta_0) - \nu_j(\theta))^2 \) is bounded by a quantity \( B \) and

\[
\sqrt{\sum_{i=1}^{J} \sum_{j=1}^{J} (\omega_{i,j} - \omega_{i,j})^2} \leq J \sqrt{(\omega_{i,j} - \omega_{i,j})^2} = O_p \left( \frac{J}{g(N)} \right)
\]

thereby giving

\[
a_1 \leq JB O_p \left( \frac{J}{g(N)} \right) = O_p \left( \frac{J^2}{g(N)} \right)
\]

which, by (C6), tends to 0 in probability. Moreover, using the results in Chapter 1, we have that

\[
a_2 \leq \| \hat{\Omega} \| \| \nu(\theta_0) - \nu(\theta) \| \| \hat{\nu} - \nu(\theta_0) \| = O_p \left( \frac{J}{g(N)} \right) \| \nu(\theta_0) - \nu(\theta) \| \| \hat{\nu} - \nu(\theta_0) \| = O_p \left( \frac{J}{\sqrt{M_j^*}} \right)
\]

Finally, we have

\[
a_3 \leq 2 \| \hat{\Omega} \| \| \nu(\theta_0) - \nu(\theta) \| \| \hat{\nu} - \nu(\theta_0) \|.
\]

By (C6) we have that \( \| \hat{\Omega} \| \leq \lambda \) and by (C1) we have \( \| \nu(\theta_0) - \nu(\theta) \| = O_p(\sqrt{J}) \). Moreover, given the results in Chapter 1 we have \( \| \hat{\nu} - \nu(\theta_0) \| = O_p(\sqrt{J}) \) which gives

\[
a_3 \leq \lambda O_p(\sqrt{J}) O_p \left( \frac{J}{\sqrt{M_j^*}} \right) = O_p \left( \frac{J}{\sqrt{M_j^*}} \right).
\]
Therefore, we have that
\[ \sup_{\theta \in \Theta} |Q_N(\theta) - Q(\theta)| \xrightarrow{P} 0. \]

Based on this and using conditions (C1), (C3), (C4), we finally have
\[ \hat{\theta} \xrightarrow{P} \theta_0 \]
thus concluding the proof.

**Proof.** Given the results on the consistency in Proposition 3.2.1, the proof of asymptotic normality of \( \hat{\theta} \) naturally follows the standard proof of asymptotic normality for GMM estimators. Here we therefore simply give a quick overview of the steps which are based on existing results (see Newey and McFadden, 1994). Let \( Q(\theta) \equiv (\nu(\theta_0) - \nu(\theta))^T \Omega(\nu(\theta_0) - \nu(\theta)) = \|\nu(\theta_0) - \nu(\theta)\|_\Omega^2 \) and \( Q_N(\theta) \equiv (\hat{\nu} - \nu(\theta))^T \hat{\Omega} \hat{\nu} - \nu(\theta)) = \|\hat{\nu} - \nu(\theta)\|_{\hat{\Omega}}^2 \). Moreover, let us define \( \hat{\Omega}_N = \frac{1}{2}(\hat{\Omega} + \hat{\Omega}^T) \). Finally, using condition (C8), let us denote \( B_N(\theta) = \partial_{\theta^T} Q_N(\theta) \), with \( B(\theta) \) being the counterpart for \( Q(\theta) \), \( H_N(\theta) = \partial_{\theta^T} B_N(\theta) \) and \( D(\theta) = \partial_{\theta^T} \nu(\theta) \). Since we have that \( Q_N(\hat{\theta}) \) minimizes \( Q_N(\theta) \) by definition, we consequently have that
\[ B_N(\hat{\theta}) = 0. \]

By taking a Maclaurin expansion of \( B_N(\hat{\theta}) \) around \( \theta_0 \) we have
\[ B_N(\hat{\theta}) = B_N(\theta_0) + H_N(\theta^*)(\hat{\theta} - \theta_0) \]
where \( \|\theta^* - \theta_0\|^2 \leq \|\hat{\theta} - \theta_0\|^2 \). Using the two equalities, rearranging and multiplying by \( \sqrt{N} \) give us
\[ \sqrt{N}(\hat{\theta} - \theta_0) = (-H_N(\theta^*)^{-1})(\sqrt{NB}_N(\theta_0)). \]

Let us focus on term \( H_N(\theta^*) \) and let us now take a Taylor expansion of \( B_N(\hat{\theta}) \) around \( \theta_0 \) which gives
\[ B_N(\hat{\theta}) = B_N(\theta_0) + H_N(\theta_0)(\hat{\theta} - \theta_0) + O(\|\hat{\theta} - \theta_0\|^2). \]

Taking the difference with the Maclaurin expansion in (C1) implies that
\[ H_N(\theta^*) - H(\theta_0) = O_p(\|\hat{\theta} - \theta_0\|) \]
and hence, by continuity of matrix inversion, we have
\[ -H_N(\theta^*)^{-1} \xrightarrow{P} H(\theta_0)^{-1}. \]

Using (C7) to (C8), we finally obtain
\[ \sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, BVB^T) \]
with \( B = H(\theta_0)^{-1}D(\theta_0)^T \Omega \), thus concluding the proof.

**C.3 Additional Results from Simulation Studies**

Here we report some additional results from the simulation studies made in Section 3.3. Table C.1 gives the estimation times in seconds and convergence rates of the RGMWM and INDI estimators under an uncontaminated setting.
C.4. Additional Results from Application on Inertial Sensor Data

In this appendix we find other results for the application on the inertial sensor measurements studied in Section 3.4.1. Figure C.1 shows the WV plots which are in the same spirit of those provided for the application in Section 3.4.2 and are interpreted in the same manner.

It can be seen how the different estimations are significantly different at the lower...
scales where the outliers have more of an influence on estimations. The model-implied WV is plotted based on the parameter estimates of the GMWM and RGMWM which can be found in Table C.2 along with their respective confidence intervals. Given the length of the signal, it can be seen how some estimates are the same between methods and the confidence intervals are extremely tight, appearing to be identical to the estimate itself due to the rounding (see, for example, the estimates and confidence intervals for $\rho_1$ and $\rho_2$). Considering these details, we can remark that the parameters of the first two autoregressive processes are significantly different from each other underlining that although the percentage of outliers is considerably low, the contamination appears to have an impact on the estimation process and robust methods should therefore be preferred.
Table C.2: State-space model estimates for the gyroscope data in static conditions. Estimated parameters with GMWM and RGMWM estimators with $\rho_i$ being the $i^{th}$ autoregressive parameter, $\nu^2_i$ the innovation variance of the $i^{th}$ autoregressive model. Confidence intervals (CI) based on the approach used in Guerrier et al. (2013b).

<table>
<thead>
<tr>
<th></th>
<th>GMWM</th>
<th>RGMWM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>CI($\cdot$, 95%)</td>
<td>Estimate</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>$8.9546 \cdot 10^{-2}$</td>
<td>$1.4816 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$\nu^2_1$</td>
<td>$(8.9546 \cdot 10^{-2} ; 8.9546 \cdot 10^{-2})$</td>
<td>$(1.4816 \cdot 10^{-1} ; 1.4816 \cdot 10^{-1})$</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>$7.8760 \cdot 10^{-5}$</td>
<td>$5.5325 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>$\nu^2_2$</td>
<td>$(7.8612 \cdot 10^{-5} ; 7.8909 \cdot 10^{-5})$</td>
<td>$(5.5211 \cdot 10^{-5} ; 5.5439 \cdot 10^{-5})$</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>$9.9831 \cdot 10^{-1}$</td>
<td>$9.9687 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$\nu^2_3$</td>
<td>$(9.9831 \cdot 10^{-1} ; 9.9831 \cdot 10^{-1})$</td>
<td>$(9.9687 \cdot 10^{-1} ; 9.9687 \cdot 10^{-1})$</td>
</tr>
<tr>
<td></td>
<td>$(2.0887 \cdot 10^{-10} ; 3.0377 \cdot 10^{-10})$</td>
<td>$(1.0466 \cdot 10^{-9} ; 9.4291 \cdot 10^{-10} ; 1.1504 \cdot 10^{-9})$</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>$9.9997 \cdot 10^{-1}$</td>
<td>$9.9997 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$\nu^2_3$</td>
<td>$(9.9995 \cdot 10^{-1} ; 9.9999 \cdot 10^{-1})$</td>
<td>$(9.9995 \cdot 10^{-1} ; 9.9999 \cdot 10^{-1})$</td>
</tr>
<tr>
<td></td>
<td>$(7.5299 \cdot 10^{-12} ; 1.6300 \cdot 10^{-11})$</td>
<td>$(1.3626 \cdot 10^{-11} ; 8.7412 \cdot 10^{-12} ; 1.8511 \cdot 10^{-11})$</td>
</tr>
</tbody>
</table>
Appendix D

Appendices for Chapter 5

D.1 Benchmarks

One of the driving design principles behind the proposed \texttt{gmwm} platform is its computational efficiency. To achieve this goal, the computational backend is written completely in \texttt{C++} using the Armadillo matrix library (see C. Sanderson, 2010). The implementation of each function is highly efficient due to the manner in which they are implemented. Specifically, multiple functions for the same task were created within \texttt{C++} and then benchmarked. The benchmarking was conducted using the “\texttt{rbenchmark}” package (\texttt{rbenchmark}) with each function being called 100 times. The function that used the least amount of time among the \texttt{C++} implementations was then included within the package. This bottom-up approach used to create the package not only delivers a very efficient implementation of the GMWM estimator but also of different wavelet-based methods and random process generation. To illustrate, we provide details on the overall computational time across a wide range of sample sizes to estimate a \texttt{3*GM()} model under both the standard and robust settings in two different modes: user-supplied and guessed parameters. The user-supplied parameters were taken to be the exact parameters whereas the guessed parameters relied on the initial starting algorithm described in Appendix D.3. The latter algorithm was based on a range of initial guesses going from 100 to 1,000,000. The time reported in Tab. D.1 and D.2 is by the number of \textit{seconds} required to estimate the model.

<table>
<thead>
<tr>
<th>Type</th>
<th>100</th>
<th>1,000</th>
<th>10,000</th>
<th>100,000</th>
<th>1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact Values</td>
<td>0.0100</td>
<td>0.0188</td>
<td>0.0300</td>
<td>0.1408</td>
<td>1.5473</td>
</tr>
<tr>
<td>Guessed 1,000</td>
<td>0.0342</td>
<td>0.0369</td>
<td>0.0594</td>
<td>0.1714</td>
<td>1.5965</td>
</tr>
<tr>
<td>Guessed 10,000</td>
<td>0.0755</td>
<td>0.0794</td>
<td>0.1250</td>
<td>0.2511</td>
<td>1.6647</td>
</tr>
<tr>
<td>Guessed 20,000</td>
<td>0.1216</td>
<td>0.1386</td>
<td>0.2005</td>
<td>0.3411</td>
<td>1.7917</td>
</tr>
</tbody>
</table>

Table D.1: Standard estimation of a \texttt{3*GM()} model using “\texttt{gmwm.imu()}” in seconds

<table>
<thead>
<tr>
<th>Type</th>
<th>1,000</th>
<th>10,000</th>
<th>100,000</th>
<th>1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact Values</td>
<td>0.0327</td>
<td>0.0816</td>
<td>0.6839</td>
<td>11.4097</td>
</tr>
<tr>
<td>Guessed 1,000</td>
<td>0.0546</td>
<td>0.1081</td>
<td>0.6875</td>
<td>11.4575</td>
</tr>
<tr>
<td>Guessed 10,000</td>
<td>0.1135</td>
<td>0.1791</td>
<td>0.7687</td>
<td>11.5385</td>
</tr>
<tr>
<td>Guessed 20,000</td>
<td>0.1666</td>
<td>0.2516</td>
<td>0.8556</td>
<td>11.5957</td>
</tr>
</tbody>
</table>

Table D.2: Robust estimation of a \texttt{3*GM()} model using “\texttt{gmwm.imu()}” in seconds
D.2 How to Import IMU Data

The first step towards the modelling of the inertial sensor stochastic error is loading the data onto the platform. This is done by:

1. Loading data in .txt or .csv format;
2. Reading in IMU binary files.

In the first case, the user must also cast the data as an “imu” object whereas this is automatically done when loading the data in the second case. Below is an example that displays each of these cases.

```r
# Case 1a: Reads in data that is separated by tab
ds = read.table("path/to/file.txt")

# Case 1b: Reads in data that is separated by comma
ds = read.csv("path/to/file.csv")

# Case 2: Cast data loaded into R to IMU object-type
sensors = imu(ds, accelerometer = 1:3, gyroscope = 1:3, freq = 100)

# Case 3: Read an imu binary file and cast object as an IMU-type
sensors = read.imu("~/path/to/file.imu","IMU_TYPE")
```

The binary file reader currently supports the binary record format used by well known commercial software such as Applanix (PosProc) and Novatel/Waypoint (IEplorer). This format is aptly described as a seven column layout where the first column of the data contains time data, columns 2-4 contain gyroscope values and columns 5-7 contain accelerometer values for axes X, Y, and Z. Upon loading, appropriate scaling factors are applied to the data. As a result, the following types of IMU can be loaded via their binary: IMAR, LN200, LN200IG, IXSEA, NAVCHIP_INT and NAVCHIP_FLT. For more details, please see the `read.imu()` help documentation.

D.3 Optimization Starting Values Algorithm

To start any optimization procedure starting values are needed and this is the case also for the GMWM procedure with the optimization in (5.2.3). The platform uses an algorithm to find appropriate starting values and this is described in three parts below: a general overview of the algorithm (Algorithm 1), a heuristic approach to determining the dominating process (Algorithm 2), and a focus on a specific step based on the required processes to be estimated (Algorithm 3).

D.4 Code Used in Chapter

The code further on was used to generate some of the results and figures included in the chapter and provides insight to the use of the “gmwm” platform. Please note that the following code was executed under version 3.0.0 of both the “gmwm” and “imudata” R package.
**D.4. Code Used in Chapter 143**

**Inputs:** An empty parameter vector $\theta$ that defines the model, the wavelet variance for the first two scales, and the slope of the data $(\min(x) - \max(x))$.

**Output:** Starting values to estimate the parameter vector $\theta$.

**Step 1:** Compute the total variance $\hat{\sigma}^2_T = \sum_{j=1}^{J} \hat{\nu}^2_j$ for all scales obtained with Equation 5.2.1.

**Step 2:** Determine the process that dominates the initial scales using Algorithm 2.

**Step 3a:** Randomly sample $G$ times from the parameter spaces depending on the type of process. At the start of each new round of guessing for all parameters, use the domination result from **Step 2** to select the starting condition of the “GM” / “AR1” process:

- **if QN dominates** then
  - Draw $R \sim U[0, 1]$  
  - if $R \leq 0.75$ then
    - Start AR1 or GM on condition 2 in Algorithm 3

- **else if WN dominates** then
  - Start AR1 or GM on condition 2 in Algorithm 3

**Step 3b:** See Algorithm 3 for details on parameter selection for each process.

**Step 4:** Select the starting parameter vector $\hat{\theta}^{(0)}$ with the smallest objective function value given by the flattening algorithm described in Stebler et al., 2014c:

$$
\hat{\theta}^{(0)} = \arg\min_{\theta \in \Theta} \left( 1 - \frac{\nu(\theta)}{\hat{\nu}} \right)^T \left( 1 - \frac{\nu(\theta)}{\hat{\nu}} \right) \quad (D-1)
$$

Algorithm 1: Starting Values Algorithm

**Inputs:** Process’ wavelet variance, and the slope of the data $(\min(x) - \max(x))$.

**Output:** Process that dominates the initial scales.

**Step 2a:** Compute the slope, $s$, between the first wavelet variance, $\hat{\nu}_1$, and the upper and lower confidence bounds of the second, $a = \left[ \frac{\hat{\eta}_2}{\chi^2_4(0.975)}, \frac{\hat{\eta}_2}{\chi^2_4(0.025)} \right]$, using $s = \frac{\log(\hat{\nu}) - \log(\hat{\nu}_1)}{\log(4)}$, where $\hat{\eta} = \max \left( (N - L_j + 1) / 2^j, 1 \right)$.

**Step 2b:** Determine whether QN, WN, or AR1/GM dominates the initial scales based on a slope heuristic:

- **if** $\text{max}(s) < -.5$ **then**
  - QN Dominates
- **else if** $\text{min}(s) > -.5$ **then**
  - AR1/GM Dominates
- **else**
  - WN Dominates

Algorithm 2: Process Domination Algorithm
**Inputs:** An empty parameter vector $\theta$ that defines the model, the process’ wavelet variance, and the slope of the data.

**Output:** A parametric model $F_\theta$

**Step 3b:** Select the appropriate realization generator based on process type

- **if AR1 or GM then**
  - **if AR/GM on condition 1 then**
    - Draw $U \sim U(0, 1/3)$ for $\phi = \frac{1 - \sqrt{1 - 3U}}{3}$
    - Draw $\sigma^2 \sim U\left(\frac{\sigma_T^2(1-\phi)^2}{2}, \sigma_T^2(1 - \phi)^2\right)$
  - **else if AR/GM on condition 2 then**
    - Draw $\phi_i \sim U\left(\max(0, 0.9, \phi_{i-1}), 0.999995\right)$
    - Draw $\sigma^2 \sim U\left(0, \frac{\hat{\sigma}_T^2(1-\phi_i)^2}{100}\right)$
  - else
    - Draw $U \sim U(0, 1/3)$ for $\phi_i = (0.999995 - \phi_{i-1}) \left(\sqrt{1 - 3U} + \phi_{i-1}\right)$
    - Draw $\sigma^2 \sim U\left(\frac{\sigma_T^2(1-\phi_i)^2}{2}, \sigma_T^2(1 - \phi_i)^2\right)$
  - Increase the condition by 1.

- **else if DR then**
  - Calculate the slope of the data: $R = \frac{\max(x) - \min(x)}{\bar{x}}$
  - Draw $\omega \sim U\left(\frac{R}{100}, \frac{R}{10}\right)$

- **else if RW then**
  - Draw $\gamma^2 \sim U\left(\frac{\hat{\sigma}_T^2}{10\gamma^2}, \frac{\hat{\sigma}_T^2}{N}\right)$

- **else if WN then**
  - **if WN Dominates then**
    - Draw $\sigma^2 \sim U\left(\frac{\hat{\sigma}_T^2}{2}, \frac{\hat{\sigma}_T^2}{4}\right)$
  - **else**
    - Draw $\sigma^2 \sim U\left(\frac{\hat{\sigma}_T^2}{100}, \frac{\hat{\sigma}_T^2}{10}\right)$

- **else if QN Dominates then**
  - Draw $Q^2 \sim U\left(\frac{\hat{\sigma}_T^2}{8}, \frac{\hat{\sigma}_T^2}{3}\right)$
  - **else**
    - Draw $Q^2 \sim U\left(\frac{\hat{\sigma}_T^2}{2\cdot10^7}, \frac{\hat{\sigma}_T^2}{100}\right)$

**Algorithm 3:** Process-specific Starting Values Algorithm
# Load R Package
library("gmwm")

# Load in MTIG data. If the data package is not installed, # then install it.
if(!require("imudata")){
  install_imudata()
  library("imudata")
}

# Read in Binary IMU Data
navchip = read.imu("~/IMU_data/navchip_4h.imu","NAVCHIP_FLT")

# 1. WAV-AR All Navchip
# Obtain the wavelet variance of all sensors # in imu object.
wv.navchip = wvar(navchip)

# Plot the Wavelet Variance of each sensor
plot(wv.navchip)

# 2. Generate models for each sensor
navchip.gx = gmwm.imu(2*GM() + WN() + DR() + QN() + RW(),
  navchip[,1])

navchip.gy = gmwm.imu(3*GM() + WN() + QN() + RW(),
  navchip[,2])

navchip.gz = gmwm.imu(2*GM() + WN() + DR() + QN() + RW(),
  navchip[,3])

navchip.ax = gmwm.imu(4*GM() + QN() + RW(), navchip[,4])

navchip.ay = gmwm.imu(4*GM() + QN() + RW(), navchip[,5])

navchip.az = gmwm.imu(3*GM() + QN(), navchip[,6])

# Robust model estimation
navchip.gy.r = gmwm.imu(3*GM() + WN() + QN() + RW(),
  navchip[,2], robust = T)

# Summary statistics
sm.navchip.gy = summary(navchip.gy, inference = T)
sm.navchip.gy.r = summary(navchip.gy.r, inference = T)

# 3. Compare Classical with Robust WV
compare.wvar(navchip.gy, navchip.gy.r)

# 4. Load and cast data for MTiG
data("imu6")
# Cast as an IMU
sensors = imu(imu6, gyros = 1:3, accels = 4:6, freq = 100)

# 5. Obtain WV and plot WV
wv.sensors = wvar(sensors)
plot(wv.sensors)

# 6. Automatic Model Selection with Overall Model Guess
models.sensors = auto.imu(sensors,
    model = 4 * AR1() + WN() + RW())

# 7. Visually see the best models
plot(models.sensors)
## List of Tables

3.1 Sample size, median computational time in seconds and convergence rates of the RGMWM and INDI estimators for the models ARMA(1,2), ARMA(3,1) and SSM in a contaminated setting. ................................................................. 50

3.2 Random Walk plus ARMA(2,1) model estimates for the PSR data. Estimated parameters with GMWM and RGMWM estimators with $\gamma^2$ being the random walk parameter, $\rho_i$ the $i^{th}$ autoregressive parameter, $\varrho$ the moving average parameter and $\sigma^2$ the innovation variance of the ARMA(2,1) model. Confidence intervals (CI) based on the approach used in Guerrier et al. (2013b). ................................................................. 56

3.3 AR(1) estimates for the mean monthly precipitation data from 1907 to 1972 taken from Hipel and McLeod (1994). Estimated parameters with ML, GMWM and RGMWM estimators with $\hat{\phi}$ being the estimated autoregressive parameter and $\hat{\sigma}^2$ the innovation variance. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b). ................................................................. 59

3.4 Exponential model estimates for the cloud image data in the third region using the GMWM and RGMWM estimators. $\hat{\phi}$ is the estimated range parameter and $\hat{\sigma}^2$ the estimated sill. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b). ................................................................. 60

3.5 Exponential model estimates for the cloud image data in the fourth region using the GMWM and RGMWM estimators. $\hat{\phi}$ is the estimated range parameter and $\hat{\sigma}^2$ the estimated sill. Confidence Intervals (CI) based on the approach used in Guerrier et al. (2013b). ................................................................. 61

4.1 MLE and GMWM estimated parameters for the SARIMA model on the prodn data. $\phi_i$ represents the $i^{th}$ autoregressive parameter, $\Theta_i$ represents the $i^{th}$ seasonal moving-average and $\sigma^2$ represents the innovation variance. 71

4.2 Estimated parameters for an AR1() model on the hydro dataset. ................. 80

4.3 Benchmarks of the WV computation in seconds. ......................................... 81

5.1 Standard vs robust parameter estimates (with corresponding standard deviations) for the NavChip Y-axis gyroscope. The robust estimator is based on an efficiency of 0.6 compared to the standard estimator. ......................... 93

C.1 Sample size, median computational time in seconds and convergence rates of the RGMWM and INDI estimators for the models ARMA(1,2), ARMA(3,1) and SSM in an uncontaminated setting. ......................................................... 137
C.2 State-space model estimates for the gyroscope data in static conditions. Estimated parameters with GMWM and RGMWM estimators with $\rho_i$ being the $i^{th}$ autoregressive parameter, $\nu_i^2$ the innovation variance of the $i^{th}$ autoregressive model. Confidence intervals (CI) based on the approach used in Guerrier et al. (2013b). 

D.1 Standard estimation of a 3*GM() model using “gmwm.imu()” in seconds . . 141
D.2 Robust estimation of a 3*GM() model using “gmwm.imu()” in seconds . . 141
List of Figures

1.1 Summary of conditions and asymptotic properties for the standard estimator $\hat{\nu}_j^2$, the robust estimator $T_N$ and the proposed estimator $\tilde{\nu}_j^2$. Normal arrows indicate links with conditions; double-headed arrows denote equivalence or implication; small full-headed arrow denotes the consequent property under the condition that $j \in J < \infty$; large full-headed arrow denotes the consequent property with $J < \lfloor \log_2(T) \rfloor$; empty-headed arrow denotes the consequent property under the condition that $J < \lfloor \log_2(\sqrt{T}) \rfloor$. The dashed lines denote approximate equivalence. ................................................ 17

1.2 RMSE* of the three WV estimators for the time series models ........................................ 21

1.3 RMSE* of the three WV estimators for the spatial models ............................................ 21

1.4 Top: Surface albedo data. Bottom: Histogram of the first backward difference of the surface albedo data (left); Histogram of the second backward difference of the surface albedo data (right). ........................................ 23

1.5 Log-log plot of the estimated standard WV and robust WV on the surface albedo data. ......................................................................................................................... 23

1.6 Heat-images of the raw cloud data for the four regions (top row), of the standard WV $\hat{\nu}$ (top-middle row), of the robust WV $\tilde{\nu}$ (bottom-middle row) and of their absolute difference (bottom row). The lighter shades indicate low values and darker shades indicate high values. The colour-coding is relative to each region (i.e. the WV has the same coding within the same region but not necessarily between regions). .................................................. 25

1.7 Left: Log-log plot of the standard and robust WV over a 1,000m² area with coordinates of central point: [4050, 4050]. Right: Log-log plot of the standard and robust WV over a 1,000m² area with coordinates of central point: [450, 450]. .......................................................... 27

2.1 Logarithm of the MSE for the GMM (red line) and GMWM (green line) on Latent Model 1 (top part of figure) and Latent Model 2 (bottom part of figure). The grey shaded areas represent the confidence intervals of the MSE for each estimator. .......................................................... 39

2.2 Summary of the results on identifiability and consistency for the MLE, GMM and GMWM estimators concerning the classes of latent models considered in this chapter. ................................................................................................. 40

3.1 Top row: RMSE* of the estimators in an uncontaminated setting. Bottom row: RMSE* of the estimators in a contaminated setting. KUNSCH represents the MAR estimator for the AR(1) and AR(2) models while it represents the INDI estimator for the other models. .......................................................... 49
3.2 Top row: RMSE* of the estimators in an uncontaminated setting. Bottom row: RMSE* of the estimators in a contaminated setting. .......................... 52
3.3 Top part: Inertial sensor time series. Bottom part: zoom-in on grey part of the time series with black points indicating extreme outliers identified through the weights of RGMWM. ................................. 54
3.4 Top figure: Saving rates time series with different types of points indicating outliers identified through the weights of the RGMWM. Bottom figure: log-log scale WV plots for saving rates series; Top left: classic estimated WV superposed with model-implied WV based on the parameters estimated through the GMWM. Top right: classic and robust estimated WV with respective confidence intervals superposed. Bottom left: classic and robust model-implied WV based on the GMWM and RGMWM estimates respectively. Bottom right: robust estimated WV superposed with model-implied WV based on the parameters estimated through the RGMWM. .......................... 57
3.5 (Top panel) Monthly precipitation series from 1907 to 1972 taken from Hipel and McLeod (1994). (Bottom panels) Estimated autocorrelation function (left) and estimated partial autocorrelation function (right) of the precipitation series. .......................................................... 59
4.1 Top-left: Estimated WV for white noise (orange line) and estimated WV for quantization noise (blue line) with confidence intervals (shaded areas). Top-right: Estimated WV for random walk (orange line) and estimated WV for drift (blue line) with confidence intervals (shaded areas). Bottom-left: Estimated WV for an AR(1) process with confidence intervals (shaded areas). Bottom-right: Estimated WV for an SARIMA(1,0,1) × (2,1,1)12 process with confidence intervals (shaded areas). ........................................... 69
4.2 Comparison between standard and robust WV for the prodn data. .......... 70
4.3 Estimated WV from the prodn data (blue line) with confidence intervals (shaded area) and WV implied by estimated SARIMA model (orange line) 72
4.4 Left: Breakdown of underlying processes and resulting latent process (bottom) from model specified in (4.4.2). Right: WV implied by the fitted model compared to the estimated WV. Additional lines show the WV implied by the parameters of each of the underlying models. ....................... 75
4.5 Left: Plot of Annual Nile river flow from 1871 - 1970. Right: WV implied by the selected local level model (orange line) and estimated WV (blue line) with confidence intervals (shaded area). ........................ 77
4.6 Left: Simulated time series with contaminated observations. Right: Comparison of resulting standard and robust WV. ................................. 79
4.7 Comparison of the standard (blue line) and robust (orange line) estimated WV on the hydro data with shaded areas representing the respective confidence intervals. ...................................................... 80
4.8 Median estimation times of the methods on different sample sizes. ........ 82
5.1 Plots produced by using the function plot() on the WV issued from the function wvar() on the NavChip IMU. (a) Plot with parameter split=TRUE; (b) Plot with parameter split=FALSE. ........................................... 91
5.2 Plot produced by the function compare.wvar() on objects containing the standard and robust WV respectively. ................................. 92
5.3 WV implied by the estimated $3*AR1()+WN()+QN()+RW()$ model compared to the empirical WV of the NavChip Y-axis gyroscope. 94

5.4 Results of running `auto.imu()` on the “mtig” data. Dashed line is the implied WV while the solid line is the empirical WV. 95

A.1 Top: theoretical risk function for a series of Dirac points. Bottom: theoretical risk function for a Gaussian contamination with different scale parameters105

C.1 Log-log scale Wavelet Variance plots for the inertial sensor series. Top left: classic estimated WV superposed with model-implied WV based on the parameters estimated through the GMWM. Top right: classic and robust estimated WV with respective confidence intervals superposed. Bottom left: classic and robust model-implied WV based on GMWM and RGMWM estimates respectively. Bottom right: robust estimated WV superposed with model-implied WV based on the parameters estimated through the RGMWM. 137
Bibliography


— (2012). Generalized Method of Wavelet Moments (GMWM) Modelling Software. 1.1. EPFL. CH-1015 Lausanne, Switzerland.


