Modeling Time Series via Copula and Extreme Value Theory

by

Zifeng Zhao

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2018

Date of final oral examination: 12/08/2017

The dissertation is approved by the following members of the Final Oral Committee:

Zhengjun Zhang, Professor, Statistics

Peng Shi, Associate Professor, Risk and Insurance

Kenneth D. West, Professor, Economics

Chunming Zhang, Professor, Statistics

Edward W. Frees, Professor, Risk and Insurance

Acknowledgments

The life of a Ph.D. student is not easy and consists of ups and downs. Looking back upon the four and half years I spent at the University of Wisconsin-Madison, I realize how lucky I am to have an extremely supportive circle of friends and colleagues, without whom this journey would have been more difficult.

First, I want to express my sincere gratitude to my advisor Professor Zhengjun Zhang, for the continuous guidance, inspiring ideas and enormous support he has provided me throughout my years as a Ph.D. student. Professor Zhang introduced Extreme Value Theory and Copula to me, which serve as the foundation of my research and will keep benefiting me throughout my academic life. Professor Zhang encouraged me to apply to various conferences and guided me to win several travel awards and paper awards. Moreover, Professor Zhang introduced me to other brilliant researchers, which not only broadened my views but also provided valuable collaboration opportunities. I deeply appreciate his great help during my life in Madison and the process of my job hunting.

I would like to thank Professor Peng Shi, with whom I have worked on various research projects. Professor Shi is an outstanding researcher in both statistics and insurance, and serves as a perfect example for applying statistics to solve important real world problems.

I would like to thank Professor Kenneth D. West, with whom I have collaborated in Econometrics. Although Professor West is a world-renowned economist and a prominent researcher, he remains a very hard-working and humble individual who serves as a perfect role model for a junior researcher like me.

I thank Professor Chunming Zhang and Professor Edwards W. Frees for being on my committee and for giving me support and invaluable ideas to complete this dissertation.

I am indebted to all my friends during my time as a graduate student, who have made my life in Madison pleasant and memorable. I also appreciate all the help from the staff in the department.

Finally, I want to thank my family. I would like to thank my parents, Weizheng Zhao and Qingmei Xu, for their love and support throughout my life. I also want to give a special thank you to my wife, Margaryta Bondarenko, for her enormous support, genuine advice and endless love.

Contents

List of Tables

List of Figures

Abstract

Throughout my Ph.D. life, I mainly work on research topics about time series. This thesis consists of three representative works I have done for statistical modeling of time series. Modeling time series is a fundamental task in statistics, with extensive applications in finance, economics, climate etc. One significant characteristics of time series is that the data have complicated dependence structures.

For univariate time series, there is temporal dependence, where the past influences the future (autocorrelation). The modeling of univariate time series is a relatively well-studied research area, especially in financial applications. However, for the modeling of extreme events, such as maximum daily loss of a high-frequency trading, the suitable models are still rare. An accurate time series model for extreme events is essential for understanding the behavior of extreme risk and thus managing the extreme risk. Parts of the thesis (**Chapter 2**) are dedicated to the development of a dynamic extreme-value-theory (EVT) based time series model which can capture the time-varying behavior of maxima in financial time series.

For multivariate time series, there are both temporal dependence of each component univariate time series and cross-sectional dependence across all the component univariate time series. To accurately capture the behavior of multivariate time series, it is essential for the time series model to be highly flexible such that it is capable of modeling various dependence structures, such as nonlinear dependence, tail dependence and asymmetric dependence. Essentially, we need to construct flexible dependence structures, i.e. multivariate distribution functions. In the literature, Copula is one of the most widely used methods for generating sophisticated multivariate distribution functions. Parts of the thesis (**Chapter 3** and **4**) focus on the construction of novel copulas and copula-based multivariate time series models.

A complicated statistical model may be computationally expensive to estimate via the conventional maximum likelihood estimator (MLE), which can also be infeasible if the model is semiparametric. Parts of the thesis, specifically the parameter estimation sections of **Chapter 2** to **4**, propose efficient estimation procedures for the complex models based on stepwise estimation and composite likelihood estimation, and study the theoretical properties of irregular MLE and semiparametric sequential estimators.

Chapter 1

Introduction

1.1 Modeling Time Series of Maxima

The study of extreme events in financial markets is always one of the main foci in risk management. Maximum observations, as the representation of extreme behavior, are of particular interest. For example, mutual fund managers are keen to assess the potential maximum daily loss across all stocks in their managed portfolio; the level of potential intra-day maximum loss is important to high-frequency traders. By Fisher–Tippett–Gnedenko theorem, the generalized extreme value distribution (GEV) can be used to characterize the behavior of maxima, making extreme value theory (EVT) a widely researched and practiced approach for risk management in financial industry (e.g. Embrechts et al., 1999, McNeil and Frey, 2000, Laurini and Tawn, 2009). Besides the Maxima-GEV methodology, the other main methodology of EVT is the peak-over-threshold (POT), which is based on generalized Pareto distribution (GPD). By Pickands–Balkema–de Haan theorem, GPD can be used to approximate the conditional behavior of random variable after it exceeds certain high thresholds (e.g. Balkema and de Haan, 1974, Picklands, 1975, Davison and Smith, 1990). Under EVT framework, Maxima-GEV and POT-GPD are closely related and can often reveal the same information, especially when used to model tail index ξ. See Chapters 4 and 7 in Coles (2001) for more details about the connection between Maxima-GEV and POT-GPD.

1.1.1 Autoregressive Conditional Fréchet Model

As mentioned in Diebold et al. (1998), most applications of EVT focus on modeling extreme events in time series with a static approach under equilibrium distribution. However, the behavior of the underlying time series may change through time. For example, financial time series tends to exhibit structural changes and time-varying dynamics such as volatility clustering. To accommodate the dynamics of extreme events and study the conditional behavior of tail risk in financial markets, there have been several recent studies of dynamic POT-GPD models. For example, Smith and Goodman (2000) and Chavez-Demoulin et al. (2014) use Bayesian method to update the time-varying GPD parameters. Kelly (2014) and Kelly and Jiang (2014) build a dynamic tail model with POT-GPD for panel data. Massacci (2016) and Zhang and Schwaab (2017) employ a generalized autoregressive score[1] type of observation-driven dynamics for the GPD parameters. These studies show strong evidence of the time-varying behavior of extreme events in financial markets, especially for the tail index ξ.

One advantage of Maxima-GEV approach over POT-GPD approach is that it offers a direct modeling of maxima in time series, which is of particular importance and is the primary focus of this paper. Unlike the dynamic GPD models, there is little research on dynamic GEV models. Bali and Weinbaum (2007) design a time-varying GEV to estimate the realized volatility in an empirical study of market risk, however, theoretical results are not provided.

Chapter 2 introduces a novel dynamic generalized extreme value (GEV) framework for modeling the time-varying behavior of maxima in financial time series. Specifically, an autoregressive conditional Fréchet (AcF) model is proposed in which the maxima are modeled by a Fréchet distribution with time-varying scale parameter (volatility) and shape parameter (tail index) conditioned on past information. The AcF provides a direct and accurate modeling of the time-varying behavior of maxima and offers a new angle to study the tail risk dynamics in financial markets. Probabilistic properties of AcF are studied, and an irregular maximum likelihood estimator is used for model estimation, with its statistical properties investigated. Simulations show the flexibility of AcF and confirm the reliability of its estimators. Two real data examples on cross-sectional stock returns and high-frequency foreign exchange returns are used to demonstrate the AcF modeling approach,

[1] See Harvey and Chakravarty (2008) and Creal et al. (2013) for more details about generalized autoregressive score model.

where significant improvement over the static GEV has been observed for market tail risk monitoring and conditional VaR estimation. Empirical result of AcF is consistent with the findings of the dynamic peak-over-threshold (POT) literature that the tail index of financial markets varies through time.

1.2 Modeling Multivariate Time Series

Modeling multivariate time series is the fundamental task for many statistical applications, see Patton (2012) for applications in economics, Erhardt et al. (2015) for applications in climate monitoring and Smith (2015) for applications in energy markets. Roughly speaking, there are two types of dependence for multivariate time series. The first type is the temporal dependence for each of its component univariate time series. The second type is the cross-sectional dependence across all of its component univariate time series. To capture the behavior of multivariate time series, it is essential for the statistical model to have an accurate modeling of both the cross-sectional and the temporal dependence. This requires the time series model to be highly flexible such that it is capable of modeling various dependence structures, including nonlinear dependence, tail dependence and asymmetric dependence.

In the literature, the copula model is one of the most widely used methods for generating flexible dependence structures. Copulas are d-dimensional distribution functions on $[0,1]^d$ with uniform margins. By Sklar (1959)'s theorem, any multivariate distribution $\mathbf{F}$ can be separated into its marginals $(F_1, \ldots, F_d)$ and its copula C, where the copula captures all the scale-free dependence in the multivariate distribution. In particular, suppose there is a random vector $\mathbf{Y} \in \mathbb{R}^d$ such that $\mathbf{Y}$ follows $\mathbf{F}$, then we have $\mathbf{F}(\mathbf{y}) = C(F_1(y_1), \ldots, F_d(y_d))$, where $\mathbf{y} = (y_1, \ldots, y_d)'$ is the realization of $\mathbf{Y}$. If all the marginals of $\mathbf{F}$ are absolutely continuous, the copula C is unique.

Various copulas have been proposed in the literature, see Joe (2014) and Nelsen (1999) for a summary. Copula-based models for multivariate distributions are widely used in a variety of applications, see Frees and Valdez (1998) in actuarial science and insurance, Cherubini et al. (2004) in finance, and Genest and Favre (2007) in hydrology.

1.2.1 Max-copula Model

Due to its tractability, interpretability, and flexibility in modeling non-extremal joint behavior, the Gaussian copula is arguably the most widely used copula. While Gaussian copulas perform well in many areas of applications, the financial market may turn out to be an exception. One of the most significant characteristics of financial data is its tail dependence, i.e., during a crisis, asset prices tend to move together. The failure of Gaussian copulas to capture this tail dependence in the pricing of CDOs and related securities is considered one of the prominent causes of the recent financial crisis, e.g., see Coval et al. (2009) and Salmon (2012) for more details.

The Gaussian copula's inability to model joint tail events inspires more research in the construction of copulas that can offer more sophisticated dependence structures. One direction is to exploit and extend the linear structure in Gaussian factor models by changing the distribution of latent factors from Gaussian to other distributions like *skewed-t*, e.g., Hull and White (2004), Murray et al. (2013), and Oh and Patton (2015). This framework is capable of offering more sophisticated dependence like tail dependence and tail asymmetry. Also, it is particularly attractive in high-dimensional applications thanks to the factor structure. Another promising direction is to use vine copulas, which builds high-dimensional copulas based on a sequentially iterative pairwise construction of bivariate copulas, see Aas et al. (2009), Min and Czado (2010), and Almeida et al. (2012) for more details. Vine copulas can offer flexible dependence relationships and can be represented in graphs, which helps the modeler to visually understand the dependence structure. Combining the "latent factor" idea with the pairwise construction idea of C-vine copulas, Krupskii and Joe (2013) proposed a factor copula model, where instead of imposing a Gaussian linear structure, bivariate copulas are used to specify the dependence between latent factors and observed variables. Theorem A1 in §1 of the supplementary material establishes a weak equivalence between the factor copula model based on the C-vine copula and the linear factor copula model under the additive model framework.

A much different yet important direction is to construct new sophisticated copulas based on existing ones. The idea is that the new copula inherits various merits from its parents and thus offers more versatile dependence structures. In the literature, mixture of distributions is a long existing technique for generating new distributions based on existing ones. The mixture copula has

a closed form *c.d.f.* and is interpretable.

Inspired by Cui and Zhang (2016) and the mixture technique, **Chapter 3** presents a novel nonlinear framework for the construction of flexible new copulas from existing copulas based on a straightforward "pairwise max" rule. The newly constructed max-copula has a closed form and has strong interpretability. Compared to the classical "linear symmetric" mixture copula, the max-copula can be viewed as a "nonlinear asymmetric" framework. It is capable of modeling asymmetric dependence and joint tail behavior while also offering good performance in non-extremal behavior modeling. Max-copulas that are based on single-factor and block-factor models are developed to offer parsimonious modeling for structured dependence, especially in high-dimensional applications. Combined with semiparametric time series models, the max-copula can be used to develop flexible and accurate models for multivariate time series. A new semiparametric composite maximum likelihood method is proposed for parameter estimation, where the consistency and asymptotic normality of estimators are established. The flexibility of the max-copula and the accuracy of the proposed estimation procedure are illustrated through extensive numerical experiments. Real data applications in Value at Risk estimation and portfolio optimization for financial risk management demonstrate the max-copula's promising ability to accurately capture joint movements of high-dimensional multivariate stock returns under both normal and crisis regime of the financial market.

1.2.2 Multivariate D-vine Copula Model

Most applications of copulas under time series setting focus on the modeling of the cross-sectional dependence of multivariate time series, see the semiparametric copula-based multivariate dynamic (SCOMDY) models in Patton (2006) and Chen and Fan (2006a). Under the SCOMDY framework, conventional univariate time series models, e.g. ARMA and GARCH (Engle, 1982, Bollerslev, 1986), are used to capture the temporal dependence, in particular conditional mean and variance, of each component univariate time series. A parametric copula and nonparametric marginal distributions are used to specify the cross-sectional dependence across the innovations of all the component univariate time series.

There are also applications where copulas are used for modeling the temporal dependence of a univariate time series. Darsow et al. (1992) provide a necessary and sufficient condition for a copula-based time series to be a Markov process. Joe (1997), Chen and Fan (2006b), Domma et al.

(2009), Ibragimov (2009) and Beare (2010) consider copula-based stationary Markov chains, where copulas and flexible marginal distributions are used to specify the transitional probability of the Markov chains. Most copula-based time series models focus on first-order Markov chains, since there are many flexible and tractable bivariate copulas. For higher-order Markov chains, the result is rare and mostly is based on elliptical copulas such as Gaussian copula and t copula due to their tractability.

As mentioned in Smith (2015), despite that copulas have been employed in either cross-sectional or temporal dependence modeling of time series, there is few multivariate time series model that uses copulas to simultaneously characterize both the cross-sectional and temporal dependence. One exception is Smith (2015) itself, where the author stacks the multivariate time series into a univariate time series and designs copula-based dependence structures for the resulted univariate time series. The proposed model in Smith (2015) is flexible and can capture nonlinear and asymmetric tail dependence, and it outperforms the conventional Vector AR model in various applications. One potential drawback is that the model is complicated and may not be easily implemented by practitioners.

Chapter 4 proposes a novel semiparametric D-vine copula based multivariate time series model (mDvine), which enables the simultaneous copula-based modeling for both the temporal and cross-sectional dependence of a multivariate time series. Due to the use of D-vine copula, the proposed mDvine is tractable and can generate flexible dependence structures. For the construction of mDvine, a univariate D-vine time series model (uDvine) is first designed for the modeling of each component univariate time series. The uDvine generalizes the existing first-order copula-based Markov chain to an arbitrary-order Markov chain. The uDvine provides flexible models for marginal behavior of univariate time series and is capable of generating sophisticated temporal dependence structures such as nonlinear dependence, asymmetric dependence and tail dependence. Using the same idea of SCOMDY framework in Chen and Fan (2006a), the mDvine is constructed based on uDvine, where a parametric copula is employed to specify the cross-sectional dependence among all the uDvine-based component univariate time series. Probabilistic and statistical properties of uDvine and mDvine are studied in detail. A sequential model selection procedure and a two-stage MLE are proposed for the inference and estimation of the mDvine. The proposed procedures are robust and computationally fast. The consistency and asymptotic normality of the MLE is

established and affirmed by extensive numerical experiments. Real data applications on the Australian electricity market and the Ireland wind speed data demonstrate mDvine's promising ability for modeling multivariate time series, where significant improvement over conventional time series models is observed.

Chapter 2

Autoregressive Conditional Fréchet Model for Time Series of Maxima

2.1 Motivation

The study of extreme events in financial markets is always one of the main foci in risk management. Maximum observations, as the representation of extreme behavior, are of particular interest. For example, mutual fund managers are keen to assess the potential maximum daily loss across all stocks in their managed portfolio; the level of potential intra-day maximum loss is important to high-frequency traders. By Fisher–Tippett–Gnedenko theorem, the generalized extreme value distribution (GEV) can be used to characterize the behavior of maxima, making extreme value theory (EVT) a widely researched and practiced approach for risk management in financial industry (e.g. Embrechts et al., 1999, McNeil and Frey, 2000, Laurini and Tawn, 2009). Besides the Maxima-GEV methodology, the other main methodology of EVT is the peak-over-threshold (POT), which is based on generalized Pareto distribution (GPD). By Pickands–Balkema–de Haan theorem, GPD can be used to approximate the conditional behavior of random variable after it exceeds certain high thresholds (e.g. Balkema and de Haan, 1974, Picklands, 1975, Davison and Smith, 1990). Under EVT framework, Maxima-GEV and POT-GPD are closely related and can often reveal the same information, especially when used to model tail index ξ. See Chapters 4 and 7 in Coles (2001) for more details about the connection between Maxima-GEV and POT-GPD.

As mentioned in Diebold et al. (1998), most applications of EVT focus on modeling extreme events in time series with a static approach under equilibrium distribution. However, the behavior of the underlying time series may change through time. For example, financial time series tends to exhibit structural changes and time-varying dynamics such as volatility clustering. To accommodate the dynamics of extreme events and study the conditional behavior of tail risk in financial markets, there have been several recent studies of dynamic POT-GPD models. For example, Smith and Goodman (2000) and Chavez-Demoulin et al. (2014) use Bayesian method to update the time-varying GPD parameters. Kelly (2014) and Kelly and Jiang (2014) build a dynamic tail model with POT-GPD for panel data. Massacci (2016) and Zhang and Schwaab (2017) employ a generalized autoregressive score[1] type of observation-driven dynamics for the GPD parameters. These studies show strong evidence of the time-varying behavior of extreme events in financial markets, especially for the tail index ξ.

One advantage of Maxima-GEV approach over POT-GPD approach is that it offers a direct modeling of maxima in time series, which is of particular importance and is the primary focus of this paper. Unlike the dynamic GPD models, there is little research on dynamic GEV models. Bali and Weinbaum (2007) design a time-varying GEV to estimate the realized volatility in an empirical study of market risk, however, theoretical results are not provided.

In this paper we are mainly interested in modeling time series of maxima $\{Q_t\}$, where $\{Q_t\}$ is a univariate time series of maxima based on a set of underlying financial time series $\{X_{it}\}_{i=1}^{p}$. There are mainly two types of $\{Q_t\}$. The first type is the time series of cross-sectional maxima, where $\{X_{it}\}_{i=1}^{p}$ are a set of panel time series, and we are interested in modeling the cross-sectional maxima $Q_t = \max_{1\leq i\leq p} X_{it}$. Such problems arise in many applications, including modeling the maximum daily loss across a group of stocks in a portfolio. The second type is the time series of intra-period maxima, where $\{X_{it}\}_{i=1}^{p}$ denote the p intra-period observations for a univariate time series within period t, and we are interested in modeling the intra-period maxima $Q_t = \max_{1\leq i\leq p} X_{it}$. For example, one may be interested in the intra-day maxima of high-frequency trading losses that occur on the same day.

It is worth noting that there is an important difference between "maxima" and "extreme event".

[1]See Harvey and Chakravarty (2008) and Creal et al. (2013) for more details about generalized autoregressive score model.

"Maxima" has a clear definition of being the maximum of a set of observations, while "extreme event" is a more vague term, typically defined as rare observations over a high threshold. A "maxima" may not necessarily be an "extreme event" though most likely it is. As a time series of maxima, $\{Q_t\}$ is observed at every t. On the other hand, an extreme event may not be observed at each t, or there may be several extreme events within a time period. There is ample research on extremal process that offers stochastic characterization and modeling of "extreme event" in stationary process (see Resnick, 1987, Basrak et al., 2002, Basrak and Segers, 2009, Drees et al., 2015, for more details). However, in this paper the focus is the direct modeling of the maxima $\{Q_t\}$ process and its time-varying behavior.

An important byproduct from dynamic modeling of maxima $\{Q_t\}$ is the tail index ξ_t, which is arguably the most important indicator for financial market tail risk. Thus, a better modeling of maxima can help obtain a more accurate assessment of the current market risk level and offer more insight into the potential market extreme movement.

With the aim of modeling the time-varying behavior of maxima in financial time series, in this paper we introduce a novel dynamic conditional GEV framework, in which parameters (μ, σ, ξ) of a conditional GEV are allowed to vary through time with a GARCH-like[2] autoregressive mechanism. Due to the heavy-tailed nature of financial data, Fréchet distribution is widely used for modeling maxima originated from financial time series. Thus, we propose an autoregressive conditional Fréchet (AcF) model that allows for an observation-driven time evolution of the scale parameter σ and the tail index ξ of a Fréchet (Type-II GEV) distribution. Since the scale parameter and the tail index play the key role in characterizing the tail behavior of Fréchet distribution, AcF provides a more flexible and applicable model for the time-varying behavior of maxima in financial time series.

The main contributions of this paper are in the following aspects. From a statistical point of view, this paper provides the first complete treatment of a dynamic GEV model. The AcF is a direct approach to modeling the time-varying behavior of maxima in financial time series. Probabilistic properties of the model and statistical properties of its estimator are investigated and developed. They make the paper theoretically sound and help lay the foundation for further development of dynamic models under EVT framework. From an econometric point of view, the

[2]See Engle (1982) and Bollerslev (1986) for more details about GARCH model.

newly designed AcF offers a new angle to study the time-varying behavior of tail risk in financial markets and serves as a promising alternative to the dynamic POT-GPD methodology in the literature. Real data applications show that the tail index of financial market is indeed time-varying. Compared to the static GEV, AcF captures the dynamics of maxima more adequately and offers more promising performance in detecting potential market extreme movement and providing more accurate conditional VaR prediction for maxima.

The rest of the paper is organized as follows. In Section 2.2 we present a detailed description of AcF and investigate its probabilistic properties. A maximum likelihood estimator (MLE) is used for estimation and its statistical properties as an irregular MLE are derived in Section 2.3. Simulation studies are presented in Section 2.4 to demonstrate the flexibility of AcF and to evaluate the performance of MLE. Section 2.5 presents two real data applications, one on market tail risk monitoring based on cross-sectisonal maximum loss of stock markets and one on conditional VaR estimation of intra-day maximum loss from high-frequency foreign exchange trading. We conclude the paper in Section 2.6. The Appendix contains the proofs of the theorems and other technical materials.

2.2 Autoregressive Conditional Fréchet Model

2.2.1 Background

As a time series of maxima, $\{Q_t\}$ cannot be directly modeled by conventional time series models like ARMA or GARCH. By Fisher–Tippett–Gnedenko theorem, we know that under certain condition, marginally $\{Q_t\}$ can be accurately approximated by a GEV distribution with three parameters (μ, σ, ξ), the location, the scale, and the shape parameter, respectively. The common practice in the literature of modeling maxima $\{Q_t\}$ is to treat it as *i.i.d.* data and model it by a GEV distribution. The obvious limitation is that the time dependency between $\{Q_t\}$ has been completely ignored in this approach, which can potentially cause a huge loss in model efficiency if $\{Q_t\}$ has a strong dependency across time. To overcome this drawback, we propose a dynamic GEV framework, under which a conditional evolution scheme is designed for the parameters (μ_t, σ_t, ξ_t) of GEV, so that time dependency of $\{Q_t\}$ can be captured. Due to the heavy-tailedness of financial data, $\{Q_t\}$ marginally can be well modeled by Fréchet distribution (i.e. Type II GEV distribution),

which corresponds to the GEV of which $\xi > 0$. For the rest of the paper, we focus on the case that conditionally Q_t can be modeled by a Fréchet distribution with parameters $(\mu_t, \sigma_t, \alpha_t)$ where $\alpha_t = 1/\xi_t$ as often used for the parametrization of Fréchet distribution:

$$Q_t = \mu_t + \sigma_t Y_t^{1/\alpha_t}, \tag{2.1}$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables and $(\mu_t, \sigma_t, \xi_t) \in \mathcal{F}_{t-1} = \sigma(Q_{t-1}, Q_{t-2}, \ldots)$. Here μ_t and σ_t are location-scale parameters and α_t is the shape parameter, also called tail index of Fréchet distribution.

The scale parameter σ_t should not be taken exactly as volatility since the conditional variance of Q_t depends on both σ_t and α_t. However, as shown later in Proposition 2.2.1, σ_t can be closely related to the volatility process of the underlying time series $\{X_{it}\}_{i=1}^p$ and thus requires a dynamic treatment due to the well-known volatility clustering in financial time series. Moreover, Harvey (2013) observes that if volatility clustering is not accounted for, movements of the tail parameter α_t can be potentially confounded with movements of the scale parameter σ_t.

The tail index α_t is the essential parameter since it governs the underlying tail behavior of $\{Q_t\}$ process and plays the most important role in quantifying the potential tail risk. To demonstrate the necessity of dynamic modeling of α_t, we perform an ad-hoc moving-window GEV analysis on the cross-sectional maxima of negative daily log-returns (i.e. daily losses) of the component stocks in S&P100 index which includes 100 leading U.S. stocks. The observation period is from January 1, 2000 to December 31, 2014 with 3773 trading days. For each trading day t, we record the maximum daily loss across the 100 stocks and denote it by Q_t. Hence the time series $\{Q_t\}$ has 3773 observations. For each t such that $500 \leq t \leq 3273$, a GEV model is fitted using $\{Q_k\}_{k=t-499}^{t+500}$, the observations within a 1000-day (approximately 4 years) local window centered at t. We plot the estimated tail index $\hat{\alpha}_t$ in Figure 2.1 along with the estimated tail index $\hat{\alpha}$ obtained by directly fitting the static GEV model with the entire series, treating them as *i.i.d.* observations.

It can be clearly seen that compared to the static estimation (dashed line), the tail index estimated with smaller moving window (solid curve) changes quite drastically throughout the years, indicating an insufficiency of the static GEV model. A similar finding of varying tail index has been reported in Kelly and Jiang (2014) and Massacci (2016). An interesting phenomenon is that the

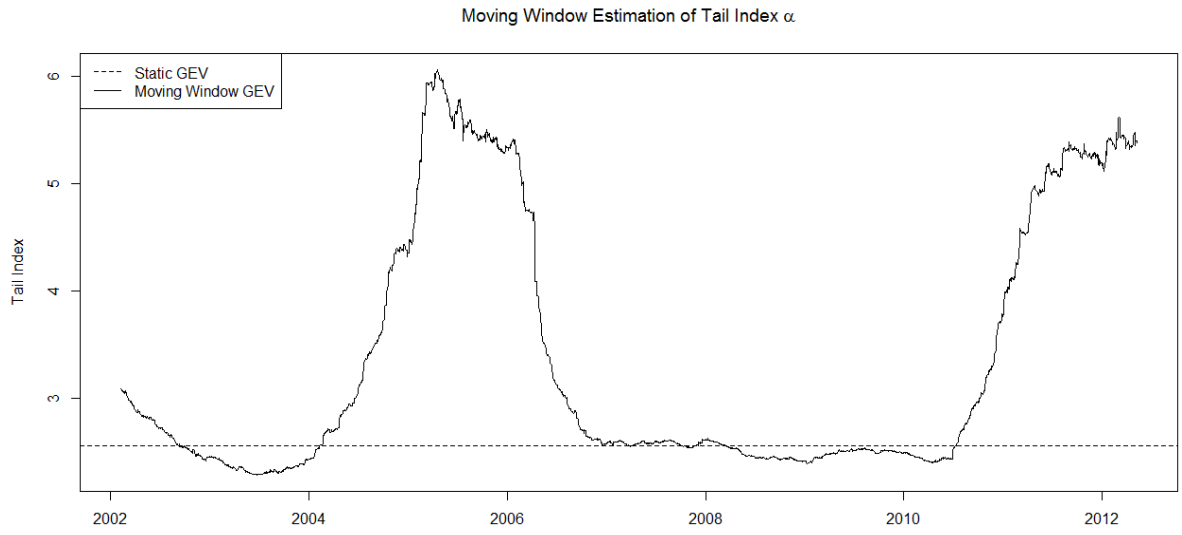

Figure 2.1: *Tail index $\hat{\alpha}_t$ estimated by moving window of size 1000 (solid curve) v.s. tail index $\hat{\alpha}$ estimated by the static GEV model based on total observations (dashed line).*

static GEV seems to give a significantly lower estimation of the tail index than the moving-window GEV. An under-estimation of tail index over-estimates the tail risk, which in turn may result in higher reserve requirements and other expenses for financial institutions.

2.2.2 Model specification

For parsimony, we set μ_t to be constant, which is the common practice in the extreme value analysis, and concentrate on the dynamics of σ_t and α_t, which are the key parameters of modeling tail behavior. We impose an autoregressive structure on the time-varying parameters (σ_t, α_t) similar to the approach of GARCH model in Bollerslev (1986), autoregressive conditional density model (ACD) in Hansen (1994), and autoregressive conditional duration model in Engle and Russell (1998).

Specifically, the autoregressive conditional Fréchet (AcF) model assumes the form

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \eta_1(Q_{t-1}), \tag{2.2}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \eta_2(Q_{t-1}), \tag{2.3}$$

where $\beta_1, \gamma_1 \geq 0$ and the two terms $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are the observation-driven factors for $\{\log \sigma_t\}$ and $\{\log \alpha_t\}$. The log transform is used to ensure the positivity of the parameters.

We further assume that $\eta_1(\cdot)$ is a continuous increasing function and $\eta_2(\cdot)$ is a continuous decreasing function of Q_{t-1}. One salient feature of the maxima series $\{Q_t\}$ in many applications, especially in financial time series, is the clustering of extreme events in time. It has been observed that large-valued maxima tend to happen around the same period in many applications. One possible explanation is that an extreme event observed at time $t-1$ (large Q_{t-1}) causes the distribution of Q_t to have larger scale (large σ_t) and heavier tail (small α_t), resulting in a larger tail risk of Q_t. An increasing $\eta_1(\cdot)$ and a decreasing $\eta_2(\cdot)$ ensure that larger Q_{t-1} is followed by larger σ_t and smaller α_t. Together with the autoregressive scheme of $\{\log\sigma_t\}$ and $\{\log\alpha_t\}$ (i.e. $\gamma_1, \beta_1 \geq 0$), this evolution dynamics offers a joint modeling of both volatility clustering for $\{\sigma_t\}$ process and heavy-tail clustering for $\{\alpha_t\}$ process.

There are many choices of the continuous monotone functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$. In this paper we use the simple exponential function $a_0\exp(-a_1 x)$. It is a simplified version of the widely used logistic function $\frac{L}{1+a_0\exp(-a_1 x)}$. Due to its monotonicity, differentiability, and boundedness, the logistic function is employed in many studies of observation-driven time series models (e.g. Hansen, 1994, Lundbergh et al., 2003, Boutahar et al., 2008, Hall et al., 2016). The simplification here is due to $Q_t > \mu$, hence there is no need to have the 1 in the denominator of the logistic function for boundedness. We set $a_1 > 0$ to ensure boundedness and let the sign of a_0 control monotonicity of the exponential function.

For the rest of the paper, we consider the following model:

$$
\begin{aligned}
Q_t &= \mu + \sigma_t Y_t^{1/\alpha_t}, && (2.4)\\
\log\sigma_t &= \beta_0 + \beta_1\log\sigma_{t-1} - \beta_2\exp(-\beta_3 Q_{t-1}), && (2.5)\\
\log\alpha_t &= \gamma_0 + \gamma_1\log\alpha_{t-1} + \gamma_2\exp(-\gamma_3 Q_{t-1}), && (2.6)
\end{aligned}
$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables, $0 \leq \beta_1 \neq \gamma_1 < 1$, $\beta_2 > 0$, $\beta_3 > 0$, $\gamma_2 > 0$, and $\gamma_3 > 0$.

One alternative for $\eta_1(\cdot)$ and $\eta_2(\cdot)$ is the widely used generalized autoregressive score (GAS) models by Harvey and Chakravarty (2008) and Creal et al. (2013), which has been successfully employed in the literature of dynamic POT-GPD models, see Massacci (2016) and Zhang and Schwaab (2017). However, as shown in Section 2.7.5, in our dynamic GEV context, the $\eta_1(\cdot)$ and

$\eta_2(\cdot)$ implied under the GAS framework are complicated and may not be monotone, and thus may suffer from lack of interpretability.

The exponential function is simple, flexible and at the same time intuitive and interpretable. Although the exponential function implies an upper bound for the $\{\sigma_t\}$ and $\{\alpha_t\}$ process[3], as is demonstrated in Hansen (1994) for logistic functions, the boundedness does not affect the flexibility of the model but facilitates numerical and technical tractability. Moreover, as shown in Section 2.5, the current model can flexibly capture dynamics of both the scale parameter and the tail index, and offers an accurate modeling of the maxima in financial time series. Extensions of AcF to allowing μ_t to vary more freely can be implemented by imposing an ARMA structure on μ_t with added complexity and potential model instability. A further justification of AcF is given in Section 2.2.4 under a general factor model setting.

To our best knowledge, this is the first formal presentation of dynamic GEV model that offers a complete dynamic treatment for both the scale parameter σ_t and the tail index α_t. In contrast to the static GEV, AcF is a time series model of the *conditional* maxima. Given all the past information $\mathcal{F}_{t-1}$, the conditional distribution of maxima Q_t is Fréchet(μ, σ_t, α_t), where (μ, σ_t, α_t)$\in \mathcal{F}_{t-1}$.

Remark 1: AcF can be easily extended to include q_1 autoregressive terms of $\log \sigma_t$ and $\log \alpha_t$, and q_2 lagged terms of $\eta(Q_t)$, similar to that of GARCH(q_1, q_2) model. Similar theoretical properties can be derived and similar estimation procedure can be used. Our empirical experience shows that the extension does not necessarily improve the performance of the model, but instead induces instability in estimation. A similar phenomenon has been observed in Creal et al. (2013) and Zhang and Schwaab (2017) for POT-GPD model. In this paper we focus on AcF(1,1) model.

Remark 2: The choice of $\eta_1(\cdot)$ and $\eta_2(\cdot)$ may require further consideration when the model is used for other applications, as the exponential function used here is designed to accommodate volatility clustering and heavy tail clustering for financial applications. An increasing $\eta_1(\cdot)$ produces volatility clustering and a decreasing $\eta_2(\cdot)$ produces heavy-tail clustering. However, if it is observed that an extreme event tends to be followed by a period of 'normal' activities, then $\eta_2(\cdot)$ may be assumed to be an increasing function. In general, as long as $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are continuous bounded

[3]As shown in Section 2.7.5, the $\eta_1(\cdot)$ and $\eta_2(\cdot)$ implied by GAS also give an upper bound on the $\{\sigma_t\}$ and $\{\alpha_t\}$ process.

functions, the probabilistic properties shown below still hold and the same estimation procedure can be applied.

Remark 3: We have assumed that the conditional distribution of Q_t is of Fréchet type since the main focus here is on financial applications. It can be extended to other types of GEVs. In some cases a proper transformation can be used. For example, if a random variable X follows Gumbel(μ, σ), then $\exp(X)$ is Fréchet with location parameter 0, scale parameter $\exp(\mu)$ and tail index $1/\sigma$. Hence, if $\{X_{it}\}_{i=1}^{p}$ are in the Domain of Attraction of Gumbel distribution (Type I GEV), an exponential transformation of the data can be modeled with AcF.

2.2.3 Stationarity and ergodicity

The evolution schemes (2.5) and (2.6) can be written as

$$\begin{aligned}\log \sigma_t &= \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3(\mu + \sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})),\\ \log \alpha_t &= \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3(\mu + \sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})),\end{aligned}$$

where $\{Y_t\}$ is an *i.i.d.* sequence of unit Fréchet random variables. Hence $\{\sigma_t, \alpha_t\}$ form a homogeneous Markov chain in $\mathbb{R}^2$. The following theorem gives a general sufficient condition under which the process is stationary and ergodic.

Theorem 2.2.1. *For an AcF with $\beta_2, \beta_3, \gamma_2, \gamma_3 > 0, \beta_0, \gamma_0, \mu \in \mathbb{R}$, and $0 \le \beta_1 \ne \gamma_1 < 1$, the latent process $\{\sigma_t, \alpha_t\}$ is stationary and geometrically ergodic.*

The proof can be found in Section 2.7.1. Since $\{Q_t\}$ is a coupled process of $\{\sigma_t, \alpha_t\}$ through (2.4), $\{Q_t\}$ is also stationary and ergodic. Unlike GARCH model, the stationarity of AcF mainly requires the autoregressive coefficient $0 \le \beta_1, \gamma_1 < 1$ with no restriction on the parameters associated with the shock process Q_{t-1}. This is due to the boundedness of $\eta_1(Q_{t-1})$ and $\eta_2(Q_{t-1})$.

2.2.4 AcF under a factor model setting

In this section, we illustrate that the limiting form of maxima Q_t under a general factor model framework leads to an AcF model. Assume $\{X_{it}\}_{i=1}^{p}$ follow a general factor model,

$$X_{it} = f_i(Z_{1t}, Z_{2t}, \ldots, Z_{dt}) + \sigma_{it}\varepsilon_{it},$$

where $\{X_{it}\}_{i=1}^{p}$ are observed time series at time t, $\{Z_{1t}, Z_{2t}, \ldots, Z_{dt}\}$ consist of observed and unobserved factors, $\{\varepsilon_{it}\}_{i=1}^{p}$ are *i.i.d.* random noises that are independent with the factors $\{Z_{it}\}_{i=1}^{d}$, and $\{\sigma_{it}\}_{i=1}^{p} \in \mathcal{F}_{t-1}$ are the conditional volatilities of $\{X_{it}\}_{i=1}^{p}$. The function $f_i : \mathbb{R}^d \to \mathbb{R}$ is a Borel function.

This general factor model has been widely used for analyzing high dimensional panel data. The (dynamic) factor models of Bai and Ng (2002), Geweke (1977), Stock and Watson (2011), Lam and Yao (2012), and many others assume unobservable factors. Asset pricing models of Sharpe (1964), Mossin (1966), Fama and French (1993), and others use observable factors.

One fundamental characteristic of many financial time series is that they are often heavy-tailed. To incorporate this observation, we make the common assumption that the random noise $\{\varepsilon_{it}\}_{i=1}^{p}$ are *i.i.d.* random variables in the Domain of Attraction of Fréchet distribution (Leadbetter et al., 1983). Specifically, we adopt the following definition:

Definition 1 (Leadbetter et al. (1983))**.** *A random variable ε is in the Domain of Attraction of Fréchet distribution with tail index α if and only if $x_F = \infty$ and $1 - F_\varepsilon(x) \sim l(x)x^{-\alpha}$, $\alpha > 0$, where F_ε is the cdf of ε, $l(x)$ is a slowly-varying function and $x_F = \sup\{x : F_\varepsilon(x) < 1\}$. Here and after, for two positive functions $m_1(x)$ and $m_2(x)$, $m_1(x) \sim m_2(x)$ means $\frac{m_1(x)}{m_2(x)} \to 1$, as $x \to \infty$.*

Distributions in Domain of Attraction of Fréchet distribution include a broad class of random variables such as Cauchy, Lévy, Pareto and t distributions. To facilitate algebraic derivation, we further assume that for ε_{it}, $l_t(x) \to K_t$ as $x \to \infty$, where $K_t \in \mathcal{F}_{t-1}$ is a positive constant. This is a rather weak assumption with all the aforementioned random variables satisfying this condition. Since K_t can be incorporated into each σ_{it}, without loss of generality, we set $K_t = 1$ in the following. Under a dynamic model, we assume that the conditional tail index α_t of ε_{it} evolves through time according to certain dynamics (e.g. (2.6)) and $\alpha_t \in \mathcal{F}_{t-1}$.

We also assume that

$$\sup_{1\le p<\infty}\sup_{1\le i\le p}|f_i(Z_{1t},Z_{2t},\ldots,Z_{dt})|<\infty \quad a.s.$$

Notice here the supremum is taken over p with the number of latent factors d fixed. This is a mild assumption and it includes all the commonly encountered factor models. For example, if the factor model takes a linear form, $f_i(Z_{1t},\ldots,Z_{dt})=\sum_{j=1}^{d}\beta_j^{(i)}Z_{jt}$, a sufficient condition for the assumption to hold would be $\sup_{1\le p<\infty}\sup_{1\le i\le p}\|\beta^{(i)}\|<\infty$.

We further assume that

$$\lim_{p\to\infty}\sum_{i=1}^{p}\sigma_{it}^{\alpha_t}=\infty \text{ and } \lim_{p\to\infty}\sup_{1\le i\le p}\frac{\sigma_{it}^{\alpha_t}}{\sum_{j=1}^{p}\sigma_{jt}^{\alpha_t}}=0.$$

Intuitively, it means the magnitudes of conditional volatility σ_{it} are comparable to each other and there is no single X_{it} that dominates the total volatility. For example, if $\sigma_{it}=c_i\sigma_t$ and all c_i's are in a compact positive interval, then the assumption holds.

Given the assumptions, the following result gives the asymptotic conditional distribution of maxima $Q_t=\max_{1\le i\le p}X_{it}$ when p goes to infinity.

Proposition 2.2.1. *Given $\mathcal{F}_{t-1}$, denote $a_{pt}=0$ and $b_{pt}=(\sum_{i=1}^{p}\sigma_{it}^{\alpha_t})^{1/\alpha_t}$, we have, as $p\to\infty$,*

$$\frac{Q_t-a_{pt}}{b_{pt}}\xrightarrow{d}\Psi_{\alpha_t}(x), \tag{2.7}$$

where $\Psi_{\alpha_t}(x)$ is a Fréchet type random variable with tail index α_t and $\Psi_{\alpha_t}(x)=\exp(-x^{-\alpha_t})$.

The proof of Proposition 2.2.1 can be found in Section 2.7.2. Proposition 2.2.1 shows that under the framework of the general factor model and some mild conditions, the conditional distribution of maxima $Q_t=\max_{1\le i\le p}X_{it}$ can be well approximated by a Fréchet distribution. In terms of stochastic representation, (2.7) can be rewritten as $Q_t\approx\sigma_tY_t^{1/\alpha_t}$, where Y_t is a unit Fréchet random variable and $\sigma_t=b_{pt}$. To be more flexible and accurate in finite samples, a location parameter μ_t can be included. That is,

$$Q_t\approx\mu_t+\sigma_tY_t^{1/\alpha_t}, \tag{2.8}$$

where $(\mu_t,\ \sigma_t,\ \alpha_t)$ are time-varying parameters. Setting $\mu_t=\mu$ for parsimonious modeling, we

obtain the dynamic structure of $\{Q_t\}$ specified in (2.4).

Remark 4: In the general factor model, the cross-sectional dependence among X_{it}'s, such as tail dependence, can be introduced by the factor structures. Notice that the independence assumption on ε_{it}'s is not essential for Proposition 2.2.1. Based on the results of maxima in stationary series in Leadbetter et al. (1983), similar and more elaborate results can be derived if we impose a stationarity assumption or block independence assumption on ε_{it}'s.

Remark 5: The assumption that $\{X_{it}\}_{i=1}^{p}$ share the same tail index α_t may seem to be strong, however, van Oordt and Zhou (2016) found it reasonable in financial applications. See Kelly (2014) and Kelly and Jiang (2014) for a similar assumption. A more involved version of Proposition 2.2.1 is also available upon request for handling the case that $\{X_{it}\}_{i=1}^{p}$ may have different tail behavior.

Remark 6: The general factor model considered in this section is just an example whose limiting form of maxima coincides with our proposed AcF model. In this paper we are not focusing on the general factor model. Instead, we focus on AcF models.

2.3 Irregular Maximum Likelihood Estimation

We denote all the parameters in the model by $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu)$ and denote $\Theta_s = \{\theta | \beta_0, \gamma_0, \mu \in \mathbb{R}, 0 \leq \beta_1, \gamma_1 < 1, \beta_2, \beta_3, \gamma_2, \gamma_3 > 0\}$. In the following, we assume that all allowable parameters are in Θ_s and denote the true parameter by $\theta_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0, \mu_0)$.

The conditional p.d.f. of Q_t given $(\mu, \sigma_t, \alpha_t)$ is

$$f_t(\theta) = f(Q_t|\sigma_t, \alpha_t) = \alpha_t \sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t+1)} \exp\left\{-\sigma_t^{\alpha_t} (Q_t - \mu)^{-\alpha_t}\right\}.$$

Hence, by conditional independence, the log-likelihood function with observations $\{Q_t\}_{t=1}^{n}$ is

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) = \frac{1}{n}\sum_{t=1}^{n} \left[\log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1)\log(Q_t - \mu) - \sigma_t^{\alpha_t} (Q_t - \mu)^{-\alpha_t}\right],$$

where $\{\sigma_t, \alpha_t\}_{t=1}^{n}$ can be obtained recursively through (2.5) and (2.6), with an initial value (σ_1, α_1).

Notice here the true value of (σ_1, α_1), denoted as (σ_1^0, α_1^0), is unknown since the state variables $\{\sigma_t, \alpha_t\}$ is a hidden processes. Fortunately, with $0 \leq \beta_1, \gamma_1 < 1$, the influence of (σ_1, α_1) on

future (σ_t, α_t) decays exponentially as t increases, hence its impact on parameter estimation will be minimum with a sufficiently large sample size. Theorems 2.3.1 and 2.3.2 below show that the consistency and asymptotic normality of MLE do not depend on whether (σ_1^0, α_1^0) is known and the asymptotic distribution does not depend on the initial value (σ_1, α_1). For simplicity, we use the estimated $(\hat{\sigma}, \hat{\alpha})$ from the static GEV as the initial value for (σ_1, α_1).

Denote the log-likelihood function based on an arbitrary $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ as $\tilde{L}_n(\theta)$. Theorems 2.3.1 and 2.3.2 show that there always exists a sequence $\hat{\theta}_n$, which is a local maximizer of $\tilde{L}_n(\theta)$, such that $\hat{\theta}_n$ is consistent and asymptotically normal, regardless of the initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$.

Theorem 2.3.1. *(Consistency) Assume the parameter space Θ is a compact set of Θ_s. Suppose the observations $\{Q_t\}_{t=1}^n$ are generated by a stationary and ergodic AcF with true parameter θ_0 and θ_0 is in the interior of Θ, then there exists a sequence $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ such that $\hat{\theta}_n \to_p \theta_0$ and $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$. Hence $\hat{\theta}_n$ is consistent.*

By the differentiability of $\tilde{L}_n(\theta)$ with respect to θ, the sequence $\hat{\theta}_n$ is also the solution to the score function $\frac{\partial \tilde{L}_n}{\partial \theta}(\theta) = 0$. Theorem 2.3.1 guarantees the existence of a sequence of consistent MLE $\hat{\theta}_n$ and is a result about the local behavior of the likelihood function $\tilde{L}_n(\theta)$ near the true parameter value θ_0. The uniqueness of MLE remains an open question due to the complication brought by μ. The same difficulty also applies to the MLE of the static GEV as is noted in Smith (1985). Proposition 2.3.1 gives a partial answer to the asymptotic uniqueness of MLE.

Proposition 2.3.1. *(Asymptotic uniqueness) Denote $V_n = \{\theta \in \Theta | \mu \leq cQ_{n,1} + (1-c)\mu_0\}$ where $Q_{n,1} = \min_{1 \leq t \leq n} Q_t$, under the conditions in Theorem 2.3.1, for any fixed $0 < c < 1$, there exists a sequence of $\hat{\theta}_n = \arg\max_{\theta \in V_n} \tilde{L}_n(\theta)$ such that, $\hat{\theta}_n \to_p \theta_0$, $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$, and $P(\hat{\theta}_n$ is the unique global maximizer of $\tilde{L}_n(\theta)$ over $V_n) \to 1$.*

Since $\tilde{L}_n(\theta)$ is defined on $Q_t > \mu$, the parameter space for the maximization of $\tilde{L}_n(\theta)$ is actually $\Theta_n = \{\theta \in \Theta | \mu < Q_{n,1}\}$. Note that for any $0 < c < 1$, $V_n \subseteq \Theta_n$ since $\mu_0 < cQ_{n,1} + (1-c)\mu_0 < Q_{n,1}$. Proposition 2.3.1 states that there is an asymptotic unique MLE over V_n, where V_n can be made arbitrarily close to Θ_n by the fact that $Q_{n,1} \searrow \mu_0$ *a.s.* and by setting c close to 1. In practice, we take $\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \tilde{L}_n(\theta)$. Numerical experiments confirms its good performance under finite sample.

Theorem 2.3.2. *(Asymptotic normality) Under the conditions in Theorem 2.3.1, we have* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M_0^{-1})$*, where* $\hat{\theta}_n$ *is that in Theorem 2.3.1 and* M_0 *is the Fisher Information matrix evaluated at* θ_0*. Further, the sample variance of plug-in estimated score functions* $\{\frac{\partial}{\partial\theta} l_t(\hat{\theta}_n)\}_{t=1}^n$ *is a consistent estimator of* M_0*.*

The proofs of Theorems 2.3.1 and 2.3.2 and Proposition 2.3.1 can be found in Section 2.7.3. The main technical difficulty is that the MLE here is irregular in the sense that μ affects the support of the observations Q_t (since $Q_t > \mu$), so regularity conditions of standard MLE are violated. Another technical complication is caused by the fact that the true initial value (σ_1^0, α_1^0) is unknown. In the literature, Smith (1985) proves consistency and asymptotic normality for such irregular MLE for a wide range of distributions, including Fréchet distribution, with *i.i.d.* observations. See Dombry (2015), Bücher and Segers (2016), and Bücher and Segers (2017) for a recent development on this topic. Our proof extends the theoretical result to a dynamic model for dependent time series under stationary and ergodic conditions. We note that this is the first formal treatment for statistical properties of MLE under dynamic EVT framework. The technical tools developed here can also be used for dynamic POT-GPD models.

2.4 Simulation Study

2.4.1 Convergence of maxima in factor model

In this section, we conduct numerical experiments to investigate the finite sample behavior of Q_t described in Proposition 2.2.1. Specifically, we study the convergence of the marginal distribution of Q_t to its Fréchet limit under a one-time period factor model. To simplify notation, we drop the time index t in this section. We simulate data from the following one-factor linear model,

$$X_i = \beta_i Z + \sigma_i \varepsilon_i, \qquad i = 1, \ldots, p,$$

where $Z \sim N(0,1)$ is the latent factor, β_i's are *i.i.d.* random coefficients generated from a uniform distribution $U(-2,2)$ and ε_i's are *i.i.d.* t-distributions with degrees of freedom ν. The σ_i's are *i.i.d.* random variables generated from a mixture of uniform distribution $\frac{1}{2}U(0.5,1.5) + \frac{1}{2}U(0.75,1.25)$ such that most σ_i's are moderate in $(0.75,1.25)$ and the ratio between maximum and minimum σ_i's

are 3. This setting roughly matches the pattern of volatilities of different stocks in S&P100 index. For t-distribution, ν corresponds to the tail index α in Definition 1. We set $Q = \max_{1\leq i\leq p}(X_i)$.

We compare the finite sample empirical distribution of Q and its corresponding Fréchet limit stated in Proposition 2.2.1 under different ν and p. For each (ν, p) combination, 1000 sets of *i.i.d.* $\{X_i\}_{i=1}^p$ are generated, resulting in 1000 sampled $Q = \max_{1\leq i\leq p}(X_i)$. Figure 2.2 plots the empirical cdf of the normalized Q in (2.7) along with the corresponding limiting Fréchet distribution. It is clearly seen that as p increases, the empirical distribution of Q approaches its Fréchet limit. A large ν requires larger p for accurate approximation. We also conduct experiments with t-distributed latent factors Z and observe similar results.

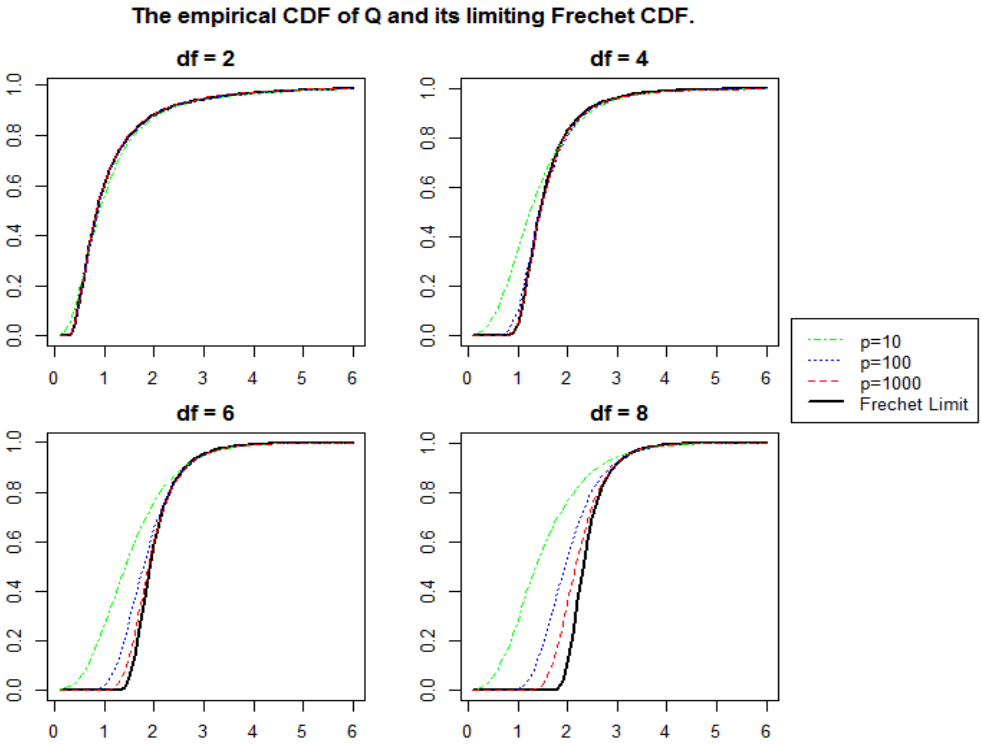

Figure 2.2: *Finite sample empirical distribution of the maxima Q and its corresponding Fréchet limit, with different combinations of p and degrees of freedom (df) of the t-distribution in the factor model.*

2.4.2 AcF estimation for conditional VaR of maxima

In this section, we investigate the temporal approximation ability of AcF to the maxima $\{Q_t\}$ process from a general factor model in terms of 1-day conditional Value at Risk (cVaR). cVaR is the most commonly used measure for tail risk in financial applications. For $0 < q < 1$, cVaR_t^q is defined as the $1-q$ extreme quantile of Q_t given all past information $\mathcal{F}_{t-1}$, where q is often taken to be 0.1, 0.05 or 0.01. Here, we model the $\{Q_t\}$ process using AcF and calculate the corresponding cVaR_t^q for Q_t using the fitted AcF.

Specifically, we simulate the $\{Q_t\}$ process from a similar one-factor linear model as in Section 2.4.1,

$$X_{it} = 0.009(\beta_i Z_t + \sigma_i \varepsilon_{it}), \quad i = 1, \ldots, p; \;\; t = 1, \ldots, T,$$

where $Z_t \sim N(0,1)$ is the latent factor, β_i's are *i.i.d.* random coefficients generated from $U(-2,2)$, σ_i's are *i.i.d.* random variables generated from a mixture of uniform distribution $\frac{1}{2}U(0.5,1.5) + \frac{1}{2}U(0.75,1.25)$ and ε_{it}'s are *i.i.d.* t-distributions with degrees of freedom ν_t. The multiplier 0.009 is used to control the magnitude of X_{it} to be at the same level of typical stock returns. We fix $p = 100$ and change the observation length T throughout this section. For each day t we obtain $Q_t = \max_{1 \le i \le p} X_{it}$. We allow ν_t to evolve, following

$$\log \nu_t = \gamma_0 + \gamma_1 \log \nu_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1}),$$

which resembles the tail index evolution scheme in AcF. Note that the volatility of ε_{it} also evolves implicitly through time due to the dynamics of ν_t. We set the parameters to be $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (-0.1, 0.9, 0.3, 5)$ such that the typical range of ν_t is $[2, 6]$.

We use AcF to model the simulated $\{Q_t\}_{t=1}^{T}$ process and assess the goodness of approximation by AcF's out-sample performance on predicting 1-day cVaR for Q_t. Specifically, we first fit AcF based on the training set $\{Q_t\}_{t=1}^{T_1}$. Then using the fitted AcF, we calculate cVaR_t^q for each Q_t on the test set $\{Q_t\}_{t=T_1+1}^{T_1+T_2}$. The true $\{Q_t\}_{t=T_1+1}^{T_1+T_2}$ are then compared with the $\{\text{cVaR}_t^q\}_{t=T_1+1}^{T_1+T_2}$ and the number of violations are recorded. A violation happens when the observed daily maxima Q_t is larger than the corresponding cVaR_t^q given by AcF. If AcF approximates the tail behavior of $\{Q_t\}$ process well, the expected proportion of violations in the test set should be close to q.

We set $T_1 = 1000, 2000, 5000$, $T_2 = 100$ and $q^0 = 0.1, 0.05, 0.01$. For each combination of (T_1, T_2, q^0), we repeat the experiment 500 times. The ith experiment gives a realized violation percentage q_i and we report the average percentage, $\bar{q} = \sum_{i=1}^{500} q_i/500$, in Table 2.1. We also report the p-value for testing $E(q_i) = q^0$ using one-sample Z-test based on $\{q_i\}_{i=1}^{500}$. It is clear that the 1-day

T_1	$\bar{q}$ $(q^0 = 0.1)$	p-value	$\bar{q}$ $(q^0 = 0.05)$	p-value	$\bar{q}$ $(q^0 = 0.01)$	p-value
1000	0.095	0.104	0.049	0.758	0.012	0.093
2000	0.096	0.229	0.049	0.748	0.012	0.110
5000	0.097	0.340	0.051	0.613	0.012	0.077

Table 2.1: *The performance of AcF on approximation of 1-day conditional VaR for* $\{Q_t\}$ *process with independent errors* ε_{it}.

cVaR given by AcF performs well. Also, a larger training set tends to produce better performance.

To further investigate the performance of AcF when ε_{it}'s are dependent, we repeat the experiment for the case where ε_{it}'s are generated from multivariate t-distributions. Specifically, we assume ε_{it}'s are generated from 10 different multivariate t-distributions of size 10. There are 45 pairwise correlations in the correlation matrix of each multivariate t-distribution, 30 of them are generated from $U(0, 0.3)$, 10 are from $U(0.3, 0.4)$ and 5 are from $U(0.4, 0.5)$. For each day t, the 100 ε_{it}'s are generated independently from the 10 multivariate t-distributions with degrees of freedom ν_t and corresponding correlation matrices. Note that marginally each ε_{it} is still a t-distribution with degrees of freedom ν_t. We keep all the other settings unchanged and report the result in Table 2.2. Again, AcF performs well in terms of 1-day cVaR.

T_1	$\bar{q}$ $(q^0 = 0.1)$	p-value	$\bar{q}$ $(q^0 = 0.05)$	p-value	$\bar{q}$ $(q^0 = 0.01)$	p-value
1000	0.094	0.053	0.048	0.308	0.012	0.104
2000	0.097	0.270	0.047	0.154	0.011	0.460
5000	0.096	0.081	0.048	0.222	0.011	0.189

Table 2.2: *The performance of AcF on approximation of 1-day conditional VaR for* $\{Q_t\}$ *process with dependent errors* ε_{it}.

All together, it indicates that AcF can accurately approximate the tail behavior of the maxima $\{Q_t\}$ process that originates from a general factor model and AcF is robust under misspecification of the scale parameter.

2.4.3 Performance of the maximum likelihood estimator

To study the finite sample performance of the MLE, we simulate data from an AcF with the following parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu) = (-0.050, 0.96, -0.051, 6.68, -0.068, 0.89, 0.33, 5.33, -0.069)$. This set of parameters is the MLE obtained from an analysis of the S&P100 returns using AcF, shown in Section 2.5.1. Under this setting, the typical range of α_t is [2, 8] and the typical range of σ_t is [0.06, 0.21].

We investigate the performance of MLE and the corresponding confidence intervals with sample sizes $n = 1000, 5000, 10000$. For each sample size, we conduct 500 experiments. Table 2.3 shows the average of the estimates, the standard deviation from the 500 experiments, and the percentage of estimates that fall into the various confidence intervals based on the asymptotic theory. It is seen

$N = 1000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
$True\ value$	-0.068	0.890	0.330	5.33	-0.050	0.960	-0.051	6.68	-0.069
$Mean$	-0.060	0.884	0.346	6.28	-0.051	0.956	-0.054	5.88	-0.066
$S.D.$	0.029	0.028	0.058	1.93	0.028	0.019	0.023	3.25	0.011
90% $C.I.$	81	82	90	91	85	81	75	78	88
95% $C.I.$	84	88	93	94	87	87	79	80	95
99% $C.I.$	88	92	97	97	95	94	87	85	98
$N = 5000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
$True\ value$	-0.068	0.890	0.330	5.33	-0.050	0.960	-0.051	6.68	-0.069
$Mean$	-0.066	0.889	0.332	5.52	-0.051	0.959	-0.052	6.53	-0.069
$S.D.$	0.014	0.012	0.029	0.88	0.012	0.008	0.009	1.83	0.005
90% $C.I.$	88	87	90	85	88	87	88	87	86
95% $C.I.$	92	96	93	94	92	91	93	93	94
99% $C.I.$	95	99	98	99	98	98	97	97	99
$N = 10000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
$True\ value$	-0.068	0.890	0.330	5.33	-0.050	0.960	-0.051	6.68	-0.069
$Mean$	-0.067	0.890	0.330	5.44	-0.050	0.960	-0.051	6.55	-0.069
$S.D.$	0.010	0.007	0.018	0.61	0.007	0.005	0.006	1.37	0.003
90% $C.I.$	90	88	88	85	89	89	86	89	90
95% $C.I.$	93	94	94	94	92	94	93	94	98
99% $C.I.$	98	100	100	99	97	98	98	98	99

Table 2.3: *Numerical results for performance of MLE with sample size* $1000, 5000, 10000$. *Mean and S.D. are the sample mean and standard deviation of the MLE's obtained from 500 simulations. 90% C.I. reports the coverage rate of the 90% C.I. constructed from the estimated Fisher Information matrix; 95% C.I. and 99% C.I. report corresponding coverage rates.*

that MLE is consistent and its accuracy increases with the sample size. Also, the coverage rate of

the asymptotic confidence interval is quite close to the target, validating the asymptotic properties presented in Section 2.3.

2.5 Real Data Applications

In this section, we present two real data applications of AcF, one on the cross-sectional maxima of negative log-returns of stocks in two major U.S. stock indices and one on the intra-day maxima of negative log-returns from high-frequency foreign exchange trading. In both cases, AcF shows its superiority over the static GEV for modeling maxima and its ability to reveal the time-varying nature of the financial market tail risk. Moreover, AcF demonstrates its potential usefulness as a market tail risk measure and an early warning signal for potential extreme movement in the financial market.

2.5.1 Cross-sectional maxima of the negative daily log-returns of stocks in S&P100 index and DJI30 index

In this section, we consider the cross-sectional maxima of the negative daily log-returns (i.e. daily losses) of component stocks in S&P100 Index (hereafter S&P100) and Dow Jones Industrial Average Index (hereafter DJI30) respectively. For both indices, the time horizon we consider here is from January 1, 2000 to December 31, 2014. The S&P100 index includes 100 leading U.S. stocks and represents about 45% of the market capitalization of the U.S. equity market. The DJI30 index is a major U.S. stock index, consisting of 30 largest publicly owned companies based in U.S. Both indices are arguably the most important U.S. financial market indicators. To maintain a better management of financial risk, it is essential for financial institutions, especially mutual funds and banks, to understand the cross-sectional tail risks of S&P100 and DJI30.

We present the AcF modeling result for S&P100 in detail. For each trading day t we calculate the negative daily log-returns of each component stock in S&P100 to obtain $\{X_{it}\}_{i=1}^{100}$ and then obtain the daily cross-sectional maxima $Q_t = \max_{1\leq i\leq 100} X_{it}$. The time series $\{Q_t\}$ has 3773 observations and is shown in the bottom panel of Figure 2.3.

The estimation result of AcF is summarized in Table 2.4. The estimated autoregressive parameter $\hat{\beta}_1$ for the scale parameter $\{\sigma_t\}$ process is 0.96, which suggests a strong persistence of the $\{\sigma_t\}$

S&P100	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	-0.068	0.890	0.328	5.33	-0.050	0.961	-0.051	6.68	-0.069
S.D.	0.014	0.013	0.063	1.27	0.006	0.004	0.0072	1.01	0.006
DJI30	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	0.023	0.895	0.261	16.32	-0.052	0.964	-0.047	7.38	-0.059
S.D.	0.016	0.013	0.041	3.529	0.005	0.004	0.0066	0.813	0.006

Table 2.4: *MLE for cross-sectional maxima of negative daily log-returns for S&P100 (top) and DJI30 (bottom) from January 1, 2000 to December 31, 2014.*

series. The estimated scale parameter $\{\hat{\sigma}_t\}$ (solid line) is plotted in Figure 2.4. For comparison, we also fit a GARCH(1,1) model for each component stock in S&P100 and plot the daily average volatility given by the GARCH models (dashed line) across the 100 stocks in Figure 2.4. The two series move very closely with each other with an overall correlation of 0.918. It suggests that AcF's dynamic scale parameter σ_t is an accurate measure of market volatility.

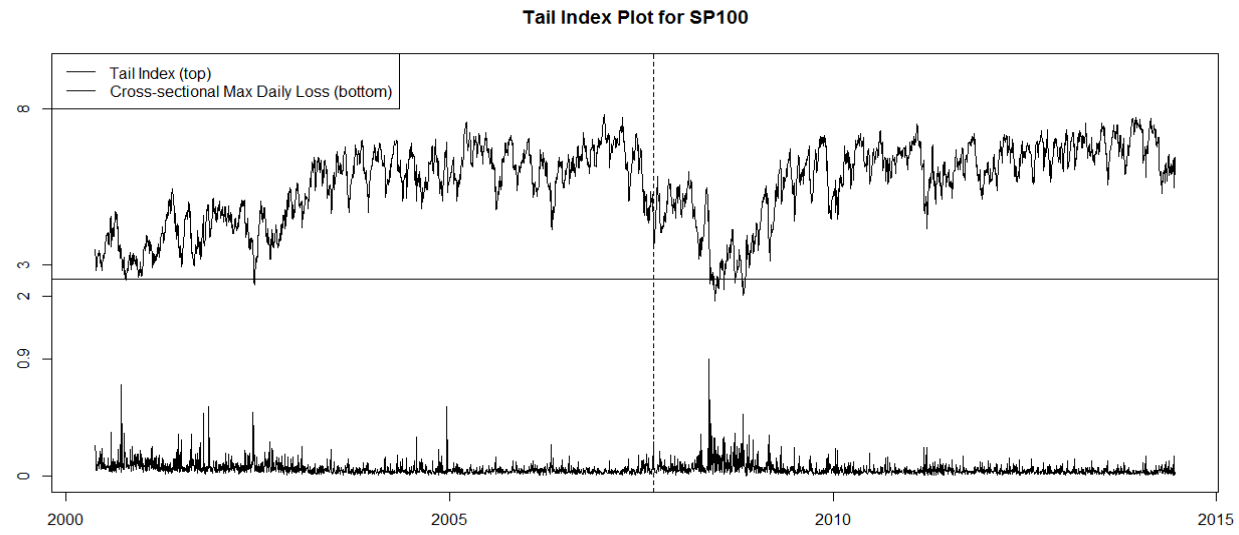

Figure 2.3: *Estimated tail index $\{\hat{\alpha}_t\}$ (top) and cross-sectional maximum negative daily returns $\{Q_t\}$ (bottom) from January 1, 2000 to December 31, 2014 for S&P100 Index. The horizontal solid line is the estimated $\hat{\alpha}$ from the static GEV. The vertical dashed line marks the date when $\hat{\alpha}_t$ is at its lowest for the first time over the past four years.*

The estimated autoregressive parameter $\hat{\gamma}_1$ for the tail index $\{\alpha_t\}$ process is 0.89, indicating a strong persistence in the tail index process. The estimated tail index $\{\hat{\alpha}_t\}$ is shown in the top panel of Figure 2.3. The estimated tail index by AcF is roughly in the range of 2 to 8, which agrees with the empirical finding of Massacci (2016) via a dynamic POT-GPD model. The two periods

of persistent small tail index ($\alpha_t < 4$) coincide with the early 2000's U.S. recession and the 2008 financial crisis. Note that the difference between $\alpha_t = 2$ and $\alpha_t = 8$ is very significant. A Fréchet type random variable has its k-th moment if and only if $\alpha > k$. It is also noted that almost all $\hat{\alpha}_t$'s are greater than 2, hence the conditional mean and variance of the cross-sectional maxima always exist, which agrees with the existing literature (e.g. Hansen, 1994).

The stationary mean of $(\alpha_t, \sigma_t, \mu)$ of the estimated AcF is $(5.73, 0.099, -0.069)$. We also fitted the static GEV model to the data, assuming the Q_t's are *i.i.d.* observations. The estimated parameters are $(\hat{\alpha}, \hat{\sigma}, \hat{\mu}) = (2.56, 0.058, -0.025)$. It is seen that the estimated tail index of the static GEV model is suspiciously low (see Figures 2.1 and 2.3). It is clear that the static GEV fails to adequately model the time-varying tail risk. On the other hand, the estimated tail index $\hat{\alpha}_t$ given by AcF matches the general pattern of the estimated tail index obtained by the moving-window GEV in the ad-hoc analysis shown in Figure 2.1.

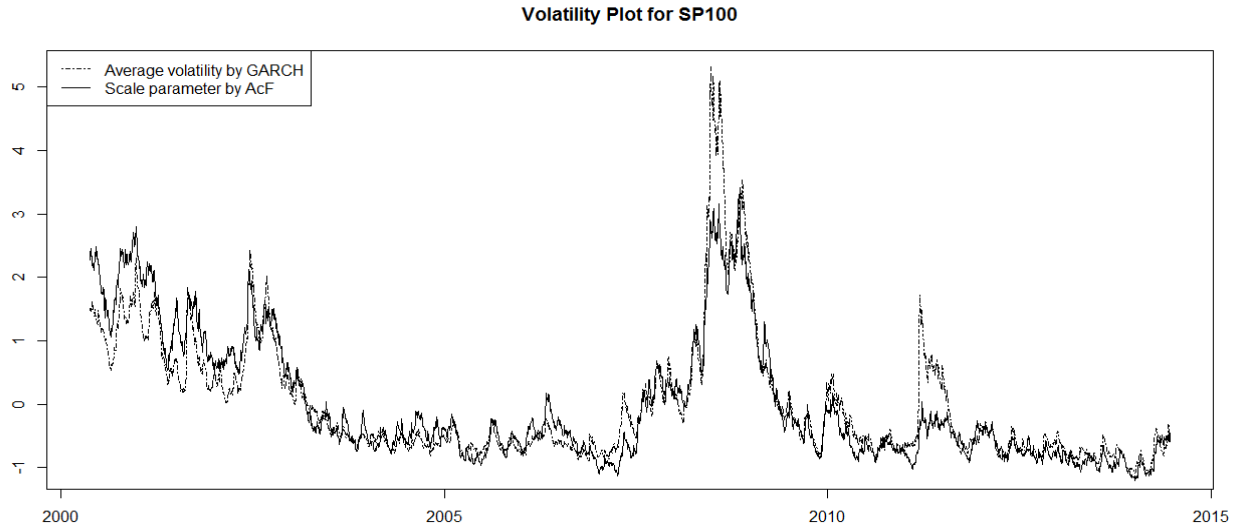

Figure 2.4: *Estimated scale parameter $\{\hat{\sigma}_t\}$ by AcF (solid line) v.s Estimated average volatility by GARCH (dashed line) from January 1, 2000 to December 31, 2014. Both series are standardized to be zero mean and unit variance for comparison.*

As shown in Figure 2.3, there is a clear negative association between the daily maxima $\{Q_t\}$ series and the estimated tail index $\{\hat{\alpha}_t\}$, making $\hat{\alpha}_t$ a useful measure of the underlying market tail risk. We also observe an interesting feature. The vertical dashed line in Figure 2.3 marks the date when tail index $\hat{\alpha}_t$ is at its lowest for the first time over the past four years. The actual date,

December 13, 2007, is in the very beginning stage of the 2008 financial crisis. It shows the ability of AcF to capture the extreme tail movement of financial market in its early stage.

We have also applied the same procedure to DJI30. The estimation result of AcF is shown in Table 2.4 and is similar to the one obtained for S&P100. Due to limited space, we only present its estimated tail index plot (Figure 2.5) here. The typical range of tail index for DJI30 is [2.5, 10], with a slight up-shift compared to the one of S&P100, indicating that the tail risk of DJI30 is lower than that of S&P100. This is reasonable considering that the companies in DJI30 are more stable and well-established than those in S&P100. The correlation between the estimated scale parameter $\{\hat{\sigma}_t\}$ from AcF and the average volatility obtained by fitting GARCH model to each stock is 0.909. It again indicates that the evolution scheme of AcF's scale parameter captures the dynamics of the underlying stock market volatility very well. The date when the tail index of DJI30 is at its

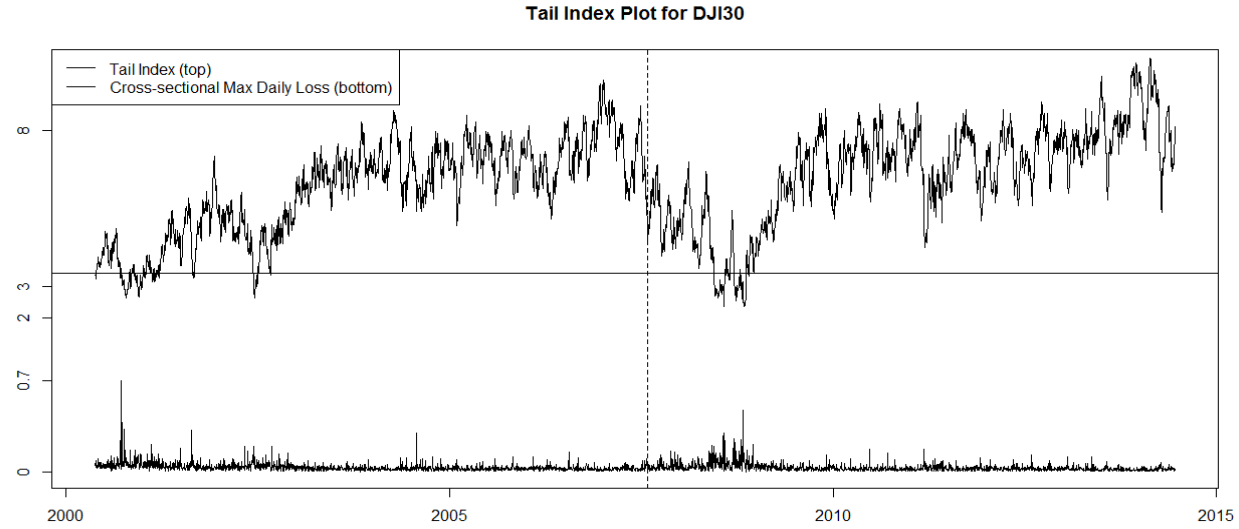

Figure 2.5: *Estimated tail index* $\{\hat{\alpha}_t\}$ *(top) and cross-sectional maximum negative daily return* $\{Q_t\}$ *(bottom) from January 1, 2000 to December 31, 2014 for DJI30 Index. The horizontal solid line is the estimated* $\hat{\alpha}$ *from the static GEV. The vertical dashed line marks the date when* $\hat{\alpha}_t$ *is at its lowest for the first time over the past four years.*

lowest for the first time over the past four years is November 09, 2007, which is very close to the one in S&P100. Together, it shows that AcF has the potential to detect extreme tail risks in the stock market and shows the possibility of using α_t as a risk measure and indicator. More studies are needed to test the notion that an abnormal downwards movement of the tail index, such as reaching a four-year running minimum, can be used as an early warning signal of possible market

crisis.

2.5.2 Intra-day maxima of 3-minute negative log-returns for USD/JPY foreign exchange rate

In this section, we consider the modeling of intra-day maxima of 3-minute negative log-returns from USD/JPY exchange trading. Specifically, we collect the historical 3-minute intra-day exchange rate of USD/JPY from January 1, 2008 to June 26, 2013. The 3-minute negative log-returns $\{X_{it}\}_{i=1}^{p}$ are obtained and intra-day maxima Q_t are calculated. The total length of the series is 1616. The maxima $\{Q_t\}$ series is shown in Figure 2.6(a).

We fit AcF to the intra-day maxima series. Estimated parameters with their standard deviations are shown in Table 2.5. Similar to the result in the stock market, the estimated autoregressive parameter $\hat{\beta}_1$ for $\{\sigma_t\}$ is 0.89, showing a strong persistence of the $\{\sigma_t\}$ series; while the autoregressive parameter $\hat{\gamma}_1$ for $\{\alpha_t\}$ is 0.59, indicating a less persistent tail index series for the foreign exchange market. The stationary mean of $(\alpha_t, \sigma_t, \mu)$ is $(3.47, 0.167, -0.051)$ under the estimated AcF, while $(\hat{\alpha}, \hat{\sigma}, \hat{\mu}) = (3.25, 0.180, -0.068)$ for the static GEV model. The static GEV model gives a relatively smaller estimated tail index.

	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	0.448	0.587	0.658	20.84	-0.120	0.890	-0.195	6.59	-0.051
S.D.	0.144	0.123	0.203	4.52	0.016	0.012	0.024	0.955	0.010

Table 2.5: *MLE for intra-day maxima of 3-minute negative log-returns for USD/JPY from January 1, 2008 to June 26, 2013.*

The estimated tail index $\{\hat{\alpha}_t\}$ is shown in Figure 2.6(b). It is seen that the tail index is small around 2009, showing a riskier foreign exchange market during the financial crisis. Compared to the tail index of stock market, the tail index series here is also more volatile due to the smaller autoregressive parameter $\hat{\gamma}_1$. The range of the tail index is roughly [3, 5], which suggests that the high-frequency trading of USD/JPY has relatively high risks, as observed by Malinowski et al. (2015). We simulate a $\{Q_t\}$ series of length 10000 from the estimated AcF and compare its stationary marginal distributions with the observed series using a quantile-quantile plot in Figure 2.6(c). It confirms that AcF is a suitable model for the series.

We further test the out-sample performance of AcF for predicting 1-day cVaR^q for the intra-day

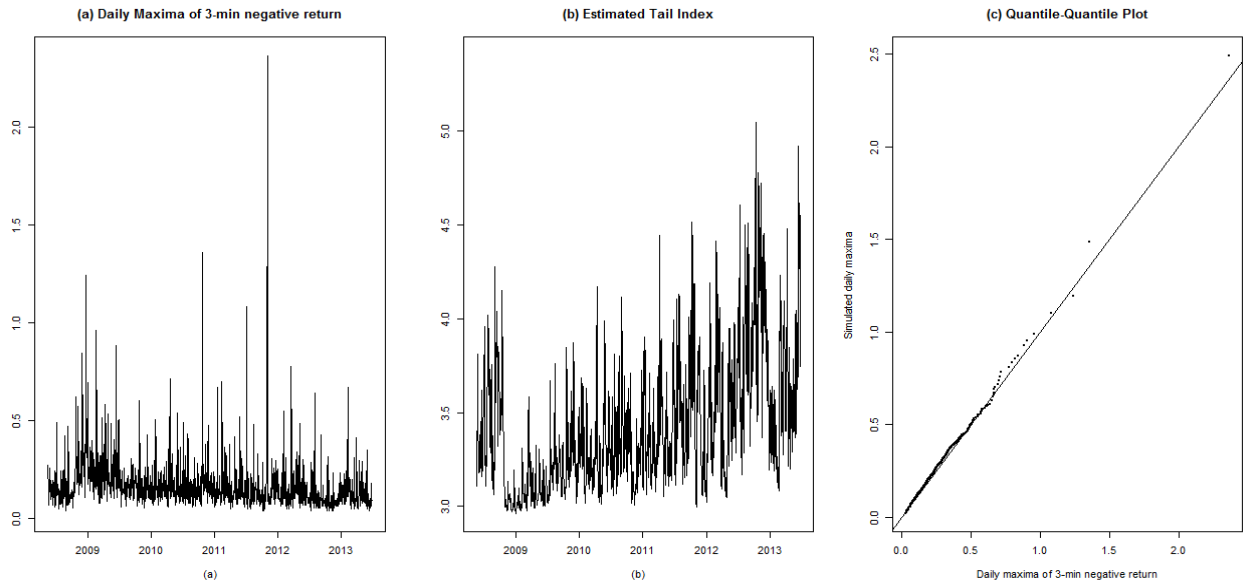

Figure 2.6: *(a) Daily maxima of 3-minute negative log-returns of USD/JPY from January 1, 2008 to June 26 2013; (b) Estimated tail index* $\{\hat{\alpha}_t\}$ *from the fitted AcF; (c) Quantile-quantile plot of real data and simulated data from the fitted AcF.*

maxima. First, we fit AcF using the 1000 observations where $1 \leq t \leq 1000$ (roughly 4 years). For the rest 616 observations where $1001 \leq t \leq 1616$, based on the fitted AcF and past information $\mathcal{F}_{t-1}$, we calculate their 1-day cVaR$^{q^0}$ at $q^0 = 0.1, 0.05, 0.01, 0.005, 0.001$. The true daily maxima are then compared with the estimated cVaR and the number of violations is recorded. For comparison, we also fit the static GEV model using the first 1000 observations and calculate the corresponding 1-day cVaR for the rest 616 observations.

Table 2.6 shows the comparison results. For each q^0, the table presents the number of expected violations ($616q^0$) and the number of actual violations. We also report the p-value of a binomial test for the hypothesis that the actual violation probability and the corresponding q^0 are the same. It is clearly seen that the 1-day cVaR based on AcF performs extremely well, with large p-values for all levels of q^0. On the other hand, the static GEV tends to produce much more conservative cVaR estimates. The comparison clearly demonstrates the time-varying nature of the tail index and the importance of having a dynamic structure such that current market condition is incorporated in the estimation of cVaR.

		AcF		static GEV	
$q^0(\%)$	*Expected Violation*	*Violation*	*p-value*	*Violation*	*p-value*
10	61.6	60	0.89	32	0.00
5.0	30.8	35	0.41	17	0.01
1.0	6.2	8	0.41	2	0.10
0.5	3.1	4	0.56	1	0.39
0.1	0.6	0	1.00	0	1.00

Table 2.6: *Result of 1-day cVaR calculated from AcF and static GEV for intra-day maxima of negative log-returns of USD/JPY exchange rate.*

2.6 Conclusion and Future Research

In this paper, we propose a general dynamic GEV framework for the modeling of time series of maxima. By allowing time-varying scale parameter and tail index of a conditional Fréchet distribution, AcF provides a direct modeling of dynamics of maxima in financial time series and offers a new angle to study the tail risk dynamics in financial markets. Probabilistic properties of AcF are investigated. We implement a maximum likelihood estimator for AcF and investigate its asymptotic properties, using a set of unique technical tools due to the irregularity of the MLE. The real data examples illustrate the efficacy of AcF in practice and its potential broad use in financial risk management and other types of tail risk monitoring.

The AcF can be extended to incorporate other structures and phenomenon observed in real applications. One potential extension is to assume a parametric dynamic structure for the location parameter μ. An ARMA model is a natural choice and is currently being studied. The choices for function $\eta(\cdot)$ may need further study when it comes to different applications. Another natural extension is to construct multivariate AcF for multivariate maxima. Incorporating the idea of Chen and Fan (2006a), an extreme value copula can be used to "link" different univariate AcF models and thus provide a natural model for multivariate time series of maxima. One limitation of AcF is that it can not model time series with marginal distributions of Weibull type, a distribution used in certain applications. Further investigation is required.

2.7 Appendix

2.7.1 Proof of stationarity and ergodicity

Proof of Theorem 2.2.1: The proof of Theorem 2.2.1 follows closely the result of Chan and Tong (1994) on non-linear dynamic system. In the following, we assume $\{\sigma_t, \alpha_t\}$ comes from an AcF with parameter $\theta \in \Theta$ as specified in Theorem 2.2.1. Without loss of generality, in the following proof, we assume $\mu = 0$. In AcF, $(\log \sigma_t, \log \alpha_t)$ forms a non-linear dynamic system according to the following equation,

$$\begin{aligned} \log \sigma_t &= \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})), \\ \log \alpha_t &= \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})). \end{aligned}$$

To fit $\{\log \sigma_t, \log \alpha_t\}$ into the framework of Chan and Tong (1994), we reparametrize the autoregressive equations as follows:

$$\begin{aligned} \log \sigma_t &= [\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}] + \left[z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}}))\right], \\ \log \alpha_t &= [\gamma_0 + z_2 + \gamma_1 \log \alpha_{t-1}] + \left[\gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})) - z_2\right], \end{aligned}$$

where z_1 is a positive constant such that $0 < z_1 < \beta_2$ (e.g. we can set $z_1 = \beta_2/2$) and z_2 is a positive constant such that $0 < z_2 = \gamma_2 \exp(\frac{\gamma_3}{\beta_3} \log(\frac{z_1}{\beta_2})) < \gamma_2$. The reason for defining z_1, z_2 as above will be made more clear in the proof of Lemma 2.7.2. Let $\mathbf{X}_t = (\log \sigma_t, \log \alpha_t)$ and

$$\begin{aligned} T(\mathbf{X}_{t-1}) &= [\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}, \gamma_0 + z_2 + \gamma_1 \log \alpha_{t-1}], \\ S(\mathbf{X}_{t-1}, Y_{t-1}) &= \left[z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})), \gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})) - z_2\right], \end{aligned}$$

we can rewrite the nonlinear dynamic system of $(\log \sigma_t, \log \alpha_t)$ as

$$\mathbf{X}_t = T(\mathbf{X}_{t-1}) + S(\mathbf{X}_{t-1}, Y_{t-1}),$$

where $\{Y_t\}$ is a sequence of *i.i.d.* Fréchet innovations.

Following the terminology in Chan and Tong (1994), $T(\cdot)$ admits a compact attractor $\Lambda = \left(\frac{\beta_0-z_1}{1-\beta_1}, \frac{\gamma_0+z_2}{1-\gamma_1}\right)$, which is a singleton in $\mathbb{R}^2$, and the domain of attraction for Λ is $\mathbb{R}^2$. In other words, for any $x \in \mathbb{R}^2$, we have that the iterates $T^n(x) \to \Lambda$ as $n \to \infty$. We further define $G = \left(\frac{\beta_0-\beta_2}{1-\beta_1}, \frac{\beta_0}{1-\beta_1}\right) \times \left(\frac{\gamma_0}{1-\gamma_1}, \frac{\gamma_0+\gamma_2}{1-\gamma_1}\right)$, which is an open rectangle in $\mathbb{R}^2$.

Lemma 2.7.1. *G is absorbing for $\mathbf{X}_t$.*

Proof. We only prove the result for $\log \alpha_t$, the proof for $\log \sigma_t$ is the same. Suppose $\log \alpha_t > \frac{\gamma_0}{1-\gamma_1}$, then $\log \alpha_{t+1} = \gamma_0 + \gamma_1 \log \alpha_t + \gamma_2 \exp(-\gamma_3 Q_t) > \gamma_0 + \gamma_1 \frac{\gamma_0}{1-\gamma_1} = \frac{\gamma_0}{1-\gamma_1}$. Similarly, we can show that $\log \alpha_{t+1} < \frac{\gamma_0+\gamma_2}{1-\gamma_1}$ if $\log \alpha_t < \frac{\gamma_0+\gamma_2}{1-\gamma_1}$. □

To prove the geometric ergodicity of $\mathbf{X}_t$, we only need to verify conditions (a)-(e) of Theorem 1 in Chan and Tong (1994). The verification of conditions (a), (b), (c) and (e) is trivial and thus is omitted. We verify condition (d) here.

Lemma 2.7.2. *For any $x \in G$, 0 is in the support of $|S(x, Y_{t-1})|$ where $|\cdot|$ is the norm of the vector. And there exists a continuous and positive function $r(x)$ for $x \in G$, such that the second step transition probability for $\mathbf{X}_t$, $P^2(x, dy)$, has an absolutely continuous component whose probability density function is positive over $B(T^2(x), r(x))$ where $B(x, \delta)$ denotes the open ball in G with center at x and radius equal to δ.*

Proof. Since $\sigma_{t-1}, \alpha_{t-1} > 0$ and $0 < Y_{t-1} < \infty$, it is easy to see that for any $\mathbf{X}_{t-1}$, there always exists a unique Y^*_{t-1} depending on $\mathbf{X}_{t-1}$ such that $Q^*_{t-1} = \sigma_{t-1}(Y^*_{t-1})^{1/\alpha_{t-1}} = -\frac{1}{\beta_3}\log(\frac{z_1}{\beta_2})$. By the definition of z_1 and z_2 in $|S(\mathbf{X}_{t-1}, Y_{t-1})|$, it can be verified that given $\mathbf{X}_{t-1}$, Y^*_{t-1} is the unique value that makes $S(\mathbf{X}_{t-1}, Y^*_{t-1}) = 0$. Hence for any $x \in G$, 0 is in the support of $|S(x, Y_{t-1})|$. In the following, we denote $Q^* = -\frac{1}{\beta_3}\log(\frac{z_1}{\beta_2})$.

We now show that there exists a positive constant $r(x) = C$ such that $P^2(x, dy)$ has an absolutely continuous component whose probability density function is positive over $B(T^2(x), C)$. Given $\mathbf{X}_{t-1}$, we have for $\mathbf{X}_{t+1} = (\log \sigma_{t+1}, \log \alpha_{t+1})$,

$$\begin{aligned}
\log \sigma_{t+1} &= T^2(\mathbf{X}_{t-1})[1] + [z_1 - \beta_2 \exp(-\beta_3 Q_t)] + \beta_1 [z_1 - \beta_2 \exp(-\beta_3 Q_{t-1})], \\
\log \alpha_{t+1} &= T^2(\mathbf{X}_{t-1})[2] + [\gamma_2 \exp(-\gamma_3 Q_t) - z_2] + \gamma_1 [\gamma_2 \exp(-\gamma_3 Q_{t-1}) - z_2],
\end{aligned}$$

where $T^2(\mathbf{X}_{t-1})[1]$ and $T^2(\mathbf{X}_{t-1})[2]$ denote the first and second component of $T^2(\mathbf{X}_{t-1})$ respectively. Given $\mathbf{X}_{t-1}$, $\mathbf{X}_{t+1}$ is a vector function of (Q_{t-1}, Q_t), thus we denote $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q_{t-1}, Q_t)$. At $(Q^*_{t-1}, Q^*_t) = (Q^*, Q^*)$, we have $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*) = T^2(\mathbf{X}_{t-1})$. It is easy to verify that the determinant of the Jacobian matrix of $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*)$ at (Q^*, Q^*) is $\exp(-(\beta_3 + \gamma_3)Q^*)\beta_2\beta_3\gamma_2\gamma_3(\beta_1 - \gamma_1)$, which is not zero since $\theta \in \Theta$.

By the Inverse Function Theorem, we know that an inverse function to $\mathbf{F}_{\mathbf{X}_{t-1}}(\cdot)$ exists in an open neighborhood of $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*) = T^2(\mathbf{X}_{t-1})$. By the nature of the vector function $\mathbf{F}_{\mathbf{X}_{t-1}}(\cdot)$, $\mathbf{X}_{t-1}$ does not affect the size of the open neighborhood. Thus for all $\mathbf{X}_{t-1} \in G$, we can find a uniform C such that $B(T^2(\mathbf{X}_{t-1}), C)$ is a subset of the open neighborhood. The rest of the proof simply follows from the fact that (Y_{t-1}, Y_t) are *i.i.d.* unit Fréchet random variables and there is a one-to-one relationship between (Y_{t-1}, Y_t) and (Q_{t-1}, Q_t) given $\mathbf{X}_{t-1}$. □

Now we have verified all five conditions of Theorem 1 in Chan and Tong (1994). Hence $\{\log \sigma_t, \log \alpha_t\}$, as a Markov chain on $G \in \mathbb{R}^2$, is stationary and geometrically ergodic. □

2.7.2 Proof of conditional distribution of Q_t in general factor model

Proof of Proposition 2.2.1: In the following, we drop the time index t for notation simplicity. The conditioning on $\mathcal{F}_{t-1}$ is implicit here. The proof follows standard procedure in the extreme value literature by deriving the cdf of $(Q_p - a_p)/b_p$ directly. Here, $Q_p = \max_{1\le i\le p} X_i$, $a_p = 0$ and $b_p = (\sum_{i=1}^p \sigma_i^\alpha)^{1/\alpha}$. We have

$$
\begin{aligned}
&P\left(\frac{Q_p - a_p}{b_p} \le x\right) = P\left(\frac{\max_{1\le i\le p} X_i - a_p}{b_p} \le x\right) = P\left(\max_{1\le i\le p} X_i \le a_p + b_p x\right) \\
&= P\left(f_i(Z_1, Z_2, \ldots, Z_d) + \sigma_i\varepsilon_i \le a_p + b_p x, \text{ for all } 1 \le i \le p\right) \\
&= P\left(\varepsilon_i \le b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i, \text{ for all } 1 \le i \le p\right) \\
&= E\left[P\left(\varepsilon_i \le b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i, \text{ for all } 1 \le i \le p \middle| Z_1, \ldots, Z_d\right)\right] \\
&= E\left(\prod_{i=1}^p F_\varepsilon\left(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i\right)\right),
\end{aligned}
$$

where the last equality follows from the independence between ε_i's and the latent factors Z_i's. By the assumption that

$$\sup_{1\le p<\infty}\sup_{1\le i\le p}|f_i(Z_1,Z_2,\ldots,Z_d)|<\infty \quad a.s.$$

and

$$\lim_{p\to\infty}\sum_{i=1}^{p}\sigma_i^{\alpha}=\infty \text{ and } \lim_{p\to\infty}\sup_{1\le i\le p}\frac{\sigma_i^{\alpha}}{\sum_{j=1}^{p}\sigma_j^{\alpha}}=0,$$

it is easy to see that, for any fixed $x>0$,

$$\lim_{p\to\infty}\inf_{1\le i\le p}(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)=\infty \quad a.s.$$

Together with the assumption that F_ε is in the Domain of Attraction of Fréchet distribution, we have uniformly for all i,

$$\begin{aligned}&F_\varepsilon(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)\sim\\&1-l(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)^{-\alpha}\quad a.s.,\end{aligned}$$

where $\sim$ has the same meaning as in Definition 1. Together with the fact that $\lim_{x\to\infty}l(x)=1$, we have

$$\begin{aligned}&\sum_{i=1}^{p}l(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)^{-\alpha}\\&=\frac{1}{b_p^{\alpha}}\sum_{i=1}^{p}l(b_px/\sigma_i-f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)(x/\sigma_i-b_p^{-1}f_i(Z_1,Z_2,\ldots,Z_d)/\sigma_i)^{-\alpha}\to x^{-\alpha}\quad a.s.,\end{aligned}$$

where the last equality follows from the fact that $b_p\to\infty$ and $\sup_{1\le p<\infty}\sup_{1\le i\le p}|f_i(Z_1,Z_2,\ldots,Z_d)|<\infty$. By the bounded convergence theorem, we have for any fixed $x>0$,

$$P\left(\frac{Q_p-a_p}{b_p}\le x\right)=P\left(\max_{1\le i\le p}X_i\le a_p+b_px\right)\to\exp\left(-x^{-\alpha}\right),\quad \text{as } p\to\infty.$$

□

2.7.3 Proof of consistency and asymptotic normality

To facilitate the proof of Theorems 2.3.1, 2.3.2 and Proposition 2.3.1, we first give several technical lemmas (Lemmas 2.7.3 to 2.7.15). As mentioned in Section 2.3, the main technical difficulty is that the location parameter μ_0 is unknown and the support of Q_t depends on μ_0, so that the standard argument for MLE cannot be directly applied. Also, the true initial value (σ_1^0, α_1^0) is unknown. New uniform convergence results about the log-likelihood function $\tilde{L}_n(\theta)$, its first and second order derivatives need to be established. The main result on uniform convergence is stated in Lemma 2.7.14. Part of the proof follows that in Francq and Zakoian (2004) for MLE of GARCH model.

In the following, we assume the conditions in Theorem 2.3.1 hold, i.e. the parameter space Θ is a compact set of Θ_s and the observations $\{Q_t\}_{t=1}^n$ come from a stationary and ergodic AcF with true parameter θ_0 where θ_0 is in the interior of Θ. We use $Y_{n,k}$ and $Q_{n,k}$ to denote the kth order statistics of $\{Y_t\}_{t=1}^n$ and $\{Q_t\}_{t=1}^n$. In the following, $\tau_n \sim n^{-r}$ means $\tau_n/n^{-r} \to 1$ as $n \to \infty$. We denote the upper bound of β_1, γ_1 in Θ by $C_b < 1$ and use C to denote a generic positive constant.

We first prove the identifiability of AcF in Lemma 2.7.3, which states that each parameter value θ defines a unique AcF.

Lemma 2.7.3. *(Identifiability) If $Q_t(\theta) = Q_t(\theta_0)$ a.s. for all t, then $\theta = \theta_0$. Here a.s. is for the infinite product space generated by $\{\ldots, Y_{-1}, Y_0, Y_1, Y_2, \ldots\}$, where Y_i's are i.i.d. unit Fréchet random variables.*

Proof. We denote $\sigma_t = \sigma_t(\theta)$, $\alpha_t = \alpha_t(\theta)$ and $\sigma_t^0 = \sigma_t(\theta_0)$, $\alpha_t^0 = \alpha_t(\theta_0)$. Suppose there exist θ and θ_0 such that $Q_t(\theta) = Q_t(\theta_0)$ *a.s.*, then

$$\mu_0 + \sigma_t^0 Y_t^{1/\alpha_t^0} = \mu + \sigma_t Y_t^{1/\alpha_t}, \ a.s.$$

Since $Y_{n,1} \searrow 0$ *a.s.*, by the boundedness of (σ_t, α_t) and (σ_t^0, α_t^0), we have $\mu = \mu_0$. After rearrangement,

$$Y_t^{1/\alpha_t^0 - 1/\alpha_t} = \sigma_t/\sigma_t^0, \ a.s.$$

Denote $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, \ldots)$, we know that $Y_t \perp \mathcal{F}_{t-1}$ and $\alpha_t, \alpha_t^0 \in \mathcal{F}_{t-1}$, so the above equation holds if and only if $\sigma_t(\theta) = \sigma_t(\theta_0)$ and $\alpha_t(\theta) = \alpha_t(\theta_0)$ *a.s.* From the autoregressive equation of

$\log \alpha_t$, we know that if $\alpha_t(\theta) = \alpha_t(\theta_0)$ *a.s.*, we have

$$\gamma_0^0 + \gamma_1^0 \log \alpha_{t-1} + \gamma_2^0 \exp(-\gamma_3^0 Q_{t-1}) = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1}).$$

After rearrangement, we have

$$\gamma_0^0 - \gamma_0 + (\gamma_1^0 - \gamma_1) \log \alpha_{t-1} = \gamma_2 \exp(-\gamma_3 Q_{t-1}) - \gamma_2^0 \exp(-\gamma_3^0 Q_{t-1}).$$

By the same argument as above, since $\alpha_{t-1} \in \mathcal{F}_{t-2}$ and $Q_{t-1} \notin \mathcal{F}_{t-2}$, we must have $\gamma_0 = \gamma_0^0$, $\gamma_1 = \gamma_1^0$, $\gamma_2 = \gamma_2^0$ and $\gamma_3 = \gamma_3^0$. Similarly, we can prove that $\beta_0 = \beta_0^0$, $\beta_1 = \beta_1^0$, $\beta_2 = \beta_2^0$ and $\beta_3 = \beta_3^0$. □

Given parameter θ and an initial value (σ_1, α_1), $\{\sigma_t, \alpha_t\}_{t=1}^n$ can be recovered recursively by their autoregressive equations. In the following, we use $\sigma_t(\theta)$, $\alpha_t(\theta)$ (or σ_t, α_t for simplicity) to denote the scale parameter series and the tail index series based on θ and true initial (σ_1^0, α_1^0), and use $\tilde{\sigma}_t(\theta)$, $\tilde{\alpha}_t(\theta)$ (or $\tilde{\sigma}_t$, $\tilde{\alpha}_t$ for simplicity) to denote the ones based on θ and an arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$. We denote the unobserved true hidden process by $\sigma_t(\theta_0), \alpha_t(\theta_0)$ (or σ_t^0, α_t^0 for simplicity).

By the compactness of Θ and the boundedness of $-\beta_2 \exp(-\beta_3 Q_{t-1}), \gamma_2 \exp(-\gamma_3 Q_{t-1})$, there exist uniform lower bound and the upper bound of $\{\sigma_t, \alpha_t\}$ and $\{\tilde{\sigma}_t, \tilde{\alpha}_t\}$ for all $\theta \in \Theta$. We denote the lower bound by (σ_L, α_L) and upper bound by (σ_U, α_U). The uniform boundedness plays a key role in the following proof.

Given (σ_t, α_t), the conditional log-likelihood function $l_t(\theta)$ of Q_t is,

$$l_t(\theta) = \log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\sigma_t} \right)^{-\alpha_t}.$$

By conditional independence, the log-likelihood function

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) = \frac{1}{n} \sum_{t=1}^{n} \log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\sigma_t} \right)^{-\alpha_t}.$$

We use $\tilde{l}_t(\theta)$ and $\tilde{L}_n(\theta)$ to denote the corresponding log-likelihood functions when $(\tilde{\sigma}_t, \tilde{\alpha}_t)$ are used.

Lemma 2.7.4 gives the result about the behavior of score function and Fisher information matrix at the true parameter θ_0 given true initial value (σ_1^0, α_1^0).

Lemma 2.7.4. *Under the conditions in Theorem 2.3.1,* $E_{\theta_0}(\frac{\partial}{\partial\theta}l_t(\theta_0)) = 0$ *and for* M_0*, the Fisher information matrix at* θ_0*, we have* $M_0 = Var_{\theta_0}(\frac{\partial}{\partial\theta}l_t(\theta_0)) = -E_{\theta_0}(\frac{\partial^2}{\partial\theta\partial\theta^T}l_t(\theta_0))$ *and* M_0 *is well defined and positive definite.*

Proof. At $\theta = \theta_0$, all the first order partial derivatives $\frac{\partial}{\partial\theta}l_t(\theta_0)$ and second order partial derivatives $\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta_0)$ can be simplified and written as functions of $\{Y_t, \sigma_t^0, \alpha_t^0\}_{t=1}^n$. The formulas for $\frac{\partial}{\partial\theta}l_t(\theta)$ and $\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta)$ are postponed to Section 2.7.4 due to their complexity.

By the fact that (σ_t^0, α_t^0) is bounded between $[\sigma_L, \sigma_U] \times [\alpha_L, \alpha_U]$, $Y_t \perp \mathcal{F}_{t-1} = \sigma(Y_s, s \leq t-1)$ and $(\sigma_t^0, \alpha_t^0) \in \mathcal{F}_{t-1}$, it is easy to prove that $E_{\theta_0}(\frac{\partial}{\partial\theta}l_t(\theta_0)) = 0$, $M_0 = \mathrm{Var}_{\theta_0}\left(\frac{\partial}{\partial\theta}l_t(\theta_0)\right) = -E_{\theta_0}(\frac{\partial^2}{\partial\theta\partial\theta^T}l_t(\theta_0))$ and M_0 is well defined, i.e. $M_0 < \infty$.

To prove that M_0 is positive definite, notice that $M_0 = \mathrm{Var}_{\theta_0}(\frac{\partial}{\partial\theta}l_t(\theta_0))$, so we only need to show that there does not exist a $c \in \mathbb{R}^9$ such that $c^T\frac{\partial}{\partial\theta}l_t(\theta_0) = 0$ *a.s.* The argument is the same as the one used in Lemma 2.7.3, where the essential idea is that $Y_t \perp \mathcal{F}_{t-1}$. □

The result of Lemma 2.7.4 is standard and expected, since θ_0 is the true parameter and we assume the data come from a correctly specified model. Lemma 2.7.5 gives moment conditions for functions of $\{Q_t\}_{t=1}^n$, which serve as building blocks for the proof of latter lemmas.

Lemma 2.7.5. *Under the conditions in Theorem 2.3.1, we have (a) for any* $\alpha > 0$, $\frac{1}{n}\sum_{t=1}^n(Q_t - \mu_0)^{-\alpha} \to_p E_{\theta_0}(Q_1 - \mu_0)^{-\alpha} < \infty$*, (b) for any positive integer* k, $\frac{1}{n}\sum_{t=1}^n[\log(Q_t - \mu_0)]^k \to_p E_{\theta_0}[\log(Q_1 - \mu_0)]^k < \infty$.

Proof. By the boundedness of scale parameter $\{\sigma_t^0\}$ and tail index $\{\alpha_t^0\}$, we have $Q_t - \mu_0 > \sigma_L \min(Y_t^{1/\alpha_L}, Y_t^{1/\alpha_U})$, so $E_{\theta_0}(Q_t - \mu_0)^{-\alpha} < \infty$ for any $\alpha > 0$ since Y_t^{-1} follows exponential distribution. The result of (a) follows from the ergodicity of AcF and the Law of Large Numbers.

For (b), we have $|\log(Q_t - \mu_0)|^k = |\log\sigma_t + 1/\alpha_t \log Y_t|^k \leq 2^k(C + 1/\alpha_L^k|\log Y_t|^k)$. It is known that $\log Y_t$ follows a Gumbel distribution thus $E_{\theta_0}(|\log Y_t|^k) < \infty$ for any positive integer k. The result of (b) follows from the ergodicity of AcF and the Law of Large Numbers. □

As mentioned above, the main technical difficulty is that the support of Q_t depends on the unknown location parameter μ_0. Lemma 2.7.6 to Lemma 2.7.14 aim to solve this difficulty by establishing uniform convergence between $\frac{1}{n}\sum_{t=1}^n h(Q_t - \mu_n)$ and $\frac{1}{n}\sum_{t=1}^n h(Q_t - \mu_0)$ for μ_n within

a neighborhood of μ_0, where $h(\cdot)$ denotes some generic function that appears in the first and second order derivatives of $\tilde{L}_n(\theta)$. The main result is stated in Lemma 2.7.14.

Lemma 2.7.6 gives an asymptotic bound on the distance between $Q_{n,1}$ and μ_0, stating that $Q_{n,1}$ converges to μ_0 at a rate that is slower than polynomial.

Lemma 2.7.6. *Under the conditions in Theorem 2.3.1,* $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$.

Proof. Notice that when $Y_t < 1$, we have $Q_t - \mu_0 = \sigma_t Y_t^{1/\alpha_t} \geq \sigma_L Y_t^{1/\alpha_L}$. Since $Y_{n,1} < 1$ *a.s.* as $n \to \infty$, it is obvious that $Q_{n,1} - \mu_0 \geq \sigma_L Y_{n,1}^{1/\alpha_L}$ *a.s.* as $n \to \infty$. The result follows from the fact that $(\log n) Y_{n,1} \to_p 1$. □

Lemma 2.7.7 gives the foundation for the uniform convergence result of first and second order derivatives of $L_n(\theta)$ given in Lemma 2.7.11, Lemma 2.7.14.

Lemma 2.7.7. *Denote* $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}$, $\alpha > 0$ *or* $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n \log(Q_{n,k}-\mu)$ *or* $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}[\log(Q_{n,k}-\mu)]^m$ *for* $m = 1, 2, 3$. *Under the conditions in Theorem 2.3.1, given positive sequence* τ_n, *s.t.* $\tau_n \sim n^{-r}, r > 0$, *the following result holds uniformly over* $|\mu_n - \mu_0| < \tau_n$,

$$|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq O_p(\tau_n).$$

Proof. We prove the result for (a) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}$, (b) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n \log(Q_{n,k}-\mu)$ and (c) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}\log(Q_{n,k}-\mu)$, a similar argument with more involved calculus can be used for the proof of others. By Lemma 2.7.6, we know that $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$, so $(Q_t - \mu_n)^{-\alpha}$ and $\log(Q_t - \mu_n)$ are asymptotically well defined for $|\mu_n - \mu_0| < \tau_n$. (a) For $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}$, assume that $\mu_n > \mu_0$, we have

$$|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n}\sum_{k=1}^n |(Q_{n,k}-\mu_n)^{-\alpha} - (Q_{n,k}-\mu_0)^{-\alpha}| \leq \frac{1}{n}\sum_{k=1}^n \frac{(\alpha+1)|\mu_n - \mu_0|}{\min\{Q_{n,k}-\mu_n, Q_{n,k}-\mu_0\}^{\alpha+1}}$$
$$\leq \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha+1}{(Q_{n,k}-\mu_n)^{\alpha+1}} = \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha+1}{(Q_{n,k}-\mu_0+\mu_0-\mu_n)^{\alpha+1}} \leq \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha+1}{(Q_{n,k}-\mu_0-\tau_n)^{\alpha+1}},$$

where the second inequality follows from the fact that $a - a^{\alpha+1} \leq (\alpha+1)(1-a)$ for all $\alpha > 0$ and $0 < a < 1$.

Since $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$, for any fixed $0 < \rho < 1$, we have $P(\rho(Q_{n,1} - \mu_0) > \tau_n) \to 1$,

so $P(\rho(Q_{n,k}-\mu_0)>\tau_n, \text{ for all } 1\le k\le n)\to 1$. With probability goes to 1, we have

$$\frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k}-\mu_0-\tau_n)^{\alpha+1}}\le\frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{[(Q_{n,k}-\mu_0)(1-\rho)]^{\alpha+1}}=O_p(\tau_n),$$

which follows from Lemma 2.7.5 (a). For $\mu_n<\mu_0$, the proof is similar but easier.

(b) For $S_n^\alpha(\mu)=n^{-1}\sum_{k=1}^{n}\log(Q_{n,k}-\mu)$, assume that $\mu_n>\mu_0$, we have

$$\begin{aligned}|S_n^\alpha(\mu_n)-S_n^\alpha(\mu_0)|&\le\frac{1}{n}\sum_{k=1}^{n}|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)|\\&=\frac{1}{n}\sum_{k=1}^{n}\log\left(1+\frac{\mu_n-\mu_0}{Q_{n,k}-\mu_n}\right)\le\frac{\tau_n}{n}\sum_{k=1}^{n}\frac{1}{Q_{n,k}-\mu_n}=O_p(\tau_n),\end{aligned}$$

where the last inequality follows from the fact that $\log(1+x)<x$ when $x>0$ and the last equality follows from the result for $S_n^\alpha(\mu)=n^{-1}\sum_{k=1}^{n}(Q_{n,k}-\mu)^{-\alpha}$. For $\mu_n<\mu_0$, the proof is similar but easier.

(c) For $S_n^\alpha(\mu)=n^{-1}\sum_{k=1}^{n}(Q_{n,k}-\mu)^{-\alpha}\log(Q_{n,k}-\mu)$, assume that $\mu_n>\mu_0$, we have

$$\begin{aligned}|S_n^\alpha(\mu_n)-S_n^\alpha(\mu_0)|&\le\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)|\\&+\frac{1}{n}\sum_{k=1}^{n}\left|(Q_{n,k}-\mu_n)^{-\alpha}-(Q_{n,k}-\mu_0)^{-\alpha}\right||\log(Q_{n,k}-\mu_0)|.\end{aligned}$$

For the first term in the sum,

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)|\\&=\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}\log\left(1+\frac{\mu_n-\mu_0}{Q_{n,k}-\mu_n}\right)\\&\le\frac{\tau_n}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-(\alpha+1)}=O_p(\tau_n),\end{aligned}$$

where the last equality follows from the result for $S_n^\alpha(\mu)=n^{-1}\sum_{k=1}^{n}(Q_{n,k}-\mu)^{-\alpha}$. For the second

term in the sum,

$$\frac{1}{n}\sum_{k=1}^{n}\left|(Q_{n,k}-\mu_n)^{-\alpha}-(Q_{n,k}-\mu_0)^{-\alpha}\right|\,|\log(Q_{n,k}-\mu_0)|$$
$$\leq \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k}-\mu_n)^{\alpha+1}}|\log(Q_{n,k}-\mu_0)| \leq \tau_n\left(\frac{1}{n}\sum_{k=1}^{n}\frac{(\alpha+1)^2}{(Q_{n,k}-\mu_n)^{2\alpha+2}}\right)^{1/2}\left(\frac{1}{n}\sum_{k=1}^{n}|\log(Q_{n,k}-\mu_0)|^2\right)^{1/2}$$
$$= O_p(\tau_n),$$

where the last inequality follows from the Cauchy-Schwartz inequality and the last equality follows from Lemma 2.7.5 and the result for $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k}-\mu)^{-\alpha}$. For $\mu_n < \mu_0$, the proof is similar but easier. □

Lemmas 2.7.8 and 2.7.9 provide the impact of parameter difference $|\theta-\theta_0|$ on $|\alpha_t-\alpha_t^0|$ and $|\sigma_t-\sigma_t^0|$ uniformly over t.

Lemma 2.7.8. *Denote* $\Phi=(\gamma_0,\gamma_1,\gamma_2,\gamma_3)$ *and* $\Phi_0=(\gamma_0^0,\gamma_1^0,\gamma_2^0,\gamma_3^0)$*, if* $\|\Phi-\Phi_0\|<\tau_n$ *and* $\tau_n\searrow 0$*, under the conditions in Theorem 2.3.1, we have*

$$(a)\ \sup_{1\leq t\leq n}\left|\alpha_t-\alpha_t^0\right|=O(\tau_n),\quad (b)\ \sup_{1\leq t\leq n}\left|\frac{\partial\alpha_t}{\partial\Phi}-\frac{\partial\alpha_t^0}{\partial\Phi}\right|=O(\tau_n),\quad (c)\ \sup_{1\leq t\leq n}\left|\frac{\partial^2\alpha_t}{\partial\Phi_i\partial\Phi_j}-\frac{\partial^2\alpha_t^0}{\partial\Phi_i\partial\Phi_j}\right|=O(\tau_n),$$

uniformly over $\|\Phi-\Phi_0\|<\tau_n$.

Proof. We only prove (a), the proof for others is similar but more involved. Using the fact that a continuously differentiable function is Lipschitz continuous on a compact set, we only need to prove that $\sup_{1\leq t\leq n}\left|\log\alpha_t-\log\alpha_t^0\right|=O(\tau_n)$. By repeatedly applying the autoregressive relation, we can get

$$\log\alpha_t=\gamma_0\sum_{k=1}^{t-1}\gamma_1^{k-1}+\gamma_2\sum_{k=1}^{t-1}\gamma_1^{k-1}\exp(-\gamma_3 Q_{t-k})+\gamma_1^{t-1}\log\alpha_1^0.$$

We have

$$|\log\alpha_t-\log\alpha_t^0|\leq\left|\gamma_0\sum_{k=1}^{t-1}\gamma_1^{k-1}-\gamma_0^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}\right|+\left|\gamma_1^{t-1}\log\alpha_1^0-(\gamma_1^0)^{t-1}\log\alpha_1^0\right|$$
$$+\left|\gamma_2\sum_{k=1}^{t-1}\gamma_1^{k-1}\exp(-\gamma_3 Q_{t-k})-\gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}\exp(-\gamma_3^0 Q_{t-k})\right|.$$

By the fact that $\sum_{k=1}^{t}\gamma_1^{k-1} < 1/(1-\gamma_1) \le 1/(1-C_b)$ and $\left|\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1} - \sum_{k=1}^{t-1}(\gamma_1)^{k-1}\right| \le \left|\frac{1}{1-\gamma_1^0} - \frac{1}{1-\gamma_1}\right| \le \frac{\tau_n}{(1-C_b)^2} = O(\tau_n)$, it is easy to see that the first two terms of the sum are $O(\tau_n)$ for any $1 \le t \le n$. For the third term, we have

$$\begin{aligned}
&\left|\gamma_2\sum_{k=1}^{t-1}\gamma_1^{k-1}\exp(-\gamma_3 Q_{t-k}) - \gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}\exp(-\gamma_3^0 Q_{t-k})\right| \\
&\le \left|\gamma_2-\gamma_2^0\right|\sum_{k=1}^{t-1}\gamma_1^{k-1}\exp(-\gamma_3 Q_{t-k}) + \gamma_2^0\sum_{k=1}^{t-1}\left|\gamma_1^{k-1}-(\gamma_1^0)^{k-1}\right|\exp(-\gamma_3 Q_{t-k}) \\
&+ \gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}\left|\exp(-\gamma_3 Q_{t-k}) - \exp(-\gamma_3^0 Q_{t-k})\right|.
\end{aligned}$$

The first two terms of the sum are $O(\tau_n)$ for any $1 \le t \le n$ by the boundedness of $\exp(-\gamma_3 Q_{t-k})$. For the third term we have,

$$\begin{aligned}
&\gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}\left|\exp(-\gamma_3 Q_{t-k}) - \exp(-\gamma_3^0 Q_{t-k})\right| \\
&= \gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1} Q_{t-k}\exp(-\gamma'_{3k} Q_{t-k})\left|\gamma_3-\gamma_3^0\right| = O(\tau_n), \text{ for any } 1 \le t \le n,
\end{aligned}$$

where $\gamma'_{3k} > 0$ is a number between γ_3 and γ_3^0 depending on Q_{t-k}, and $\gamma'_{3k} \to \gamma_3^0$ uniformly over all $k \ge 1$. By the compactness of Θ, $\gamma'_{3k} \ge C > 0$ for all $k \ge 1$. Mean value theorem is used to get the first equality and the uniform boundedness of $Q_{t-k}\exp(-\gamma'_{3k} Q_{t-k})$ is used to get the second equality. □

Lemma 2.7.9. *Denote* $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$ *and* $\Psi_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)$, *if* $\|\Psi - \Psi_0\| < \tau_n$ *and* $\tau_n \searrow 0$, *under the conditions in Theorem 2.3.1, we have*

$$\textit{(a)}\ \sup_{1\le t\le n}\left|\sigma_t - \sigma_t^0\right| = O(\tau_n),\ \ \textit{(b)}\ \sup_{1\le t\le n}\left|\frac{\partial\sigma_t}{\partial\Psi} - \frac{\partial\sigma_t^0}{\partial\Psi}\right| = O(\tau_n),\ \ \textit{(c)}\ \sup_{1\le t\le n}\left|\frac{\partial^2\sigma_t}{\partial\Psi_i\partial\Psi_j} - \frac{\partial^2\sigma_t^0}{\partial\Psi_i\partial\Psi_j}\right| = O(\tau_n),$$

uniformly over $\|\Psi - \Psi_0\| < \tau_n$.

Proof. The proof is the same as the one for Lemma 2.7.8 and thus omitted. □

Lemma 2.7.10 is used for the proof of Lemma 2.7.11.

Lemma 2.7.10. *Suppose* $\tau_n \sim n^{-r}$, $r > 0$ *and* $\sup_{1\le t\le n} |\alpha_t - \alpha_t'| = O(\tau_n)$ *where* $\{\alpha_t\}$ *and* $\{\alpha_t'\}$ *represent two different series of tail index. Under the conditions in Theorem 2.3.1, we have*

$$\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\alpha_t'}\right| = O_p(\tau_n),$$

uniformly over $|\mu_n-\mu_0|<\tau_n$. *The same result holds for* $\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\alpha_t'}\right|[\log(Q_t-\mu_n)]^k$, $k=1,2$.

Proof. We only prove the result for $\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\alpha_t'}\right|$, the proof for others is the same. Assume $\alpha_t' > \alpha_t$, the proof for the other direction is the same. By mean value theorem,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\alpha_t'}\right| \le \frac{C}{n}\sum_{t=1}^{n}(Q_t-\mu_n)^{-\alpha_t^*}\left|\log(Q_t-\mu_n)\right|\tau_n\\
&\le \frac{\tau_n C}{n}\sum_{t=1}^{n}\left((Q_t-\mu_n)^{-\alpha_L}+(Q_t-\mu_n)^{-\alpha_U}\right)\left|\log(Q_t-\mu_n)\right| = O_p(\tau_n),
\end{aligned}$$

where $\alpha_t^* \in (\alpha_t, \alpha_t')$. The last equality follows from Lemma 2.7.7. □

Lemma 2.7.11 gives the uniform convergence result of the second order derivatives of $L_n(\theta)$ over a neighborhood of θ_0, which is used in the proof of Lemma 2.7.14 (a). In the following, we denote $m_{\theta_i\theta_j}(\theta_0) = -E_{\theta_0}(\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_1(\theta_0))$.

Lemma 2.7.11. *Under the conditions in Theorem 2.3.1, for all second order derivatives of* $L_n(\theta_n)$, *we have* $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta_n) \to_p -m_{\theta_i\theta_j}(\theta_0)$, *uniformly over* $\|\theta_n-\theta_0\|<\tau_n$, *where* $\tau_n \sim n^{-r}$, $r>0$.

Proof. We only prove the case for $\frac{\partial^2}{\partial\mu^2}L_n(\theta_n)$, the proof for others is similar but more involved. By the Law of Large Numbers, we know that $\frac{\partial^2}{\partial\mu^2}L_n(\theta_0) \to_p m_{\mu\mu}(\theta_0)$, so we only need to prove that $\frac{\partial^2}{\partial\mu^2}L_n(\theta_n) - \frac{\partial^2}{\partial\mu^2}L_n(\theta_0) \to_p 0$ uniformly over the claimed region.

$$\begin{aligned}
\frac{\partial^2}{\partial\mu^2}L_n(\theta_n) - \frac{\partial^2}{\partial\mu^2}L_n(\theta_0) &= \frac{1}{n}\sum_{t=1}^{n}\left[(\alpha_t+1)(Q_t-\mu_n)^{-2}-(\alpha_t^0+1)(Q_t-\mu_0)^{-2}\right]\\
&-\frac{1}{n}\sum_{t=1}^{n}\left[\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}(Q_t-\mu_n)^{-(\alpha_t+2)}-\alpha_t^0(\alpha_t^0+1)(\sigma_t^0)^{\alpha_t^0}(Q_t-\mu_0)^{-(\alpha_t^0+2)}\right].
\end{aligned}$$

We now analyze the difference term by term. For the first term,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{t=1}^{n}\left[(\alpha_t+1)(Q_t-\mu_n)^{-2}-(\alpha_t^0+1)(Q_t-\mu_0)^{-2}\right]\right|\\&\leq\frac{1}{n}\sum_{t=1}^{n}\left|(\alpha_t+1)[(Q_t-\mu_n)^{-2}-(Q_t-\mu_0)^{-2}]\right|+\frac{C\tau_n}{n}\sum_{t=1}^{n}(Q_t-\mu_0)^{-2}=O_p(\tau_n)\to 0,\end{aligned}$$

where the inequality comes from the fact that $|\alpha_t-\alpha_t^0|=O(\tau_n)$ uniformly for all $1\leq t\leq n$ by Lemma 2.7.8 (a), and the equality comes from Lemma 2.7.7 and boundedness of $\{\alpha_t\}$. For the second term,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{t=1}^{n}\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}(Q_t-\mu_n)^{-(\alpha_t+2)}-\frac{1}{n}\sum_{t=1}^{n}\alpha_t^0(\alpha_t^0+1)(\sigma_t^0)^{\alpha_t^0}(Q_t-\mu_0)^{-(\alpha_t^0+2)}\right|\\&\leq\frac{1}{n}\sum_{t=1}^{n}\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}\left|(Q_t-\mu_n)^{-(\alpha_t+2)}-(Q_t-\mu_0)^{-(\alpha_t+2)}\right|\\&+\frac{1}{n}\sum_{t=1}^{n}\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}\left|(Q_t-\mu_0)^{-(\alpha_t+2)}-(Q_t-\mu_0)^{-(\alpha_t^0+2)}\right|\\&+\frac{1}{n}\sum_{t=1}^{n}\left|\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}-\alpha_t^0(\alpha_t^0+1)(\sigma_t^0)^{\alpha_t^0}\right|(Q_t-\mu_0)^{-(\alpha_t^0+2)}.\end{aligned}$$

By Lemma 2.7.8 (a), we know that $\sup_{1\leq t\leq n}|\alpha_t-\alpha_t^0|=O(\tau_n)$. The first two terms go to zero by Lemma 2.7.7 and Lemma 2.7.10 respectively, and the last term goes to zero by the boundedness of $\{\sigma_t,\alpha_t\}$, the differentiable continuity of $\alpha_t(\alpha_t+1)\sigma_t^{\alpha_t}$ w.r.t. σ_t,α_t and Lemma 2.7.8 (a), Lemma 2.7.9 (a). □

Note that our ultimate goal is to establish uniform convergence result about $\tilde{L}_n(\theta)$. Lemmas 2.7.12 and 2.7.13 state that the impact of arbitrary initial value $(\tilde{\sigma}_1,\tilde{\alpha}_1)$ on the behavior of $\tilde{L}_n(\theta)$ is asymptotically negligible over a neighborhood of μ_0.

Lemma 2.7.12. *Under the conditions in Theorem 2.3.1, there exists a positive constant C such that for all $\theta\in\Theta$ and $t\geq 1$,*

$$\textit{(a)}\ |\alpha_t-\tilde{\alpha}_t|\leq C\cdot C_b^{t-1},\quad \textit{(b)}\ \left|\frac{\partial\alpha_t}{\partial\Phi}-\frac{\partial\tilde{\alpha}_t}{\partial\Phi}\right|\leq C\cdot tC_b^{t-1},\quad \textit{(c)}\ \left|\frac{\partial^2\alpha_t}{\partial\Phi_i\partial\Phi_j}-\frac{\partial^2\tilde{\alpha}_t}{\partial\Phi_i\partial\Phi_j}\right|\leq C\cdot t^2C_b^{t-1},$$

$$\textit{(d)}\ |\sigma_t-\tilde{\sigma}_t|\leq C\cdot C_b^{t-1},\quad \textit{(e)}\ \left|\frac{\partial\sigma_t}{\partial\Phi}-\frac{\partial\tilde{\sigma}_t}{\partial\Phi}\right|\leq C\cdot tC_b^{t-1},\quad \textit{(f)}\ \left|\frac{\partial^2\sigma_t}{\partial\Phi_i\partial\Phi_j}-\frac{\partial^2\tilde{\sigma}_t}{\partial\Phi_i\partial\Phi_j}\right|\leq C\cdot t^2C_b^{t-1}.$$

Proof. We skip the proof since it is obvious. □

Lemma 2.7.13. *Under the conditions in Theorem 2.3.1, we have* $\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\tilde{\alpha}_t}\right| \to_p 0$*, uniformly over* $|\mu_n-\mu_0|<\tau_n$*, where* $\tau_n \sim n^{-r}$*,* $r>0$*. The same result holds for*

$$\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\tilde{\alpha}_t}\right|\left[\log(Q_t-\mu_n)\right]^k,\ k=1,2.$$

Proof. We only prove the result for $\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\tilde{\alpha}_t}\right|$, the proof for others is the same. By Lemma 2.7.12 (a), we have $|\alpha_t-\tilde{\alpha}_t| \le C\cdot C_b^{t-1}$. Assume $\tilde{\alpha}_t>\alpha_t$, the proof for the other direction is the same. By mean value theorem,

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_t}-(Q_t-\mu_n)^{-\tilde{\alpha}_t}\right| \le \frac{C}{n}\sum_{t=1}^{n}(Q_t-\mu_n)^{-\alpha_t^*}\left|\log(Q_t-\mu_n)\right| C_b^{t-1}\\ &\le \frac{C}{n}\sum_{t=1}^{n}\left((Q_t-\mu_n)^{-\alpha_L}+(Q_t-\mu_n)^{-\alpha_U}\right)\left|\log(Q_t-\mu_n)\right| C_b^{t-1} \to_p 0\end{aligned}$$

where $\alpha_t^* \in (\alpha_t,\tilde{\alpha}_t)$. The result follows from Lemma 2.7.7 and that

$$E_{\theta_0}\left[\sum_{t=1}^{\infty}\left((Q_t-\mu_0)^{-\alpha_L}+(Q_t-\mu_0)^{-\alpha_U}\right)\left|\log(Q_t-\mu_0)\right| C_b^{t-1}\right]<\infty.$$

□

Lemma 2.7.14 states the main uniform convergence result used in the proof of Theorems 2.3.1 and 2.3.2.

Lemma 2.7.14. *Under the conditions in Theorem 2.3.1, (a) for all second order derivatives of* $\tilde{L}_n(\theta)$*, we have* $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) \to_p -m_{\theta_i\theta_j}(\theta_0)$*, uniformly over* $\|\theta-\theta_0\|<\tau_n$*, where* $\tau_n\sim n^{-r}$*,* $r>0$ *(b) for the score function of* $\tilde{L}_n(\theta)$*, we have* $(\tau_n^*)^{-1}\left(\frac{\partial}{\partial\theta}\tilde{L}_n(\theta_0)-\frac{\partial}{\partial\theta}L_n(\theta_0)\right)\to_p 0$ *if* $\tau_n^* n\to\infty$*, e.g.* $\tau_n^*=1/\sqrt{n}$*.*

Proof. (a) is a direct result of Lemma 2.7.11 and the fact that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta)\to_p 0$ uniformly over $\|\theta-\theta_0\|<\tau_n$. The proof of $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta)\to_p 0$ uniformly is based on Lemmas 2.7.12 and 2.7.13. The argument is the same as that in the proof of Lemma 2.7.11, thus we skip it.

We prove (b) for $\frac{\partial}{\partial\mu}\tilde{L}_n(\theta_0)$, the proof for other first order partial derivatives is similar. Let $g(\sigma_t,\alpha_t)=\alpha_t\sigma_t^{\alpha_t}$, by the fact that $|\sigma_t-\tilde{\sigma}_t|\le C\cdot C_b^{t-1}$, $|\alpha_t-\tilde{\alpha}_t|\le C\cdot C_b^{t-1}$, we have $|g(\sigma_t,\alpha_t)-$

$g(\tilde{\sigma}_t, \tilde{\alpha}_t)| \leq C \cdot C_b^{t-1}$.

$$\frac{1}{\tau_n^*}\left(\frac{\partial}{\partial\mu}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\mu}L_n(\theta_0)\right) = \frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left(\frac{\alpha_t - \tilde{\alpha}_t}{Q_t - \mu_0} - \frac{g(\sigma_t, \alpha_t)}{(Q_t-\mu_0)^{\alpha_t+1}} + \frac{g(\tilde{\sigma}_t, \tilde{\alpha}_t)}{(Q_t-\mu_0)^{\tilde{\alpha}_t+1}}\right)$$
$$= \frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left(\frac{\alpha_t - \tilde{\alpha}_t}{Q_t - \mu_0} - \frac{g(\sigma_t, \alpha_t) - g(\tilde{\sigma}_t, \tilde{\alpha}_t)}{(Q_t-\mu_0)^{\alpha_t+1}} + g(\tilde{\sigma}_t, \tilde{\alpha}_t)[(Q_t-\mu_0)^{-(\tilde{\alpha}_t+1)} - (Q_t-\mu_0)^{-(\alpha_t+1)}]\right).$$

The first term is bounded $\frac{C}{n\tau_n^*}\sum_{t=1}^{n}\frac{C_b^{t-1}}{Q_t-\mu_0}$ and the second term by $\frac{C}{n\tau_n^*}\sum_{t=1}^{n}\frac{C_b^{t-1}}{(Q_t-\mu_0)^{\alpha_t+1}}$. Both terms go to zero in probability since $n\tau_n^* \to \infty$ and $E_{\theta_0}\left[\sum_{t=1}^{\infty} C_b^{t-1}(Q_t-\mu_0)^{-\alpha}\right] < \infty$ for all $\alpha > 0$. The same argument applies to the third term after applying mean value theorem. □

Lemma 2.7.15 gives the standard Martingale CLT.

Lemma 2.7.15. *Under the conditions in Theorem 2.3.1,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta} \Rightarrow N(0, M_0^{-1}),$$

where M is the Fisher Information matrix at θ_0.

Proof. We prove this result by using CLT for martingale difference (Billingsley (1961)). It is easy to verify that,

$$E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\middle|\mathcal{F}_{t-1}\right) = 0 \text{ and } \mathrm{Var}_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\right) = M_0 < \infty.$$

So for any $\lambda \in \mathbb{R}^9$, $\{\lambda'\frac{\partial l_t(\theta_0)}{\partial\theta}, \mathcal{F}_t\}_t$ is a square-integrable stationary martingale difference. By CLT of Billingsley (1961) and Wold-Cramér device, Lemma 2.7.15 is true. □

Proof of Theorem 2.3.1: The proof mainly uses Taylor expansion. Let $\{\tau_n\}$ be any sequence s.t. $\tau_n \sim n^{-r}$ and $n^{1/2}\tau_n \to \infty$ (i.e. $0 < r < 1/2$), let $t \in \mathbb{R}$, $y \in \mathbb{R}^8$ and define $f_n(t, y) = \tau_n^{-2}\tilde{L}_n(\mu_0 + \tau_n t, \phi^0 + \tau_n y)$, where we denote $\phi^0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$.

By Taylor Expansion we have,

$$\frac{\partial}{\partial t}f_n(t,y) = \tau_n^{-1}\frac{\partial\tilde{L}_n(\mu_0+\tau_n t, \phi^0+\tau_n y)}{\partial\mu} = \tau_n^{-1}\frac{\partial\tilde{L}_n(\mu_0,\phi^0)}{\partial\mu} + \frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu^2}t + \sum_{i=1}^{8}\frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu\partial\phi_i}y_i$$
$$= \tau_n^{-1}\left(\frac{\partial\tilde{L}_n(\mu_0,\phi^0)}{\partial\mu} - \frac{\partial L_n(\mu_0,\phi^0)}{\partial t}\right) + \tau_n^{-1}\left(\frac{\partial L_n(\mu_0,\phi^0)}{\partial\mu}\right) + \frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu^2}t + \sum_{i=1}^{8}\frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu\partial\phi_i}y_i,$$

where the second equality comes from a Taylor expansion of $\frac{\partial \tilde{L}_n(\mu_0+\tau_n t,\phi^0+\tau_n y)}{\partial t}$ at (μ_0,ϕ^0), and we have $|\mu^*-\mu_0|<\tau_n t$ and $\|\phi^*-\phi^0\|<\tau_n\|y\|$. The first term goes to 0 by Lemma 2.7.14 (b) and the second term goes to 0 by Lemma 2.7.15 and the fact that $\tau_n\sqrt{n}\to\infty$. By Lemma 2.7.14 (a), the last two terms converge uniformly over $t^2+\|y\|^2\le 1$, i.e.,

$$\frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu^2}t+\sum_i\frac{\partial^2\tilde{L}_n(\mu^*,\phi^*)}{\partial\mu\partial\phi_i}y_i\to_p -m_{\mu\mu}(\theta_0)t-\sum_{i=1}^8 m_{\mu\phi_i}(\theta_0)y_i.$$

So together we have $\frac{\partial}{\partial t}f_n(t,y)=-m_{\mu\mu}(\theta_0)t-\sum_{i=1}^8 m_{\mu\phi_i}(\theta_0)y_i+o_p(1)$. Similarly, we have $\frac{\partial}{\partial y_k}f_n(t,y)=-m_{\phi_k\mu}(\theta_0)t-\sum_{i=1}^8 m_{\phi_k\phi_i}(\theta_0)y_i+o_p(1)$, for $k=1,\ldots,8$, where $o_p(1)$'s are uniformly decaying over $t^2+\|y\|^2\le 1$. Let $t^2+\|y\|^2=1$, we have

$$t\frac{\partial f_n}{\partial t}(t,y)+\sum_i y_i\frac{\partial f_n}{\partial y_i}(t,y)=-t^2m_{\mu\mu}(\theta_0)-2t\sum_{i=1}^8 y_i m_{\mu\phi_i}(\theta_0)-\sum_{i=1}^8\sum_{j=1}^8 y_iy_j m_{\phi_i\phi_j}(\theta_0)+o_p(1)<0,$$

where the negative sign follows from the fact that the Fisher Information matrix M_0 is positive definite.

By the above argument and Lemma 5 in Smith (1985), we have that with probability going to 1, f_n has a local maximum over the open set $t^2+\|y\|^2<1$, so there exists a sequence of local maximizer $\hat{\theta}_n$ of $\tilde{L}_n(\theta)$ such that $\hat{\theta}_n\to_p\theta_0$ and $\|\hat{\theta}_n-\theta_0\|\le\tau_n$, where $\tau_n\sim n^{-r}$, $0<r<1/2$. □

Proof of Theorem 2.3.2: Theorem 2.3.1 shows the existence of $\hat{\theta}_n$ with $P(\|\hat{\theta}_n-\theta_0\|\le\tau_n)\to 1$, where $\tau_n\sim n^{-r}$, $0<r<1/2$. By Lemma 2.7.14 (a), we have that the second derivatives of $\tilde{L}_n$ are asymptotically constant in this region. The result therefore follows by standard Taylor expansion argument, Lemma 2.7.14 (b) and Lemma 2.7.15. □

Proof of Proposition 2.3.1: The arguments used in the proof of Proposition 2.3.1 are similar to the ones used in the proof of Theorems 2.3.1 and 2.3.2, thus we only give an outline of the proof since the actual argument is very tedious. In the following, we use δ to denote a generic small positive value and denote $\phi=(\beta_0,\beta_1,\beta_2,\beta_3,\gamma_0,\gamma_1,\gamma_2,\gamma_3)$. As in Proposition 2.3.1, $V_n=\{\theta\in\Theta|\mu\le cQ_{n,1}+(1-c)\mu_0\}$. Note that for any $0<c<1$, we have $\mu_0<cQ_{n,1}+(1-c)\mu_0<Q_{n,1}$ and $cQ_{n,1}+(1-c)\mu_0\searrow\mu_0$ *a.s.*

Denote $\Theta_n^\delta=\{\theta\in V_n\big|\|\theta-\theta_0\|\ge\delta\}$, $\Theta_n^\mu=\{\theta\in V_n\big|\|\theta-\theta_0\|\ge\delta,\mu>\mu_0\}$ and $\Theta^\delta=\{\theta\in$

$V_n\big|\|\theta-\theta_0\|\geq\delta,\mu\leq\mu_0\}$. Note that $\Theta_n^{\delta}=\Theta_n^{\mu}\cup\Theta^{\delta}$. We first prove that,

$$\text{(I) for any } \delta>0,\ P\left(\sup_{\Theta_n^{\delta}}\tilde{L}_n(\theta)\geq\tilde{L}_n(\theta_0)\right)\to 0,\quad \text{as } n\to\infty.$$

By the same argument in Lemmas 2.7.7 and 2.7.13, it can be proved that $\sup_{\Theta_n^{\delta}}\left|\tilde{L}_n(\theta)-L_n(\theta)\right|\to_p 0$, as $n\to\infty$. By the same argument in Lemma 2.7.7, we can further prove $\sup_{\Theta_n^{\mu}}|L_n(\mu,\phi)-L_n(\mu_0,\phi)|\to_p 0$, as $n\to\infty$. Together, we have $\sup_{\Theta_n^{\delta}}\tilde{L}_n(\theta)=\sup_{\Theta_n^{\delta}}L_n(\theta)+o_p(1)=\max\left(\sup_{\Theta^{\delta}}L_n(\theta),\sup_{\Theta_n^{\mu}}L_n(\theta)\right)+o_p(1)=\max\left(\sup_{\Theta^{\delta}}L_n(\theta),\sup_{\Theta_n^{\mu}}L_n(\mu_0,\phi)\right)+o_p(1)\leq\sup_{\Theta^{\delta/2}}L_n(\theta)+o_p(1)$. The last inequality comes from the fact that $Q_{n,1}\searrow\mu_0\ a.s.$, so with probability going to 1, we have $\{\phi|\phi\in\Theta_n^{\mu}\}\subseteq\{\phi|\phi\in\Theta^{\delta/2}\}$. It is also easy to prove that $\tilde{L}_n(\theta_0)=L_n(\theta_0)+o_p(1)\to_p E_{\theta_0}(l_1(\theta_0))$. The rest of the proof for (I) follows from the proof of Proposition 2 in Dombry (2015), which is based on the standard compactness argument.

Denote $\Theta_n^{\delta c}=\{\theta\in V_n\big|\|\theta-\theta_0\|<\delta\}$, $\Theta_n^{\mu c}=\{\theta\in V_n\big|\|\theta-\theta_0\|<\delta,\mu>\mu_0\}$ and $\Theta^{\delta c}=\{\theta\in V_n\big|\|\theta-\theta_0\|<\delta,\mu\leq\mu_0\}$. Note that $\Theta_n^{\delta c}=\Theta_n^{\mu c}\cup\Theta^{\delta c}$. We now prove that there exists a $\delta^*>0$ such that

$$\text{(II) } P\left(\text{All Hessian matrices } \frac{\partial^2}{\partial\theta\partial\theta^T}\tilde{L}_n(\theta) \text{ over } \theta\in\Theta_n^{\delta^* c} \text{ is negative definite}\right)\to 1,\quad \text{as } n\to\infty.$$

By the same argument in Lemmas 2.7.7 and 2.7.13, we can prove that

$$\sup_{\Theta_n^{\delta c}}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta)\right|\to_p 0,\ \text{as } n\to\infty,$$

and

$$\sup_{\Theta_n^{\mu c}}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\mu,\phi)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\mu_0,\phi)\right|\to_p 0,\ \text{as } n\to\infty.$$

Since $\mu\leq\mu_0$ over $\Theta^{\delta c}$, it can be proved that $\sup_{\Theta^{\delta c}}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta)\right|$ is integrable. By ergodicity of AcF and Uniform Law of Large Numbers, we have

$$\sup_{\Theta^{\delta c}}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\mu,\phi)-E_{\theta_0}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_1(\mu,\phi)\right)\right|\to_p 0,\ \text{as } n\to\infty.$$

By Lemma 2.7.4, $E_{\theta_0}\left(\frac{\partial^2}{\partial\theta\partial\theta^T}l_1(\theta_0)\right)=-M_0$ is negative definite. By the continuity of $E_{\theta_0}\left(\frac{\partial^2}{\partial\theta\partial\theta^T}l_1(\theta)\right)$

w.r.t. θ over $\Theta^{\delta c}$, we can find a $\delta^* > 0$ such that $E_{\theta_0}\left(\frac{\partial^2}{\partial\theta\partial\theta^T} l_1(\theta)\right)$ is negative definite for all $\theta \in \Theta^{\delta^* c}$. Together with the above argument, we can prove (II).

By (I), with probability going to 1, the global maximizer of $\tilde{L}_n(\theta)$ over V_n is located within $\Theta_n^{\delta^* c}$. By Theorem 2.3.1, there exists a sequence $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ such that $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$. Thus $P(\hat{\theta}_n \in \Theta_n^{\delta^* c}) \to 1$. Also, we know that $\frac{\partial \tilde{L}_n}{\partial\theta}(\hat{\theta}_n) = 0$. Together with (II) and Theorem 2.6 in Makelainen et al. (1981), we can prove Proposition 2.3.1. □

2.7.4 First and second order partial derivative of $l_t(\theta)$

In this section, we give the formula for $\frac{\partial l_t(\theta)}{\partial\theta}$ and $\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T}$. Denote $\Phi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, i.e. we use Φ as a generic symbol for $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Similarly, we set $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$.

For the first order partial derivative, we have

$$
\begin{aligned}
\frac{\partial l_t(\theta)}{\partial\mu} &= \frac{\alpha_t + 1}{Q_t - \mu} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-(\alpha_t+1)}, \\
\frac{\partial l_t(\theta)}{\partial\Phi} &= \left[\frac{1}{\alpha_t} - \log\left(\frac{Q_t - \mu}{\sigma_t}\right) + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial\alpha_t}{\partial\Phi}, \\
\frac{\partial l_t(\theta)}{\partial\Psi} &= \left[\frac{\alpha_t}{\sigma_t} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial\sigma_t}{\partial\Psi}.
\end{aligned}
$$

For the second order partial derivative, we have

$$
\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\mu^2} &= (\alpha_t + 1)(Q_t - \mu)^{-2} - \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+2)}, \\
\frac{\partial^2 l_t(\theta)}{\partial\mu\partial\Phi} &= \left[\frac{1}{Q_t - \mu} - \sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+1)} + \alpha_t\sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+1)}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial\alpha_t}{\partial\Psi}, \\
\frac{\partial^2 l_t(\theta)}{\partial\mu\partial\Psi} &= -\frac{\alpha_t^2}{\sigma_t^2}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-(\alpha_t+1)}\frac{\partial\sigma_t}{\partial\Psi}, \\
\frac{\partial^2 l_t(\theta)}{\partial\Phi\partial\Psi} &= \left[\frac{1}{\sigma_t} - \frac{1}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t} + \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial\sigma_t}{\partial\Psi}\frac{\partial\alpha_t}{\partial\Phi}, \\
\frac{\partial^2 l_t(\theta)}{\partial\Psi_i\partial\Psi_j} &= \left[\frac{\alpha_t}{\sigma_t} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial^2\sigma_t}{\partial\Psi_i\partial\Psi_j} + \left[-\frac{\alpha_t}{\sigma_t^2} - \frac{\alpha_t(\alpha_t - 1)}{\sigma_t^2}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial\sigma_t}{\partial\Psi_i}\frac{\partial\sigma_t}{\partial\Psi_j}, \\
\frac{\partial^2 l_t(\theta)}{\partial\Phi_i\partial\Phi_j} &= \left[\frac{1}{\alpha_t} - \log\left(\frac{Q_t - \mu}{\sigma_t}\right) + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial^2\alpha_t}{\partial\Phi_i\partial\Phi_j} \\
&\quad + \left[-\frac{1}{\alpha_t^2} + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\left[\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]^2\right]\frac{\partial\alpha_t}{\partial\Phi_i}\frac{\partial\alpha_t}{\partial\Phi_j}.
\end{aligned}
$$

2.7.5 Observation-driven functions $\eta_1(\cdot), \eta_2(\cdot)$ implied by GAS

Under the GAS framework described in Creal et al. (2013), we give the formulas of $\eta_1(\cdot)$ and $\eta_2(\cdot)$ implied by GAS in the dynamic GEV context. Set $\tau_t = \log \alpha_t$ and $\zeta_t = \log \sigma_t$ to ensure the positivity of parameters. Given (ζ_t, τ_t), the conditional distribution of Q_t is Fréchet$(\mu, \exp(\zeta_t), \exp(\tau_t))$. The log-likelihood function $l_t(\cdot)$ of Q_t is

$$l_t(Q_t|\mu, \zeta_t, \tau_t) = \tau_t + \exp(\tau_t)\zeta_t - (\exp(\tau_t) + 1)\log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)}.$$

To derive GAS, we need to obtain the score function of $l_t(Q_t|\mu, \zeta_t, \tau_t)$ w.r.t. (ζ_t, τ_t), which is

$$\begin{aligned}
\frac{\partial l_t}{\partial \zeta_t} &= \exp(\tau_t) - \exp(\tau_t)\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)}, \\
\frac{\partial l_t}{\partial \tau_t} &= 1 - \exp(\tau_t)\log\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right) + \exp(\tau_t)\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)}\log\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right).
\end{aligned}$$

Following the recommendation in Creal et al. (2013), we take the scaling matrix function $S_t = I$ and obtain the following GAS model,

$$\begin{aligned}
Q_t &= \mu + \sigma_t Y_t^{1/\alpha_t} \\
\log \sigma_t &= \beta_0 + \beta_1 \log \sigma_{t-1} + \beta_2 \frac{\partial l_{t-1}}{\partial \zeta_{t-1}} \\
\log \alpha_t &= \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \frac{\partial l_{t-1}}{\partial \tau_{t-1}},
\end{aligned}$$

where we assume $0 \le \beta_1, \gamma_1 < 1$ and $\beta_2, \gamma_2 > 0$. Notice that the evolution scheme given by GAS is complicated. It is easy to see that $\eta_1(Q_{t-1}) = \beta_2 \frac{\partial l_{t-1}}{\partial \zeta_{t-1}}$ is a monotone increasing function of Q_{t-1}, which is expected for volatility clustering. It is also easy to prove that $\eta_2(Q_{t-1}) = \gamma_2 \frac{\partial l_{t-1}}{\partial \tau_{t-1}}$ is an increasing function of Q_{t-1} when $\frac{Q_{t-1}-\mu}{\sigma_{t-1}} < 1$ and decreasing when $\frac{Q_{t-1}-\mu}{\sigma_{t-1}} > 1$, which is quite counter-intuitive in terms of econometric meaning. On the contrary, for AcF, a larger Q_{t-1} always give a smaller α_t.

By the fact that $\frac{\partial l_t}{\partial \zeta_t} \le \exp(\tau_t)$ and $\frac{\partial l_t}{\partial \tau_t} \le 1$, it is easy to see that (σ_t, α_t) of the GAS model is upper bounded.

Chapter 3

Semiparametric Dynamic Max-copula Model for Multivariate Time Series

3.1 Motivation

Modeling the multivariate joint behavior of random variables is one of the most fundamental tasks in statistical modeling. The construction of multivariate distributions is technically difficult and most of the early multivariate modeling is restricted within the Gaussian/elliptical family. Thanks to Sklar (1959)'s theorem, which states that multivariate dependency can be separated into a copula and individual marginal distributions, the "time of the copula" has emerged for the construction of multivariate distributions. Various copulas have been proposed in the literature, see Joe (2014) and Nelsen (1999) for a summary. Copula-based models for multivariate distributions are widely used in a variety of applications, see Frees and Valdez (1998) in actuarial science and insurance, Cherubini et al. (2004) in finance, and Genest and Favre (2007) in hydrology.

Due to its tractability, interpretability, and flexibility in modeling non-extremal joint behavior, the Gaussian copula is arguably the most widely used copula. While Gaussian copulas perform well in many areas of applications, the financial market may turn out to be an exception. One of the most significant characteristics of financial data is its tail dependence, i.e., during a crisis, asset prices tend to move together. The failure of Gaussian copulas to capture this tail dependence in the pricing of CDOs and related securities is considered one of the prominent causes of the recent

financial crisis, e.g., see Coval et al. (2009) and Salmon (2012) for more details.

The Gaussian copula's inability to model joint tail events inspires more research in the construction of copulas that can offer more sophisticated dependence structures. One direction is to exploit and extend the linear structure in Gaussian factor models by changing the distribution of latent factors from Gaussian to other distributions like *skewed-t*, e.g., Hull and White (2004), Murray et al. (2013), and Oh and Patton (2015). This framework is capable of offering more sophisticated dependence like tail dependence and tail asymmetry. Also, it is particularly attractive in high-dimensional applications thanks to the factor structure. Another promising direction is to use vine copulas, which builds high-dimensional copulas based on a sequentially iterative pairwise construction of bivariate copulas, see Aas et al. (2009), Min and Czado (2010), and Almeida et al. (2012) for more details. Vine copulas can offer flexible dependence relationships and can be represented in graphs, which helps the modeler to visually understand the dependence structure. Combining the "latent factor" idea with the pairwise construction idea of C-vine copulas, Krupskii and Joe (2013) proposed a factor copula model, where instead of imposing a Gaussian linear structure, bivariate copulas are used to specify the dependence between latent factors and observed variables. Theorem 3.7.1 in the Appendix establishes a weak equivalence between the factor copula model based on the C-vine copula and the linear factor copula model under the additive model framework.

A much different yet important direction is to construct new sophisticated copulas based on existing ones. The idea is that the new copula inherits various merits from its parents and thus offers more versatile dependence structures. In the literature, mixture of distributions is a long existing technique for generating new distributions based on existing ones. The mixture copula has a closed form *c.d.f.* and is interpretable. Inspired by Cui and Zhang (2016) and the mixture technique, in this paper we propose a novel non-linear copula construction framework named the max-copula, which generates new copulas based on existing ones via a straightforward pairwise max function. The max-copula has a closed form *c.d.f.* and has strong interpretability, especially in financial applications. Compared to mixture copulas, it can actively generate more flexible dependence structures, including asymmetric dependence structures. Moreover, due to its unique "pairwise max" characteristic, it can offer a better modeling of non-extremal behavior while attaining an accurate modeling of tail dependence. Combined with semiparametric time series models, the max-copula has shown its advantages over the mixture copula: it can accurately capture asymmetric

dependence and joint extremal movements of multivariate financial time series while simultaneously offering a better modeling of non-extremal market behavior.

The remainder of the paper is structured as follows. Section 3.2 presents the max-copula model, derives its quantile and tail dependence properties and discusses the selection of component copulas for the max-copula. Unique characteristics of max-copulas are emphasized. For high-dimensional applications, single-factor and block-factor max-copulas are developed. Section 3.3 describes the composite maximum likelihood estimation (CMLE) method under the semiparametric time series setting. Numerical experiments on the flexibility of max-copulas and the performance of CMLE are conducted in Section 3.4. In Section 3.5, we present empirical applications of max-copulas in the estimation of conditional Value at Risk (VaR) for a financial portfolio and in the construction of optimal portfolios based on 30 component stocks from the Dow Jones Industrial Average. Section 3.6 concludes with a discussion of potential extensions. All the proofs can be found in the Appendix.

3.2 Max-copula Model

3.2.1 Model specification

The idea of the max-copula is partially motivated by the mixture copula. Suppose we have two copulas of the same dimension $d \geq 2$, say $\mathbf{C}_1$ and $\mathbf{C}_2$. For any $0 \leq c \leq 1$, the linear mixture $\mathbf{C} = c\mathbf{C}_1 + (1-c)\mathbf{C}_2$ is a new copula. The stochastic representation of $\mathbf{C}$ can be built as follows: Suppose $\mathbf{U}_1 = (U_{11}, \ldots, U_{1d}) \sim \mathbf{C}_1$, $\mathbf{U}_2 = (U_{21}, \ldots, U_{2d}) \sim \mathbf{C}_2$, $X \sim Bernoulli(c)$, and $(\mathbf{U}_1, \mathbf{U}_2, X)$ are mutually independent, then $\mathbf{U} = (U_1, \ldots, U_d) = \max(\mathbf{U}_1^{\frac{1}{X}}, \mathbf{U}_2^{\frac{1}{1-X}}) \sim \mathbf{C}$, where $\max(\cdot)$ is a pairwise max function, i.e., $U_i = \max(U_{i1}^{\frac{1}{X}}, U_{i2}^{\frac{1}{1-X}})$[1]. In the following, we always assume that $(\mathbf{U}_1, \mathbf{U}_2, X)$ are mutually independent.

A closer examination reveals that the joint distribution of $\mathbf{U}$ is always a copula, as long as X is a random variable on $[0, 1]$. A general distribution on the interval [0,1] is $Beta(a, b)$. Suppose $X \sim Beta(a, b)$, we obtain the following copula model,

$$\mathbf{U} = \max(\mathbf{U}_1^{\frac{1}{X}}, \mathbf{U}_2^{\frac{1}{1-X}}), \tag{3.1}$$

[1]Here we define $1/0{=}\infty$

where $\mathbf{U}_1 \sim \mathbf{C}_1$, $\mathbf{U}_2 \sim \mathbf{C}_2$ and $X \sim Beta(a, b)$. By an elementary argument, the copula $C_{\mathbf{U}}$ of $\mathbf{U}$ is

$$\mathbf{C}_{\mathbf{U}}(u_1, \ldots, u_d) = \mathbb{E}\left(\mathbf{C}_1(u_1^X, \ldots, u_d^X) \cdot \mathbf{C}_2(u_1^{1-X}, \ldots, u_d^{1-X})\right). \tag{3.2}$$

A closed form solution of (3.2) is generally not available due to the expectation on X, however, it can be computed numerically via a one-dimensional integration. In this paper, we consider the case when X is a Dirac mass on c, where $0 < c < 1$. Under this framework, the expectation can be dropped and we have

$$\mathbf{U} = \max(\mathbf{U}_1^{\frac{1}{c}}, \mathbf{U}_2^{\frac{1}{1-c}}), \quad \text{i.e.,} \quad U_i = \max(U_{1i}^{\frac{1}{c}}, U_{2i}^{\frac{1}{1-c}}), \text{ for } i = 1, 2, \ldots, d,$$

where $\mathbf{U}_1 \sim \mathbf{C}_1$, $\mathbf{U}_2 \sim \mathbf{C}_2$ and $\mathbf{C}_1$, $\mathbf{C}_2$ are two existing copulas. Given u, $u^{1/c}$ is an increasing function of c and $u^{1/(1-c)}$ is a decreasing function of c. For each U_i, we have $P(U_i = U_{1i}^{1/c}) = P(U_{1i}^{1/c} > U_{2i}^{1/(1-c)}) = c$, so c can be viewed as the weight parameter that controls the relative strength of $\mathbf{C}_1$ and $\mathbf{C}_2$ in the max-copula $\mathbf{C}$, with a large c favoring $\mathbf{C}_1$ and small c favoring $\mathbf{C}_2$. The newly constructed $\mathbf{C}$ takes the form

$$\mathbf{C}(u_1, \ldots, u_d) = \mathbf{C}_1(u_1^c, \ldots, u_d^c) \cdot \mathbf{C}_2(u_1^{1-c}, \ldots, u_d^{1-c}), \tag{3.3}$$

and we call $\mathbf{C}$ a max-copula.

Remark 1: Both the mixture copula and the max-copula can be seen as limiting cases of the general copula (3.2) where $X \sim Beta(a, b)$. Let $X \sim Beta(\frac{c}{n}, \frac{1-c}{n})$, we have $X \to_d Bernoulli(c)$ as $n \to \infty$ and $X \to_d$ Dirac mass(c) as $n \to 0$. Therefore, the general copula (3.2) can be seen as a "bridge" between the max-copula and the mixture copula and is capable of generating more flexible dependence structures. Its properties are being investigated in a separate project.

Remark 2: We derive the max-copula (3.3) under the novel stochastic representation (3.1). We note that a more general form of (3.3) can be found in Liebscher (2008) and Durante and Sempi (2015), which can be seen as a generalization of the Khoudraji device in Khoudraji (1995). To our best knowledge, the present paper provides the first thorough study of the max-copula's probabilistic properties and establishes novel semiparametric statistical inference procedures that make its

real data application feasible.

Remark 3: An obvious generalization of max-copulas is to allow different c's for different (U_{1i}, U_{2i})'s. This extension can produce non-exchangeable max-copulas even when its components $\mathbf{C}_1$ and $\mathbf{C}_2$ are exchangeable. Such extension is not obvious in corresponding mixture copulas.

For mixture copulas, whether $\mathbf{U}$ behaves like $\mathbf{U}_1$ or $\mathbf{U}_2$ does not depend on $(\mathbf{U}_1, \mathbf{U}_2)$ itself, but on an independent Bernoulli random variable X, while for max-copulas it does depend on $(\mathbf{U}_1, \mathbf{U}_2)$. Following the pairwise max rule, for each U_i in $\mathbf{U}$, we have $U_i = \max(U_{1i}^{1/c}, U_{2i}^{1/(1-c)})$, so $\mathbf{U}$ behaves like the more extreme one between $\mathbf{U}_1^{1/c}$ and $\mathbf{U}_2^{1/(1-c)}$. This direct interaction between $\mathbf{U}_1$ and $\mathbf{U}_2$ is more realistic and has more meaningful interpretability in many areas. In financial applications, we can think of $\mathbf{U}$ as the risks for multiple stocks. Marginally speaking, for each stock U_i, there are two sources of risks coming from U_{1i} and U_{2i} with c controlling the relative weight and each stock taking the larger risk from the two sources. Jointly speaking, $\mathbf{U}_1$ and $\mathbf{U}_2$ represent the two sources of joint risks with different joint behavior, e.g., $\mathbf{U}_1$ may follow a copula $\mathbf{C}_1$ with weak multivariate dependence, representing the joint risks in a "normal state" market; while $\mathbf{U}_2$ can follow a copula $\mathbf{C}_2$ with strong joint tail dependence, representing the joint risks in a "crisis state" market. Further details about the interpretation can be found in Section 3.2.3. As is seen later in real data applications, this unique characteristic of the max-copula helps it accurately capture the joint behavior of multivariate financial time series.

3.2.2 Quantile dependence and tail dependence coefficient

Suppose the bivariate random vector $(U_1, U_2) \sim \mathbf{C}$, where $\mathbf{C}$ is the max-copula derived from $\mathbf{C}_1$ and $\mathbf{C}_2$, by (3.3). The quantile dependence λ_{max}^q between (U_1, U_2) takes the closed form

$$\lambda_{max}^q = \begin{cases} P(U_1 \le q | U_2 \le q) = \dfrac{\mathbf{C}_1(q^c, q^c) \cdot \mathbf{C}_2(q^{1-c}, q^{1-c})}{q}, & q \in (0, 0.5]; \\ P(U_1 > q | U_2 > q) = \dfrac{1 - 2q + \mathbf{C}_1(q^c, q^c) \cdot \mathbf{C}_2(q^{1-c}, q^{1-c})}{1-q}, & q \in (0.5, 1). \end{cases}$$

Sibuya (1959) introduced the concept of tail dependence coefficient as a simple criteria to quantify the joint extreme behavior of two random variables. Let (X_1, X_2) be a random vector, the quantity

$$\lambda^U = \lim_{x \to x_F} P(X_2 > x | X_1 > x), \text{ where } x_F = \sup\{x : F(x) < 1\}, \text{ [2]}$$

is called the upper tail dependence coefficient, provided the limit exists. The joint distribution of (X_1, X_2) is said to have upper tail dependence if $\lambda^U > 0$ and upper tail independence if $\lambda^U = 0$ respectively. Similarly, we can define the lower tail dependence coefficient λ^L. Theorem 3.2.1 states that the tail dependence coefficients of max-copulas have nice closed form expressions and behave differently from the ones of mixture copulas. In the following, we denote the upper and lower tail dependence coefficients of $\mathbf{C}_1$ as λ_1^U and λ_1^L, and the ones of $\mathbf{C}_2$ as λ_2^U and λ_2^L.

Theorem 3.2.1. *For the max-copula $\mathbf{C}_{max}$ based on $\mathbf{C}_1$ and $\mathbf{C}_2$ with weight c, the upper and lower tail dependence coefficients are $\lambda_{max}^U = c\lambda_1^U + (1-c)\lambda_2^U$ and $\lambda_{max}^L = \lambda_1^L \cdot \lambda_2^L$. For the mixture copula $\mathbf{C}_{mix}$ based on $\mathbf{C}_1$ and $\mathbf{C}_2$ with weight c, the upper and lower tail dependence coefficients are $\lambda_{mix}^U = c\lambda_1^U + (1-c)\lambda_2^U$ and $\lambda_{mix}^L = c\lambda_1^L + (1-c)\lambda_2^L$.*

By Theorem 3.2.1, there is a clear difference between the tail behavior of max-copulas and mixture copulas. If both component copulas $\mathbf{C}_1$ and $\mathbf{C}_2$ have symmetric tail dependence (i.e., $\lambda_1^L = \lambda_1^U, \lambda_2^L = \lambda_2^U$), the mixture copula gives symmetric tail dependence, while the max-copula can clearly offer asymmetric tail dependence, which is often found more appealing in many applications. For example, a mixture of two most widely used copulas, the Gaussian and the t copula, gives symmetric tail dependence, while the corresponding max-copula offers asymmetric tail dependence with upper tail dependence and lower tail independence. Based on Theorem 3.2.1, we can obtain the following corollary.

Corollary 3.2.1. *Assume both component copulas $\mathbf{C}_1$ and $\mathbf{C}_2$ are "diagonally" symmetric (i.e., $\mathbf{C}_i(q,q) = 2q - 1 + \mathbf{C}_i(1-q, 1-q)$, for all $q \in (0,1)$, $i = 1, 2$). By symmetry we have $\lambda_1^L = \lambda_1^U = \lambda_1$ and $\lambda_2^L = \lambda_2^U = \lambda_2$. If $\min(\lambda_1, \lambda_2) < 1$ and $\max(\lambda_1, \lambda_2) > 0$, then for the upper quantile dependence λ_{max}^q and lower quantile dependence λ_{max}^{1-q} of $\mathbf{C}_{max}$, there always exists a $q^* > 0.5$ such that $\lambda_{max}^q > \lambda_{max}^{1-q}$, for all $q \in (q^*, 1)$. On the contrary, for $\mathbf{C}_{mix}$, we have $\lambda_{mix}^q = \lambda_{mix}^{1-q}$, for all $q \in (0.5, 1)$.*

[2] In this definition, X_1 and X_2 are required to have identical marginal distribution $F(x)$.

Remark 4: In Nelsen (1999), a copula $\mathbf{C}$ is said to be "radially" symmetric if $\mathbf{C}(u,v) = \bar{\mathbf{C}}(u,v)$ for all $(u,v) \in (0,1)^2$, where $\bar{\mathbf{C}}(u,v) = u+v-1+\mathbf{C}(1-u,1-v)$ is the survival copula of $\mathbf{C}$. Diagonal symmetry in Corollary 3.2.1 is a weaker condition than radial symmetry and all copulas in the elliptical copula family are radially symmetric.

Corollary 3.2.1 further demonstrates the fundamental difference between the max-copula and the mixture copula. The max-copula is an "asymmetric system" that is capable of actively generating asymmetric dependence from symmetric component copulas, while the mixture copula is a "symmetric system" that can only inherit asymmetric dependence from an asymmetric component copula.

An extreme example is when $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$. For all copulas $\mathbf{C}$ we have $\mathbf{C}_{mix} = \mathbf{C}$, while for almost all widely used copulas $\mathbf{C}$ (except extreme value copulas which are max-stable) we have $\mathbf{C}_{max} \neq \mathbf{C}$. Under the condition of Corollary 3.2.1, the upper quantile dependence of $\mathbf{C}_{max}$ is stronger than its corresponding lower quantile dependence, which is a desirable property, especially in financial risk management.

3.2.3 Choice of copulas and unique characteristics of the max-copula

Since the max-copula is based on component copulas $\mathbf{C}_1$ and $\mathbf{C}_2$, to construct a max-copula, we need to specify $\mathbf{C}_1$ and $\mathbf{C}_2$, which is a model selection problem. For bivariate max-copulas, a rule of thumb can be developed based on Theorem 3.2.1: if we want the max-copula to have upper tail dependence, at least one of the copulas need to be upper tail dependent, while if we need lower tail dependence, both of the copulas are required to have lower tail dependence. Users can choose different copulas to suit different applications. In the elliptical copula family, the Gaussian copula has no upper and lower tail dependence; the t copula has both upper and lower tail dependence. Meanwhile, in the Archimedean copula family, the Gumbel copula has upper tail dependence and the Clayton copula has lower tail dependence.

In general, the selection of max-copulas is a hard topic, as is the selection of mixture copulas. A standard procedure in the literature is to use a likelihood based information criteria (e.g., AIC, BIC) to select $\mathbf{C}_1$ and $\mathbf{C}_2$ from a pool of candidate copulas. There are various extensions of AIC/BIC for composite likelihood (which is later used for the estimation of max-copulas), e.g., Gao and Song (2010) and Varin et al. (2011). Another promising direction is to follow the work in Cai and Wang

(2014), where the authors proposed a penalized likelihood procedure via shrinkage operators for the selection of mixture copulas, which selects appropriate copulas and estimates related parameters simultaneously.

The primary purpose of this paper is to model multivariate financial time series, e.g., negative daily returns of multiple stocks. One of the most significant characteristics of the stock market is that stock returns tend to have greater dependence during a crisis and tend to behave more "independently" otherwise. In other words, negative stock returns have asymmetric tail dependence with strong upper tail dependence and weak to none lower tail dependence, see Oh and Patton (2015) for an example. Thus, in general, we want to design a max-copula that has asymmetric tail behavior with strong upper tail dependence.

As mentioned above, the Gaussian copula has no tail dependence and thus lacks the ability to capture the "joint crash" property of the stock market. However, it performs well in capturing moderate scale stock returns, i.e., the Gaussian copula can be used to model the "normal state" stock market. On the other hand, the Gumbel copula is capable of modeling joint upper tail dependence, while it may not perform well under moderate scale since it is an extreme-value copula, i.e., the Gumbel copula can be used to model the "crisis state" stock market.

Based on the above observations, to better capture the multivariate dependence structure of the stock market, in this paper we choose $\mathbf{C}_1$ to be a Gaussian copula with correlation matrix Σ and $\mathbf{C}_2$ to be a Gumbel copula with parameter α. By Theorem 3.2.1, the constructed max-copula $\mathbf{C}$ has a lower tail dependence coefficient of zero and an upper tail dependence coefficient of $(1-c)(2-2^{1/\alpha})$. Since the Gaussian copula is a special case of the elliptical copula and the Gumbel copula is a special case of the Archimedean copula, the generalization of the max-copula based on the elliptical family and the Archimedean family easily follows.

If $c = 1$, the max-copula $\mathbf{C}$ degenerates to a Gaussian copula $\mathbf{C}_1$ and there is no upper or lower-tail dependence between the U_i's. If $c = 0$, the max-copula $\mathbf{C}$ degenerates to a Gumbel copula $\mathbf{C}_2$. When $0 < c < 1$, the Gaussian copula $\mathbf{C}_1$ helps regulate the U_i's dependence structure under moderate scale and the Gumbel copula $\mathbf{C}_2$ helps achieve upper tail dependence between the U_i's. To help better understand the unique dependence structure of the max-copula, Figure 3.1 shows scatter plots of simulated data $\mathbf{U} = (U_1, U_2)$ from three sets of parameters (c, α, ρ) for a bivariate max-copula $\mathbf{C}(u_1, u_2)$. For a better visual illustration, the data are transformed to be marginally

normally distributed using $\Phi^{-1}(U_i)$, where Φ is the *c.d.f.* of standard normal distribution. In the graph, circle points correspond to data where the Gumbel copula dominates (i.e., both points of $\mathbf{U}$ come from the Gumbel copula), triangle points correspond to data where the Gaussian copula dominates (i.e., both points of $\mathbf{U}$ come from the Gaussian copula) and cross points correspond to the "mixed" case where one point of $\mathbf{U}$ comes from the Gumbel copula and one from the Gaussian copula. The three plots demonstrate the case where the Gaussian copula has positive correlation ($\rho = 0.5$) and the Gumbel copula has moderate upper tail dependence ($\alpha = 2$). From left to right, c decreases from 0.8 to 0.2, i.e., the influence of the Gaussian copula is declining and that of the Gumbel copula is increasing. As is expected, the Gumbel copula dominantly regulates the upper tail area, while the Gaussian copula mainly affects the distribution in the non-extremal area.

The unique characteristic of the max-copula is well demonstrated in Figure 3.1. Notice here in the upper tail region, for all three c's, $\mathbf{U}$ always come from the Gumbel copula due to its upper tail dependence, while in the non-extremal region, the graph is a mix of three shapes with triangle or cross points taking the dominance, depending on the magnitude of c. Since in comparison to the Gaussian copula, the Gumbel copula is not suitable for modeling non-extremal behavior, the existence of the "mixed" cross points helps achieve a better modeling of the non-extremal area by "decreasing" the influence of the Gumbel copula. This mechanism takes effect especially in the case of multivariate max-copulas (i.e., $d \geq 3$), since multiple points of $\mathbf{U} = (U_1, \ldots, U_d)$ can come from the Gaussian copula, e.g., $(U_1, \ldots U_p)$ come from the Gaussian copula and $(U_{p+1}, \ldots U_d)$ come from the Gumbel copula. This is a unique feature of the max-copula that is not shared by the mixture copula. With the unique feature, the max-copula achieves better modeling flexibility for the non-extremal behavior while attaining good modeling for tail dependence. This feature is further demonstrated through simulation experiments in Section 3.4.1 and real data applications in Section 3.5.

3.2.4 Factor structured max-copula

The number of parameters of an unstructured Gaussian copula's correlation matrix Σ increases quadratically as the dimension increases, which imposes a huge challenge on the estimation and inference of the model. To bypass this obstacle, we propose two factor structured max-copulas,

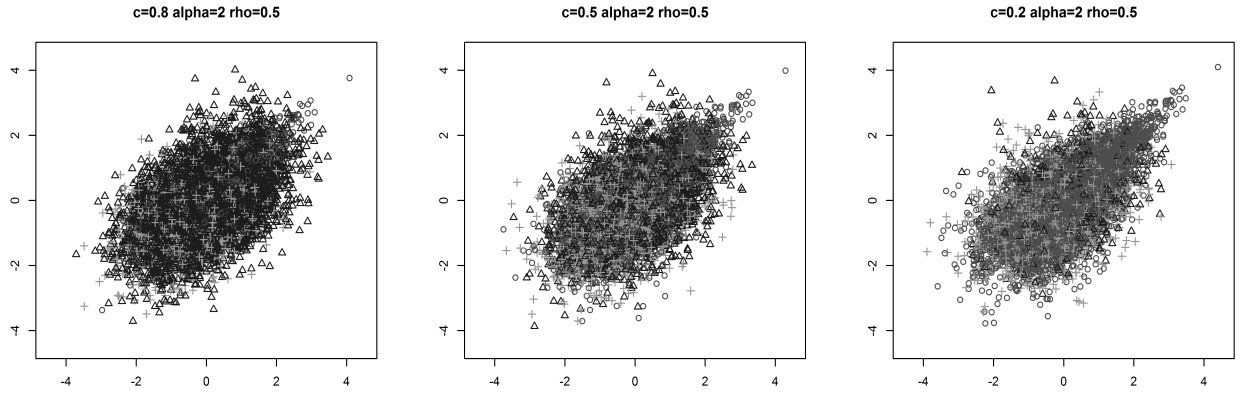

Figure 3.1: *Scatterplots of simulated two-dimensional max-copula under three sets of parameters* (c, α, ρ)*. The marginals are transformed to Gaussian scale for better illustration.*

which offer flexible multivariate dependence modeling while remaining numerically tractable.

3.2.4.1 Single-factor max-copula

For low-dimensional applications, we design a parsimonious max-copula by imposing a factor structure on the correlation matrix Σ of the Gaussian copula. For the single-factor max-copula, Σ takes a single-factor structure, where $\rho_{ij} = \beta_i \cdot \beta_j$ for all $1 \leq i, j \leq d$ and $|\beta_i| < 1$. For identification purpose, we assume $\beta_1 > 0$. The Gumbel copula belongs to the one-parameter Archimedean copula family, which offers parsimonious modeling at the cost of assuming a rather restrictive exchangeable dependence structure among variables. Such assumption is acceptable when the dimension is low, so we keep it unchanged in the single-factor max-copula. Thus, the parameters associated with the d-dimensional single-factor max-copula are $(c, \beta_1, \ldots, \beta_d, \alpha)$, which is of length $d+2$ and much smaller than $O(d^2)$.

3.2.4.2 Block-factor max-copula

In practice, especially in high-dimensional applications, situations where the multivariate observations come from several groups with similar characteristics are not uncommon. For example, in financial applications, stocks may come from different industrial sectors and it is expected that stocks from the same sector have common behavior and are more closely related. In order to have a better modeling of such data, we propose the block-factor max-copula. Assume $\mathbf{U}$ consists of p groups and for each group $i = 1, 2, \ldots, p$, it contains d_i group members. By a slight abuse of

notation, we denote $\mathbf{U} = \bigcup_{i=1}^{p}(U_{i1}, U_{i2}, \ldots, U_{id_i})$, where $\mathbf{U}$ is of dimension $d = \sum_{i=1}^{p} d_i$.

For the Gaussian copula, we take advantage of the natural group structure by imposing a block-factor structure on the correlation matrix Σ. Specifically, we assume the Gaussian copula is implied by a multivariate normal distribution, denoted by $\mathbf{Z} = \bigcup_{i=1}^{p}(Z_{i1}, Z_{i2}, \ldots, Z_{id_i})$, which has the stochastic representation:

$$Z_{ij} = \beta_i \cdot F_0 + \gamma_i \cdot F_i + \varepsilon_{ij},$$

where $i = 1, \ldots, p$, $j = 1, \ldots, d_i$, F_0 is the common factor across different groups, F_i's are group-specific factors, and ε_{ij}'s are subject-level noise. Also, all random variables are mutually independent and standard normal. Here F_0 introduces correlations across groups while F_i's are responsible for group-specific correlations. The block-factor structure requires $2p$ parameters instead of $O(d^2)$, which provides a much more parsimonious model.

We impose the group structure on the Gumbel copula by using the theory of hierarchical Archimedean copula (HAC). Intuitively, HAC can be thought as a block-factor Archimedean copula. Here we extend the Gumbel copula to the one-level hierarchical Gumbel copula, which offers different within group dependence and common between group dependence. More formally, $\mathbf{U} = \bigcup_{i=1}^{p}(U_{i1}, U_{i2}, \ldots, U_{id_i})$ is said to follow a one-level hierarchical Gumbel copula if its *c.d.f.* can be written as

$$C(\mathbf{u}) = \psi_0\left(\sum_{i=1}^{p} \psi_0^{-1}[C_i(u_{i1}, \ldots, u_{id_i})]\right),$$

where $\psi_0 = \exp(-x^{1/\alpha_0})$ is the Gumbel copula generator with parameter α_0, $C_i(\cdot)$'s are Gumbel copulas with $C_i(u_{i1}, \ldots, u_{id_i}) = \psi_i\left(\sum_{j=1}^{d_i} \psi_i^{-1}(u_{ij})\right)$, and $\psi_i = \exp(-x^{1/\alpha_i})$'s are the Gumbel copula generators with parameter α_i's. For $C(\mathbf{u})$ to be a valid copula, we require $\alpha_0 \leq \min_{1 \leq i \leq p} \alpha_i$, i.e., the between group dependence is weaker than the within group dependence. Under the one-level hierarchical Gumbel copula, the within group dependence parameter is α_i for the ith group and the between group dependence parameter is α_0 for all different groups. For more details of HAC, we refer readers to Joe (2014).

Based on the block-factor Gaussian copula $\mathbf{C}_1$ and the hierarchical Gumbel copula $\mathbf{C}_2$, we now

specify the block-factor max-copula. Assume that $\mathbf{U}_1 = \bigcup_{i=1}^{p}(U_{i1,1}, U_{i2,1}, \ldots, U_{id_i,1})$ follows a block-factor Gaussian copula $\mathbf{C}_1$ and $\mathbf{U}_2 = \bigcup_{i=1}^{p}(U_{i1,2}, U_{i2,2}, \ldots, U_{id_i,2})$ follows a one-level hierarchical Gumbel copula $\mathbf{C}_2$. We call the copula of

$$\mathbf{U} = \max(\mathbf{U}_1^{1/c}, \mathbf{U}_2^{1/(1-c)})$$

a block-factor max-copula. The parameters associated with the d-dimensional multivariate copula are $(c, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_p, \alpha_0, \ldots, \alpha_p)$, which is of length $3p+2$, where p is the number of groups of $\mathbf{U}$. For variables within the same group i, i.e., $(U_{i1}, U_{i2}, \ldots, U_{id_i})$, the block-factor max-copula reduces to a max-copula based on a Gaussian copula with exchangeable correlation $\rho = \frac{\beta_i^2+\gamma_i^2}{1+\beta_i^2+\gamma_i^2}$ and a Gumbel copula with parameter α_i. For two variables from different groups i and j, e.g., (U_{i1}, U_{j1}), the block-factor max-copula reduces to a bivariate max-copula based on a Gaussian copula with correlation $\rho = \frac{\beta_i \cdot \beta_j}{\sqrt{1+\beta_i^2+\gamma_i^2}\sqrt{1+\beta_j^2+\gamma_j^2}}$ and a Gumbel copula with parameter α_0.

For parsimony, we use the same weight c for all p groups. Additional flexibility can be obtained by imposing group-specific weights. Here we use the factor Gaussian copula and the hierarchical Gumbel copula as component copulas for the block-factor max-copula, the generalization to the factor elliptical copula and the hierarchical Archimedean copula follows readily.

3.2.5 Semiparametric dynamic max-copula model

The ultimate purpose of the max-copula is to model the joint behavior of multivariate time series. In reality, it is almost impossible to observe an *i.i.d.* sequence of multivariate time series. Furthermore, the marginals by no means can behave like uniform random variables on [0,1]. To tackle these two problems, we follow the procedure in Chen and Fan (2006a) and propose a Semiparametric Dynamic Max-copula (hereafter SDM) model. Let $\{\mathbf{Y}_t\}_{t=1}^{T}$ be a multivariate time series where $\mathbf{Y}_t$ is of dimension d and let $\mathcal{F}_{t-1}$ denote the information set at time $t-1$, i.e., the sigma field generated by $\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots\}$. We specify the SDM model as follows:

I. *Parametric dynamic component*: We assume that the dynamics of multivariate time series happen in first and second order conditional moments and we further assume that the conditional mean and variance can be correctly parametrized up to a finite-dimensional unknown parameter

λ_0, i.e.,

$$\mathbf{Y}_t = \mu_t(\lambda_0) + \sqrt{H_t(\lambda_0)}\boldsymbol{\eta}_t,$$

where $\mu_t(\lambda_0) = (\mu_{t,1}(\lambda_0), \ldots, \mu_{t,d}(\lambda_0))' = \mathbb{E}(\mathbf{Y}_t|\mathcal{F}_{t-1})$ and $H_t(\lambda_0) = diag(h_{t,1}(\lambda_0), \ldots, h_{t,d}(\lambda_0))$, in which $h_{t,j}(\lambda_0) = \mathbb{E}[(Y_{t,j} - \mu_{t,j}(\lambda_0))^2|\mathcal{F}_{t-1}],\ j = 1, \ldots, d$.

II. *Nonparametric marginal component*: We do not impose any parametric assumption on the marginals of $\boldsymbol{\eta}_t = (\eta_{t1}, \ldots, \eta_{td})'$. Instead, we only assume that each marginal of η_{tj} is continuous and denote the true marginal distribution function as F_j^0.

III. *Parametric max-copula component*: According to Sklar (1959), there is a unique copula $\mathbf{C}^0$ such that $F^0(\boldsymbol{\eta}) = \mathbf{C}^0(F_1^0(\eta_1), \ldots, F_d^0(\eta_d))$, where F^0 is the true joint distribution of $\boldsymbol{\eta}_t$. In an SDM model, we assume $\mathbf{C}^0$ is a max-copula based on component copulas $\mathbf{C}_1$, $\mathbf{C}_2$ and denote θ_0 as true parameters for the max-copula $\mathbf{C}^0$.

In the following, $\mathbf{C}^0$ is taken to be a single-factor max-copula or a block-factor max-copula depending on different applications, and we call the corresponding SDM model a single-factor SDM model or a block-factor SDM model respectively.

3.3 Semiparametric Composite Maximum Likelihood Estimation

In order to fully estimate the SDM model, we need to estimate all three components of it. We largely follow the semiparametric copula estimation framework in Genest et al. (1995) and Chen and Fan (2006a). The multivariate time series we observe are $\{\mathbf{Y}_t\}_{t=1}^T$ and the ultimate goal is to estimate the parameter θ_0 of the max-copula $\mathbf{C}^0$.

3.3.1 Estimation of the dynamic and marginal components

The parametric dynamic component and the nonparametric marginal component are treated as nuisance parameters and the estimation procedure is standard. After the parametric assumption on the dynamic component is fixed, standard MLE can be employed to estimate λ^0 using the observations $\{\mathbf{Y}_t\}_{t=1}^T$. Based on estimated $\hat{\lambda}^0$, we can construct the fitted errors $\{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T$. An empirical distribution function $\hat{F}_{Tj}$ is employed to estimate the nonparametric F_j^0, for $j = 1, \ldots, d$,

where $\hat{F}_{Tj}(\cdot) = \frac{1}{T+1}\sum_{t=1}^{T} I(\hat{\eta}_{tj} \leq \cdot)$.

3.3.2 Estimation of the max-copula component

Using estimated $\hat{F}_T = (\hat{F}_{T1}, \ldots, \hat{F}_{Td})$, we can turn the fitted errors $\{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T$ into the transformed errors $\{\hat{\mathbf{U}}_t = \hat{F}_T(\hat{\boldsymbol{\eta}}_t)\}_{t=1}^T$, which can be seen as an approximation for the unobserved true copula processes $\{\mathbf{U}_t\}_{t=1}^T$ that drive $\{\mathbf{Y}_t\}_{t=1}^T$.

In Genest et al. (1995) and Chen and Fan (2006a), the authors used standard MLE on $\{\hat{\mathbf{U}}_t\}_{t=1}^T$ to estimate the copula parameter θ_0. In our SDM model, due to the nature of max-copulas, the full log-likelihood function is analytically complicated and may also introduce numerical difficulty during optimization, especially when the dimension of the multivariate time series is high. To bypass the problem, we employ the composite maximum likelihood estimation method (CMLE). To our best knowledge, this is the first time CMLE is applied to the semiparametric dynamic modeling of time series.

Consider a d-dimensional random vector $\mathbf{U} = (U_1, \ldots, U_d)$, with *p.d.f.* $f(\mathbf{u}; \theta)$ for some unknown parameter vector θ. Denote by $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ a set of marginal events with associated likelihoods $L_k(\mathbf{u}; \theta) \propto f(\mathbf{u} \in \mathcal{A}_k; \theta)$. Following Lindsay (1988), a composite log-likelihood is the weighted sum $CL(\mathbf{u}; \theta) = \sum_{k=1}^{K} w_k \log L_k(\mathbf{u}; \theta)$, where w_k's are nonnegative weights to be chosen. For more details of CMLE, see Varin et al. (2011).

In this paper, we set $\mathcal{A}$ to be all pairwise combinations between elements of $\mathbf{U} = (U_1, \ldots, U_d)$, so $K = \frac{d(d-1)}{2}$, $\mathcal{A}_k = \{i, j\}$ for some $1 \leq i < j \leq d$ and $L_k(\mathbf{u}; \theta) = f_{ij}(u_i, u_j; \theta)$, where $f_{ij}(\cdot)$ denotes the pairwise *p.d.f.* between U_i and U_j. Also, we set $w_k = 1$ for all $k = 1, \ldots, K$. As a result, the composite log-likelihood function based on one observation $\mathbf{u} = (u_1, \ldots, u_d)$ takes the form

$$CL(\mathbf{u}; \theta) = \sum_{k=1}^{K} \log L_k(\mathbf{u}; \theta) = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \log f_{ij}(u_i, u_j; \theta).$$

3.3.2.1 Pairwise likelihood function

Suppose $\mathbf{U} = (U_1, \ldots, U_d)$ follows the max-copula $\mathbf{C}(\mathbf{u}; \theta)$. To employ CMLE, we first need to derive the pairwise copula density for each pair of (U_i, U_j). In the following, we set $i = 1$ and $j = 2$, while others can be derived similarly. To be generic, in the following, we assume the correlation

parameter of the Gaussian copula to be ρ and the parameter of the Gumbel copula to be α. Later, ρ and α can be replaced by the parameters in the single-factor or block-factor max-copula. By construction, the pairwise *c.d.f.* of (U_1, U_2) can be written as

$$\begin{aligned} C_{12}(u_1, u_2) &= C_{Gaussian}(u_1^c, u_2^c, \rho) \cdot C_{Gumbel}(u_1^{1-c}, u_2^{1-c}, \alpha) \\ &= \Phi(\Phi^{-1}(u_1^c), \Phi^{-1}(u_2^c), \rho) \cdot \exp\left[-\left(\sum_{i=1}^{2} (-\log u_i^{1-c})^\alpha \right)^{1/\alpha} \right] \\ &= \Phi(\Phi^{-1}(u_1^c), \Phi^{-1}(u_2^c), \rho) \cdot \exp\left[-(1-c)\left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right]. \end{aligned}$$

The pairwise *p.d.f.* can be obtained by taking the partial derivatives w.r.t. u_1 and u_2. The detailed expression can be found in Section 3.7.2 of the Appendix.

3.3.2.2 Composite likelihood method

Suppose we have a d-dimensional single-factor max-copula $\mathbf{C}$ with true parameter $\theta_0 = (c_0, \beta_1^0, \beta_2^0, \ldots, \beta_d^0, \alpha_0)$. Following Section 3.3.2.1, the composite log-likelihood function based on the transformed errors $\{\hat{\mathbf{U}}_t\}_{t=1}^T$ can be written as

$$\mathrm{CL}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} l_{t,ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta) = \frac{1}{T} \sum_{t=1}^{T} l_t(\hat{\mathbf{U}}_t; \theta),$$

where $l_{t,ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta) = \log f_{ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta) = \log f_{ij}(\hat{U}_{ti}, \hat{U}_{tj}; c, \beta_i, \beta_j, \alpha)$ is the pairwise log-likelihood function between U_i and U_j based on the tth transformed error $\hat{\mathbf{U}}_t = (\hat{U}_{t1}, \ldots, \hat{U}_{td})$ and $l_t(\hat{\mathbf{U}}_t; \theta) = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} l_{t,ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta)$. Denote $\hat{\theta}_T$ as the maximizer of the composite likelihood function $\mathrm{CL}_T(\theta)$ and call it the CMLE of θ.

For the block-factor max-copula, we can further separate the composite likelihood function into the within group likelihood and between group likelihood. Suppose there are p groups, each with group size d_i, and denote the tth transformed error as $\hat{\mathbf{U}}_t = \bigcup_{i=1}^{p} (\hat{U}_{t,i1}, \ldots, \hat{U}_{t,id_i})$. The composite log-likelihood function can be written as

$$\mathrm{CL}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} [\mathrm{CLWG}_t(\theta) + \mathrm{CLBG}_t(\theta)],$$

where $\mathrm{CLWG}_t(\theta)$ stands for the within group composite likelihood and $\mathrm{CLBG}_t(\theta)$ stands for the between group composite likelihood. Following the above formulation, we have

$$\mathrm{CLWG}_t(\theta) = \sum_{k=1}^{p} \sum_{i=1}^{d_k-1} \sum_{j=i+1}^{d_k} l_{t,(ki,kj)}(\hat{U}_{t,ki}, \hat{U}_{t,kj}; \theta),$$
$$\mathrm{CLBG}_t(\theta) = \sum_{k=1}^{p-1} \sum_{g=k+1}^{p} \sum_{i=1}^{d_k} \sum_{j=1}^{d_g} l_{t,(ki,gj)}(\hat{U}_{t,ki}, \hat{U}_{t,gj}; \theta),$$

where $l_{t,(ki,kj)}(\hat{U}_{t,ki}, \hat{U}_{t,kj}; \theta) = \log f_{ki,kj}(\hat{U}_{t,ki}, \hat{U}_{t,kj}; \theta) = \log f_{ki,kj}(\hat{U}_{t,ki}, \hat{U}_{t,kj}; c, \beta_k, \gamma_k, \alpha_k)$ is the pairwise log-likelihood function between U_{ki} and U_{kj} that both come from kth group, and $l_{t,(ki,gj)}(\hat{U}_{t,ki}, \hat{U}_{t,gj}; \theta) = \log f_{ki,gj}(\hat{U}_{t,ki}, \hat{U}_{t,gj}; \theta) = \log f_{ki,gj}(\hat{U}_{t,ki}, \hat{U}_{t,gj}; c, \beta_k, \gamma_k, \beta_g, \gamma_g, \alpha_0)$ is the pairwise log-likelihood function between U_{ki} and U_{gj} that come from kth group and gth group respectively.

If we re-index $\hat{\mathbf{U}}_t$ as $\hat{\mathbf{U}}_t = (\hat{U}_{t1}, \hat{U}_{t2}, \ldots, \hat{U}_{td})$ with $d = \sum_{i=1}^{p} d_i$, it is easy to verify that we have

$$\mathrm{CL}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}[\mathrm{CLWG}_t(\theta) + \mathrm{CLBG}_t(\theta)] = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{d-1}\sum_{j=i+1}^{d} l_{t,ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\hat{\mathbf{U}}_t; \theta),$$

where $l_t(\hat{\mathbf{U}}_t; \theta) = \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} l_{t,ij}(\hat{U}_{ti}, \hat{U}_{tj}; \theta)$. By re-indexing, we unify the notation of $CL_T(\theta)$ for the single-factor and block-factor max-copulas.

As mentioned before, we set $w_k = 1$ for all pairwise likelihoods. It is known that all weights lead to consistent estimation while some weights are more efficient. For the block-factor max-copula, since pairwise likelihoods come from various sources (i.e., within group and between group), it is natural to set different weights for likelihoods from different sources, which may help improve the estimation efficiency. However, since efficiency is not the main focus here and the simulation studies show satisfactory performance of the current weights, we leave the investigation of finding efficient weights as a future research question.

3.3.3 Asymptotic theory

In Genest et al. (1995) and Chen and Fan (2006a,b), the authors showed that under the semi-parametric setting, the standard MLE based on the transformed errors $\{\hat{\mathbf{U}}_t\}_{t=1}^{T}$ is consistent and asymptotically normal. It is also well known that with the classical *i.i.d.* setting, under some regularity conditions, both MLE and CMLE are consistent and asymptotically normal, while CMLE is

less efficient in terms of asymptotic covariance. Based on the above two observations, it is not surprising that CMLE is also consistent and asymptotically normal under the SDM model, as stated in Theorems 3.3.1 and 3.3.2.

Theorem 3.3.1. *Suppose the observations* $\{\mathbf{Y}_t\}_{t=1}^T$ *and the composite likelihood function* $l_t(\cdot)$ *satisfy Assumption C and D in Chen and Fan (2006a), we have* $\hat{\theta}_T \to_p \theta_0$ *as* $T \to \infty$*, i.e.,* $\hat{\theta}_T$ *is consistent.*

Before we state Theorem 3.3.2, we first introduce some notations. Denote $l(u_1, \ldots, u_d; \theta)$ $= \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} l_{ij}(u_i, u_j; \theta) = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \log f_{ij}(u_i, u_j; \theta)$, $l_\theta(u_1, \ldots, u_d; \theta) = \partial l(u_1, \ldots, u_d; \theta)/\partial\theta$, $l_j(u_1, \ldots, u_d; \theta) = \partial l(u_1, \ldots, u_d; \theta)/\partial u_j$, $l_{\theta\theta}(u_1, \ldots, u_d; \theta) = \partial^2 l(u_1, \ldots, u_d; \theta)/\partial\theta\partial\theta'$, and $l_{\theta j}(u_1, \ldots, u_d; \theta)$ $= \partial^2 l(u_1, \ldots, u_d; \theta)/\partial u_j \partial\theta$, for $j = 1, \ldots, d$. Also, denote $\{\mathbf{U}_t = (U_{t1}, \ldots, U_{td})\}_{t=1}^T$ as the unobserved true copula processes that drive the observations $\{\mathbf{Y}_t\}_{t=1}^T$. Further denote

$$A_T^0 = \frac{1}{T} \sum_{t=1}^{T} \{l_\theta(U_{t1}, U_{t2}, \ldots, U_{td}; \theta_0) + \sum_{j=1}^{d} Q(U_{tj}; \theta_0)\},$$

where $Q(U_{tj}; \theta) \equiv E_{\theta_0}\left[l_{\theta j}(\mathbf{U}_s; \theta_0)(I\{U_{tj} \le U_{sj}\} - U_{sj})\middle| U_{tj}\right], s \neq t$. Denote $B = -\mathbb{E}_{\theta_0}[l_{\theta\theta}(\mathbf{U}_t; \theta_0)]$ and $\Sigma = \mathrm{Var}_{\theta_0}[l_\theta(\mathbf{U}_t; \theta_0) + \sum_{j=1}^{d} Q(U_{tj}; \theta_0)]$.

Theorem 3.3.2. *Let* $\theta_0 \in int(\Theta)$*. Under Assumption D and N in Chen and Fan (2006a), we have (1)* $\hat{\theta}_T - \theta_0 = B^{-1}A_T^0 + o_p(T^{-1/2})$*, (2)* $\sqrt{T}(\hat{\theta}_T - \theta_0) \to_d N(0, B^{-1}\Sigma B^{-1})$*, i.e.,* $\hat{\theta}_T$ *is asymptotically normal.*

As mentioned in Chen and Fan (2006a), the additional term $Q(U_{tj}; \theta_0)$ in A_T^0 is introduced by the estimation of the marginal distribution functions $F_j^0(\cdot), j = 1, \ldots, d$, and if $F_j^0(\cdot)$ is completely known, $Q(U_{tj}; \theta_0)$ will disappear. We note that CMLE implicitly imposes an incorrect working independence assumption, which may make it less efficient in terms of asymptotic covariance. However, the benefits of CMLE are that it bypasses the analytical difficulty of the full likelihood function and lowers the computational complexity and instability of the estimation procedures, especially in high dimension.

The classical plug-in estimator for the asymptotic covariance matrix is available. However, due to analytical complexity, a parametric bootstrap procedure is proposed to estimate the asymptotic

variance for CMLE, which is consistent and gives good finite sample performance, as is shown in Section 3.7.3.2 of the Appendix.

3.4 Simulation Study

3.4.1 Comparison between the max-copula and the mixture copula

In this section, we demonstrate the advantage of max-copulas over mixture copulas in the modeling of non-extremal behavior. Specifically, we use a bivariate Gaussian copula as a "surrogate" for the non-extremal joint behavior of random vector (U_1, U_2). We compare the Kullback-Leibler (KL) distance[3] of the max-copula and the mixture copula to the Gaussian copula, while keeping the upper tail dependence coefficients of the max-copula and the mixture copula the same. We conduct experiments for two scenarios where the component copulas are either Gaussian + Gumbel copula or Gaussian + t copula. We demonstrate the simulation procedure for Gaussian + Gumbel copula in detail and report the results for both scenarios. Denote $\mathbf{C}_1(\rho)$ as a Gaussian copula and denote $\mathbf{C}_2(\alpha)$ as a Gumbel copula. Denote the max-copula based on (c, $\mathbf{C}_1(\rho)$, $\mathbf{C}_2(\alpha)$) as $\mathbf{C}_{max}(c, \rho, \alpha)$ and the corresponding mixture copula as $\mathbf{C}_{mix}(c, \rho, \alpha)$. By Theorem 3.2.1, $\mathbf{C}_{max}(c, \rho, \alpha)$ and $\mathbf{C}_{mix}(c, \rho, \alpha)$ have the same upper tail dependence coefficient. We conduct experiments for two different cases of settings.

In Case 1, we calculate the KL distance of the max-copula $\mathbf{C}_{max}(c, \rho, \alpha)$ and the mixture copula $\mathbf{C}_{mix}(c, \rho, \alpha)$ to its component Gaussian copula $\mathbf{C}_1(\rho)$. We set $\rho = 0.5$ and $\alpha = 2$ since they are the typical parameters obtained in real data applications in Section 3.5, and we change c from 0.1 to 0.9. The result is summarized in Figure 3.2 (A). As can be seen, the max-copula is always closer to the component Gaussian copula $\mathbf{C}_1(\rho)$ in terms of KL distance. It confirms the observation in Figure 3.1 that the unique pairwise max rule of the max-copula helps decrease the influence of the Gumbel copula in the non-extremal region and thus helps offer a better modeling of the non-extremal joint behavior.

In Case 2, we first set a target bivariate Gaussian copula $\mathbf{C}_{Gau}(\rho_0)$. We then seek the max-copula $\mathbf{C}_{max}(c, \rho, \alpha)$ and the mixture copula $\mathbf{C}_{mix}(c, \rho, \alpha)$ that minimize their KL distance to the target $\mathbf{C}_{Gau}(\rho_0)$ and also attain a given upper tail dependence coefficient λ^U. This is a constrained

[3] All the KL distance calculated in this section are based on Monte Carlo integration.

optimization problem where the parameters are (c, ρ, α) and the constraint is $(1-c)(2-2^{1/\alpha}) = \lambda^U$. We fix $\lambda^U = 0.24$, the same as the estimated one in Section 3.5.1. We change the target Gaussian copula from $\rho_0 = 0.1$ to $\rho_0 = 0.9$ and report the minimized KL distance attained by the max-copula and the mixture copula in Figure 3.2 (B). The max-copula attains a better ability to approximate the Gaussian copula in the range $0 < \rho_0 < 0.6$, which is the most commonly encountered range of Gaussian copulas in real data applications in Section 3.5. It indicates that for Gaussian copulas with low to medium level correlation, the max-copula has a stronger approximation ability than the mixture copula, while attaining the same level of tail dependence.

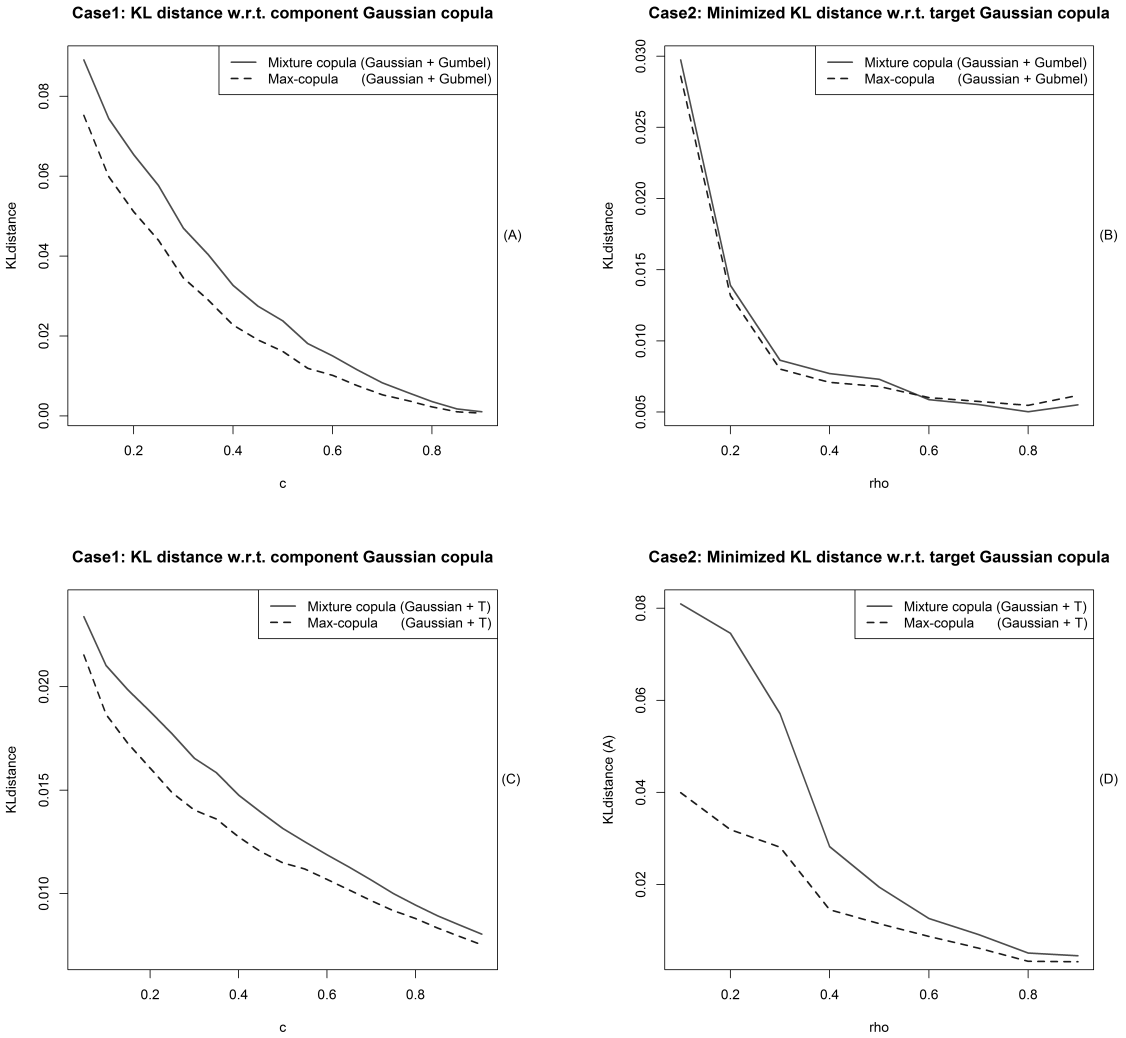

Figure 3.2: *KL distances of the max-copula and the mixture copula based on either Gaussian copula + Gumbel copula or Gaussian copula + t-copula to the pre-specified Gaussian copulas.*

We have also conducted experiments where $\mathbf{C}_1(\rho_1)$ is a Gaussian copula and $\mathbf{C}_2(\rho_2, \nu)$ is a

t-copula. We fix $\nu = 4$ for the t-copula. For Case 1, we set $\rho_1 = \rho_2 = 0.5$ and change c from 0.1 to 0.9. For Case 2, we use the same target Gaussian copula $\mathbf{C}_{Gau}(\rho_0)$ and the same λ_U. Similar results have been observed in Figure 3.2 (C) and (D), which confirm the advantage of the max-copula in achieving good non-extremal behavior modeling while attaining desired tail dependence.

3.4.2 Performance of CMLE

In this section, we examine the performance of CMLE under single-factor and block-factor SDM models. The data generating process is as follows:

$$
\begin{aligned}
&Y_{ti} = \phi_0 + \phi_1 Y_{t-1,i} + \sigma_{ti}\eta_{ti},\ t = 1, 2, \ldots, T \text{ and } i = 1, \ldots, d,\\
&\sigma_{ti}^2 = \omega + \beta\sigma_{t-1,i}^2 + \alpha\sigma_{t-1,i}^2\eta_{t-1,i}^2,\\
&\boldsymbol{\eta}_t \equiv [\eta_{t1}, \ldots, \eta_{td}]' \overset{i.i.d.}{\sim} \mathbf{F}_\eta = \mathbf{C}(\Phi, \Phi, \ldots, \Phi),
\end{aligned}
$$

where Φ is the standard normal distribution function and $\mathbf{C}$ is the max-copula. Here we assume an AR(1)-GARCH(1,1) structure for the dynamic component of $\mathbf{Y}_t$. We set the parameters to be $[\phi_0, \phi_1, \omega, \beta, \alpha] = [0.01, 0.05, 0.05, 0.85, 0.10]$, which according to Oh and Patton (2013), broadly match the values of estimation from real world financial data.

3.4.2.1 Single-factor max-copula result

We conduct experiments on single-factor SDM models with $d = 4$, where we set the true parameters $(c, \alpha, \beta_1, \beta_2, \beta_3, \beta_4) = (0.5, 2, 0.2, 0.4, 0.6, 0.8)$. Under this setting, the Gaussian copula has correlations of 0.08, 0.12, 0.16, 0.24, 0.32, 0.48, and the Gumbel copula has an upper tail dependence coefficient of 0.59.

We simulate 500 data sets of sample size $T = (1000, 2000, 5000)$ and report the sample mean and sample standard deviation based on the 500 estimators. The result is summarized in Table 3.1. As can be seen, under the single-factor SDM model, the CMLE is consistent, where both bias and variance of the estimators get smaller as sample size T gets bigger.

	c	α	β_1	β_2	β_3	β_4
T=1000	0.503 (0.080)	2.010 (0.148)	0.195 (0.105)	0.407 (0.072)	0.596 (0.075)	0.822 (0.087)
T=2000	0.494 (0.057)	1.992 (0.105)	0.193 (0.067)	0.403 (0.049)	0.601 (0.053)	0.813 (0.076)
T=5000	0.497 (0.037)	1.991 (0.075)	0.197 (0.043)	0.402 (0.033)	0.602 (0.031)	0.807 (0.040)

Table 3.1: *Semiparametric CMLE for data generated by the single-factor SDM model with* $d = 4$ *and* $(c, \alpha, \beta_1, \beta_2, \beta_3, \beta_4) = (0.5, 2, 0.2, 0.4, 0.6, 0.8)$. *The standard deviations of estimators are in brackets.*

3.4.2.2 Block-factor max-copula result

In this section, we conduct numerical investigations on block-factor SDM models. We assume $d = 20$ and $p = 4$, i.e., the observations are of dimension 20 with five subjects in each of the four groups. We set $c = 0.5, (\beta_1, \ldots, \beta_4) = (1, 1, 1.2, 1.2)$, and $(\gamma_1, \ldots, \gamma_4) = (0.8, 0.8, 1, 1)$ for the block-factor Gaussian copula, and set $(\alpha_0, \alpha_1, \ldots, \alpha_4) = (1.5, 1.75, 1.75, 2, 2)$ for the one-level hierarchical Gumbel copula. Under the current setting, the Gaussian copula has within group correlations of 0.62, 0.71 and between group correlations of $0.38, 0.40, 0.42$; the Gumbel copula has within group upper tail dependence coefficients of 0.51, 0.59 and a common between group upper tail dependence coefficient of 0.41.

We simulate 500 data sets of sample size $T = (1000, 2000, 5000)$ and report the sample mean and sample standard deviation based on the 500 estimators. The result is summarized in Table 3.2. As can be seen, the CMLE is consistent where both bias and variance of the estimators get smaller as sample size T gets bigger.

We have also conducted a simulation for a much larger block-factor SDM model with $d = 108$ and $p = 9$, i.e., the observations are of dimension 108 with twelve subjects in each of the nine groups. Though there are 29 parameters in total, the CMLE still performs well and provides decent accuracy when $T = 500$. Due to limited space, the result is provided in Section 3.7.3.1 of the Appendix.

3.5 Real Data Application

In this section, we give two real data applications of the max-copula. The first one is about the single-factor SDM model in the estimation of conditional VaR for a financial portfolio, and the second one is about the block-factor SDM model in the construction of optimal portfolios based

	c	β_1	β_2	β_3	β_4
$T = 1000$	0.501 (0.092)	1.020 (0.110)	1.031 (0.100)	1.231 (0.122)	1.236 (0.125)
$T = 2000$	0.492 (0.064)	1.018 (0.078)	1.008 (0.070)	1.216 (0.101)	1.214 (0.073)
$T = 5000$	0.499 (0.039)	1.007 (0.042)	1.010 (0.048)	1.203 (0.051)	1.206 (0.050)
	γ_1	γ_2	γ_3	γ_4	α_0
$T = 1000$	0.840 (0.133)	0.805 (0.159)	1.000 (0.158)	1.027 (0.144)	1.493 (0.084)
$T = 2000$	0.821 (0.084)	0.827(0.084)	1.002 (0.088)	1.006 (0.086)	1.500 (0.041)
$T = 5000$	0.802 (0.056)	0.810 (0.054)	1.003 (0.056)	0.998 (0.045)	1.501 (0.028)
	α_1	α_2	α_3	α_4	
$T = 1000$	1.742 (0.142)	1.745 (0.147)	2.011 (0.127)	1.992 (0.137)	
$T = 2000$	1.736 (0.065)	1.737 (0.078)	2.000 (0.071)	1.998 (0.075)	
$T = 5000$	1.746 (0.046)	1.741 (0.057)	1.999 (0.053)	1.996 (0.047)	

Table 3.2: *Semiparametric CMLE for data generated by the block-factor SDM model with* $d = 20$ *and* $p = 4$. *The standard deviations of estimators are in brackets.*

on 30 component stocks from the Dow Jones Industrial Average. For comparison purposes, in each section, we fit the data with four different copulas (the max-copula, its component Gaussian copula, its component Gumbel copula, and the corresponding mixture copula) and compare their performances.

3.5.1 Value at Risk estimation for financial portfolios

Monitoring negative returns of a portfolio is essential in financial risk management. The most common practice is to use the Value at Risk (VaR), which is a certain extreme quantile (e.g., 0.95, 0.99, 0.995) of the portfolio's negative return. Since a portfolio usually contains multiple constituents, in order to obtain an accurate estimation of its VaR, it is essential to have a good modeling of the joint behavior of the portfolio components. For a given portfolio with d-constituents $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{td})$ and weight $\mathbf{w}_t = (w_{t1}, \ldots, w_{td})$, the conditional daily VaR for day $t+1$ can be defined as certain conditional quantile of $\sum_{i=1}^{d} w_{t+1,i} Y_{t+1,i}$ given the current information $\mathcal{F}_t$. For simplicity, here we only consider the case where the weight $\mathbf{w}_t$ is constant. Following McNeil and Frey (2000), we propose a semiparametric copula framework for estimating conditional VaR of a given portfolio based on the SDM model.

We consider the analysis of negative daily stock returns of Citigroup, General Electric and Pfizer, between September 1, 1995 and August 31, 2012, which consists of 4295 observations. The parametric dynamic component of each time series $\{Y_{ti}\}_{t=1}^{T}, i = 1, 2, 3$, is set to be AR(1)+GARCH(1,1).

For each pair (i, j) of the three stocks, we calculate the sample quantile dependence $\hat{\lambda}_{ij}^q$ based on the transformed errors $\{\hat{\mathbf{U}}_t\}_{t=1}^{4295}$ and plot the average sample dependence $(\hat{\lambda}_{12}^q + \hat{\lambda}_{13}^q + \hat{\lambda}_{23}^q)/3$ in Figure 3.3. A nonparametric bootstrap on the transformed errors $\{\hat{\mathbf{U}}_t\}_{t=1}^{4295}$ with $B = 1000$ replications is used to construct pointwise 95% C.I. for the sample quantile dependence estimates. As is clearly shown, the sample quantile dependence diminishes to 0 as q approaches 0 with 0 in the 95% C.I. and stays positive as q comes near 1 with 0 outside of the 95% C.I., which implies an asymmetric tail dependence structure.

For this low-dimensional ($d = 3$) application, we set the copula component of the SDM model to be the single-factor max-copula. We first fit the SDM model using the whole dataset and summarize the result in Table 3.3. As can be seen, the Gumbel copula plays a substantial role with $1-\hat{c} = 0.42$. The estimated Gaussian copula has correlations of 0.27, 0.33 and 0.55, while the estimated Gumbel copula has an upper tail dependence coefficient of 0.58. Together with $\hat{c} = 0.58$, it implies that the three stocks have a strong upper tail dependence with $\hat{\lambda}^U = 0.24$.

c	α	β_1	β_2	β_3
0.58 (0.07)	1.98 (0.37)	0.67 (0.07)	0.82 (0.06)	0.41 (0.12)

Table 3.3: *Semiparametric CMLE for negative daily stock returns of Citigroup (β_1), General Electric (β_2), and Pfizer (β_3). The bootstrapped standard deviations of estimators are in brackets.*

To test the performance of the single-factor SDM model in estimating conditional VaR, we use the back testing technique in McNeil and Frey (2000). Specifically, on each day t, we use a window of size 1000 (i.e., days $(t - 999, \ldots, t)$) to estimate the SDM model. Based on the fitted model, we generate $B = 1000$ bootstrap samples $\{\mathbf{Y}_{t+1}^b = (Y_{t+1,1}^b, Y_{t+1,2}^b, Y_{t+1,3}^b)\}_{b=1}^B$ for day $t+1$ and estimate the conditional VaR using the B bootstrapped portfolio returns $\{\sum_{i=1}^3 w_i Y_{t+1,i}^b\}_{b=1}^B$. A violation happens when the actual daily loss is over the q-th quantile of the bootstrap sample, where q varies among 0.95, 0.99 and 0.995. Along with the single-factor max-copula, we also use the Guassian copula, the Gubmel copula, and the corresponding mixture copula for model comparison.

In total, we have a test sample of size $4295 - 1000 = 3295$ days. The back testing is conducted for seven different weights $\mathbf{w}$. The expected and actual number of violations, as well as p-values of the binomial tests for three selected weights are reported in Table 3.4 due to limited space. More details can be found in Section 3.7.4.1 of the Appendix. As can be seen, the estimated VaR

performs well, indicating that the single-factor SDM model captures the joint dynamics of the three stocks accurately. Due to limited space, the results for the other three copulas are not presented. In summary, the Gaussian copula tends to underestimate the VaR and achieves 0.949, 0.989, 0.994 for the mean VaR level across the seven portfolios; the Gumbel copula tends to overestimate the VaR and achieves 0.953, 0.991, 0.996 for mean VaR level; the mixture copula achieves a better result with mean VaR level 0.951, 0.991 and 0.996. The numbers achieved by the max-copula are 0.952, 0.990 and 0.995, which are arguably the best among all copulas, particularly at the extreme quantiles 0.99 and 0.995.

Portfolio	Expected VaR Level	Actual VaR Level	Expected Violation	Actual Violation	p-value
$w_1 = .33$	0.950	0.952	164.75	158	0.63
$w_2 = .33$	0.990	0.991	32.95	28	0.43
$w_3 = .33$	0.995	0.995	16.48	17	0.81
$w_1 = .2$	0.950	0.952	164.75	158	0.63
$w_2 = .3$	0.990	0.992	32.95	27	0.34
$w_3 = .5$	0.995	0.997	16.48	10	0.14
$w_1 = .2$	0.950	0.951	164.75	161	0.81
$w_2 = .5$	0.990	0.992	32.95	27	0.34
$w_3 = .3$	0.995	0.995	16.48	18	0.71

Table 3.4: *Performance of conditional VaR based on the single-factor SDM model, with VaR level 0.95, 0.99, and 0.995.*

The reason why the max-copula outperforms the others is illustrated more clearly in Figure 3.3, which plots the sample quantile dependence function along with those implied by the four fitted copulas. As can be seen, the Gumbel copula overestimates the upper tail dependence and also performs poorly in the non-extremal region. While the Gaussian copula has a good fit for the non-extremal region, it suffers from an underestimation of the upper tail dependence. Both the mixture copula and the max-copula perform decently with the max-copula having a slightly better fit in the upper quantile and tail area.

We further conduct a formal Goodness of Fit test on all four copulas based on the Cramér-von Mises type test in Genest et al. (2009). The estimated p-value is 0.036 for the Gaussian copula, 0.00 for the Gumbel copula, 0.066 for the mixture copula, and 0.072 for the max-copula, confirming the advantage of the max-copula.

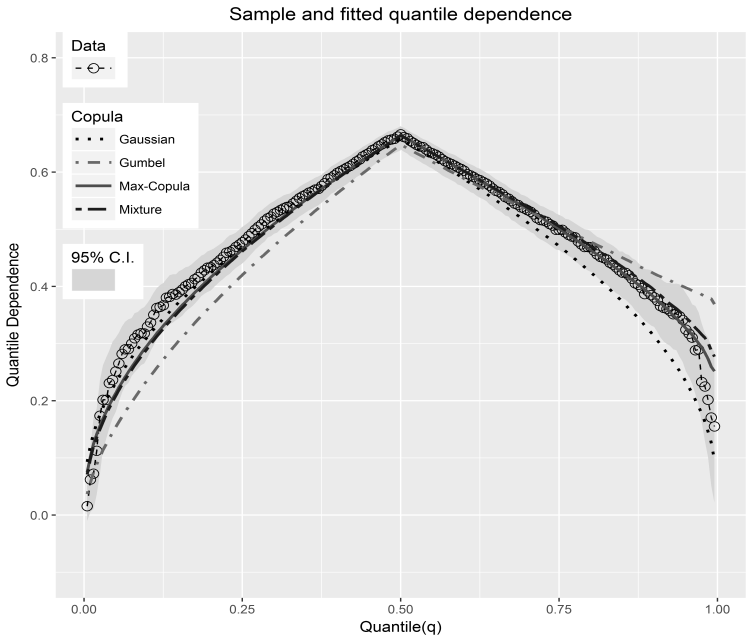

Figure 3.3: *This figure plots the average of the sample quantile dependence functions of the three stocks and a bootstrap 95% (pointwise) confidence interval for it. The quantile dependence functions based on estimated copulas (max-copula, Gaussian copula, Gumbel copula, and mixture copula) are also plotted.*

3.5.2 Optimal portfolio construction on Dow Jones Industrial Average

In this section, we extend the application of the SDM model from financial risk management to optimal portfolio construction as described in Harris and Mazibas (2013). The task here is to establish a weekly[4] optimal portfolio construction framework for the 30 component stocks of Dow Jones Industrial Average Index (hereafter DJI30). The DJI30 is a major U.S. stock index which consists of 30 large publicly owned companies based in the U.S. The 30 companies come from six industrial sectors, according to the Standard Industrial Classification (SIC) system. Detailed group information can be found in Section 3.7.4.2 of the Appendix. It is known that stocks from the same industrial sector have common behavior and closer relationships. Because of this natural group structure among the 30 stocks, the block-factor SDM model is used for this high-dimensional ($d = 30$) application. To capture the leverage effect in the conditional volatility of stock returns, we set the parametric dynamic component to be AR(1)+GJR-GARCH(1,1). More details about GJR-

[4]Compared to daily-frequency portfolio optimization, weekly-frequency portfolio optimization (i.e., we update the portfolio weight every trading week) is more realistic due to the transaction cost in real trading. Here a week refers to a trading week, which typically contains the five workdays in a week.

GARCH(1,1) can be found in Glosten et al. (1993).

The initiative of this application is to demonstrate the block-factor SDM model's ability to capture both normal and extreme joint movements of multiple stocks, which can be utilized for the construction of more risk-rewarding portfolios. To serve this end, we choose the observation period from January 02, 2004 to December 30, 2011, which consists of $T = 2015$ days and roughly 403 trading weeks. It covers the so-called Financial crisis period from 2008 to 2009 and the post-crisis market rally from 2009 to 2011, which are later used as test samples to examine the performance of the proposed portfolio optimization framework under different market scenarios. Since VISA is not listed until 2008 and Verizon is the only stock in SIC 4, we remove these two stocks from the candidate pool for portfolio construction and consider the remaining 28 stocks from DJI30. In summary, the 28 stocks can be classified into five groups with group sizes of 8, 9, 3, 5 and 3 respectively.

We now describe the weekly optimal portfolio construction framework. Roughly speaking, in the proposed framework, we manage a portfolio that consists of the 28 component stocks in DJI30. At the end of each trading week, say day t, based on the estimated block-factor SDM model, we seek to forecast an optimal portfolio weight $\hat{\mathbf{w}}_t = \{\hat{w}_i^t\}_{i=1}^{28}$ for the next trading week of days $(t+1, t+2, \ldots, t+5)$ such that a certain risk measure (e.g., Variance, VaR, Expected Shortfall, etc.) of the constructed portfolio with the optimal weight $\hat{\mathbf{w}}_t$ is minimal among all possible portfolios. More formally, denote the portfolio weight as $\mathbf{w} = \{w_i\}_{i=1}^{28}$, we are solving the following constrained optimization problem:

$$\min_{\mathbf{w}} \Phi_t(\mathbf{w}), \text{ where } \mathbf{w} \geq 0, \mathbf{w}'\mathbf{1} = 1.$$

where $\Phi_t(\mathbf{w})$ denotes the selected risk measure of the portfolio with weight $\mathbf{w}$, which can be calculated based on the estimated block-factor SDM model on day t. The minimizer $\hat{\mathbf{w}}_t$ is the optimal portfolio weight.

In the following, we set $\Phi_t(\mathbf{w})$ to be the Expected Shortfall (ES) of the cumulative portfolio returns in the next trading week of days $(t+1, \ldots, t+5)$. The reason we choose ES instead of VaR is that ES is a convex function of portfolio weight $\mathbf{w}$, which makes the optimization an easier task. For a random variable Z, its $(1-\alpha)$ level ES is defined to be $ES_{1-\alpha} =$

$-\frac{1}{\alpha}\left(E(Z1_{\{Z\leq z_\alpha\}})+z_\alpha(\alpha-P(Z\leq z_\alpha))\right)$, where z_α denotes the α-quantile of Z. For more information on ES, readers are referred to McNeil et al. (2005). Conditioned on the estimated SDM model on day t, $\Phi_t(\mathbf{w})$ can be well approximated by parametric bootstrap. Using the estimated SDM model, we generate $B = 10000$ times bootstrap samples $\{\{\mathbf{Y}^b_{t+k} = (Y^b_{t+k,1},\ldots,Y^b_{t+k,28})\}^5_{k=1}\}^B_{b=1}$ for the returns of the 28 stocks on days $(t+1,\ldots,t+5)$ and approximate the ES $\Phi_t(\mathbf{w})$ by its empirical version based on the B bootstrap portfolio returns $\{\sum^5_{k=1}\sum^{28}_{i=1} w_i Y^b_{t+k,i}\}^B_{b=1}$. For the level of ES, we use $1-\alpha = 0.99$. The reason why we set $1-\alpha$ close to 1 is that we only want to avoid the extreme loss of the portfolio instead of all the risks. The intuition is that if we try to eliminate all the risks, the constructed portfolio will also lack the ability to capture upward movements of the stock market.

As in Section 3.5.1, we use a "back testing" procedure to examine the performance of the weekly optimal portfolio constructed by the block-factor SDM model. The basic steps are as follows. We initially estimate the block-factor SDM model using the first $s = 1000$ observations. Based on the estimated model, we forecast the one-week ahead out-of-sample optimal portfolio weight $\hat{\mathbf{w}}_s = \{\hat{w}^s_i\}^{28}_{i=1}$ for the next trading week of days $(s+1, s+2,\ldots,s+5)$ using the optimal portfolio construction framework described above. The performance of the constructed optimal portfolio is $\hat{r}_s = \sum^{s+5}_{t=s+1}\sum^{28}_{i=1}\hat{w}^s_i r_{i,t}$, where $r_{i,t}$ is the actual log-return of the ith stock on day t. The estimation window is then rolled forward one trading week (5 days) and the optimal portfolio weight for the next week is generated. The last iteration uses days $\{T-1004,\ldots,T-5\}$ to generate optimal portfolio weight for the trading week of days $(T-4, T-3,\ldots,T)$. The starting date of the out-of-sample test is December 21, 2007 and the ending date is December 30, 2011, consisting of 203 weeks.

Besides the block-factor max-copula, we also conduct experiments using the block-factor Gaussian copula, the hierarchical Gumbel copula, and the corresponding block-factor mixture copula. The summary statistics of the portfolio returns in the out-of-sample test set for DJI30 and optimal portfolios constructed by each of the four copulas can be found in Table 3.5. As can be seen, compared to DJI30, all the copula-constructed portfolios deliver considerably higher annualized returns and lower risks, in terms of standard deviation, VaR, and ES of the portfolio. Moreover, the portfolio constructed by the max-copula achieves the highest return and offers the best overall Sharpe ratio among all portfolios, followed by the mixture copula. We note that the Sharpe ratio

is the most commonly used measure for risk-adjusted returns in the financial industry. For a given portfolio, its Sharpe ratio is defined to be $\frac{r-r_f}{\sigma}$, where r is its expected return, σ is its standard deviation, and r_f is the risk free rate. [5]

To better demonstrate the comparison, we plot the cumulative portfolio values of the four constructed portfolios and DJI30 throughout the test set in Figure 3.4. As is clearly shown, the portfolio value constructed by the max-copula almost always stays on top. The vertical dotted line marks the date (Feb 26, 2009) when DJI30 reaches its minimum level, and we call it the minimal point. The market scenario before the minimal point can be seen as in "crisis" state and the one after the minimal point can be seen as in post-crisis "normal/rally" state. To better assess different copulas' performance under different market scenarios, separate analyses have been performed on portfolio returns before and after the minimal point. We summarize the Sharpe ratio in Table 3.5 and more detailed information is in Section 3.7.4.2 of the Appendix. As can be seen, the Gumbel copula does not perform well during the "normal/rally" state, the Gaussian copula does not perform well during the "crisis state", while the max-copula offers the best balance of performance under both market scenarios (best in "normal/rally" state and second best in "crisis state").

	TR	AR	SD	VaR95	VaR99	ES95	ES99	SR(Overall)	SR(Before)	SR(After)
Max-copula	19.86	5.16	2.11	3.34	5.45	5.32	9.03	0.047	-0.169	0.165
Mixture	16.06	4.18	2.12	3.47	5.58	5.47	9.47	0.038	-0.179	0.155
Gaussian	12.87	3.35	2.19	3.19	5.90	5.63	9.58	0.029	-0.191	0.144
Gumbel	11.25	2.92	2.31	3.84	5.78	5.70	9.16	0.024	-0.143	0.113
DJI30	1.23	0.32	3.11	4.63	9.86	7.89	13.62	0.0019	-0.254	0.142

Table 3.5: *Summary statistics of the returns of different portfolios from December 21, 2007 to December 30, 2011. "TR" stands for total return, "AR" stands for annualized return and "SR" stands for Sharpe ratio. All numbers, except SR, are in percentage.*

3.6 Conclusion and Future Research

In this paper, we have proposed and studied the max-copula, which is a novel nonlinear asymmetric copula generating framework that constructs new flexible copulas based on existing ones through a pairwise max function. The constructed max-copula enjoys tractable theoretical properties, such as closed form quantile and tail dependence functions. Moreover, it is capable of modeling

[5]Here, we set r_f to be the Federal funds rate, which is 0.00-0.25% during the test period. For simplicity we set $r_f = 0$.

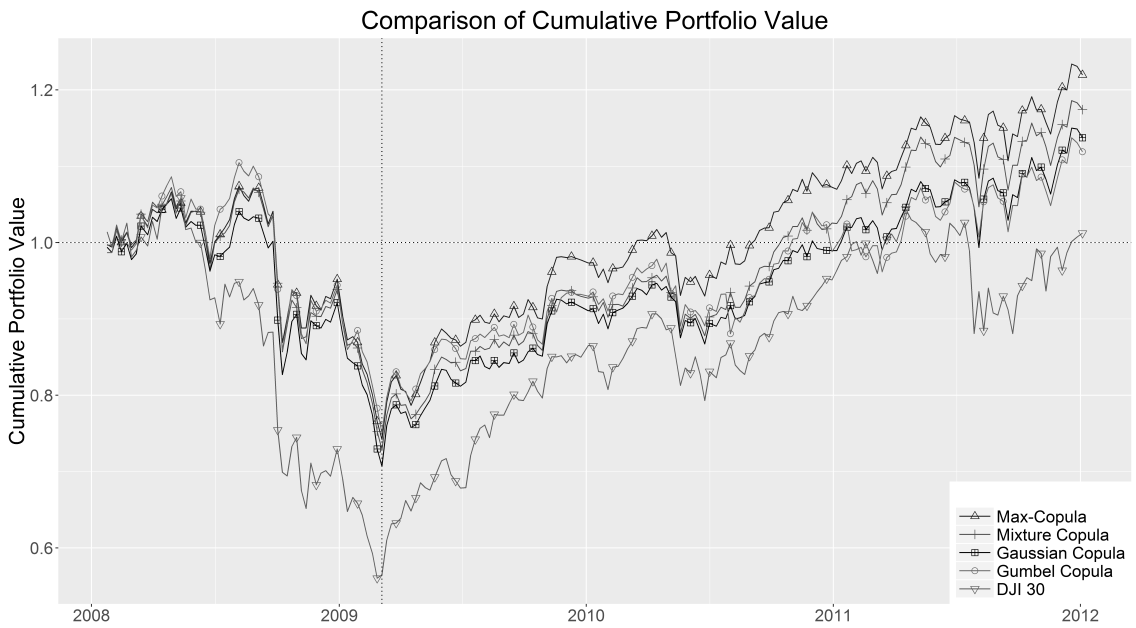

Figure 3.4: *This figure plots cumulative portfolio values of copula-based portfolios and DJI 30 over the period Dec 21, 2007 - Dec 30, 2011. The cumulative portfolio value on day t is calculated using* $\exp(\sum_{s=1}^{t} \hat{r}_s)$*, where* $\hat{r}_s$ *denotes the portfolio return on day s.*

asymmetric dependence and joint tail behavior while offering good performance in non-extremal behavior modeling. Max-copulas based on single-factor and block-factor models are developed to offer parsimonious modeling for structured dependence, especially in high-dimensional applications. Combined with semiparametric time series models, the proposed framework can further help obtain flexible and accurate models for multivariate time series. The consistency and asymptotic normality of the proposed CMLE has been affirmed through extensive numerical experiments. The max-copula's ability to model multivariate financial time series has been demonstrated by the estimation of conditional VaR for a financial portfolio and by weekly optimal portfolio constructions for the Dow Jones Industrial Average under various market scenarios.

Since the max-copula is relatively new, much work remains to be done. Its advantage and disadvantage relative to the mixture copula and the factor copula can further be explored. There are also several interesting extensions of the max-copula itself. A natural extension is to design a time-varying max-copula. Instead of setting the weight parameter c to be constant, we can design an autoregressive structure for c_t, such that c_t depends on $\{c_{t-i}, i > 0\}$ and on whether $\mathbf{C}_1$ or $\mathbf{C}_2$ takes lead on day $t-1$. Compared to constant c, a dynamic c_t may be more realistic in real data applications.

3.7 Appendix

3.7.1 Weak equivalence between linear factor copula model and factor copula based on C-vine

The idea of applying factor model in copula construction has been widely employed in the literature, e.g., Hull and White (2004) and Oh and Patton (2015). Most of the existing literature focus on constructing copulas from a linear factor model, one exception may be Krupskii and Joe (2013), in which the authors developed a general nonlinear factor copula construction framework. Under the proposed framework, bivariate copulas are used to iteratively specify the dependence relationship between latent factors and observed variables under a simplifying assumption (which is the same as the one made in vine copula, see below for more details). Because of the use of bivariate copula, the proposed framework is general and flexible, and it offers tail dependence and tail asymmetry. As is shown in Theorem 3.7.1, due to the simplifying assumption, under additive model framework, Krupskii and Joe (2013)'s factor copula model partially coincides with the copula constructed by linear factor model, e.g., Oh and Patton (2015).

Suppose we have an additive factor model such that, for $i = 1, \ldots, d$,

$$Y_i = f_i(X_1, X_2, \ldots, X_p) + \xi_i,$$

where $\{Y_i\}_{i=1}^d$ are observed random variables, $\{X_i\}_{i=1}^p$ are independent latent factors and $\{\xi_i\}_{i=1}^d$ are independent random noise. Assume that $f_i(\cdot)$'s are continuous and differentiable, $\{X_i\}_{i=1}^p$ are continuous random variables and $\{\xi_i\}_{i=1}^d$ are continuous random variables with *p.d.f.* $f_{\xi_i}(x) > 0$ for all $x \in \mathbb{R}$. Also, assume that $\{Y_i\}_{i=1}^d$ and $\{X_i\}_{i=1}^p$ fulfill the simplifying assumption in Krupskii and Joe (2013). Before presenting Theorem 3.7.1, we first state the simplifying assumption.

Definition 2. *Under the above framework, the simplifying assumption states that for each observed random variable Y_k, where $k = 1, \ldots, d$, the conditional copula between Y_k and X_i given $X_1, X_2, \ldots, X_{i-1}$ does not depend on the values of $X_1, X_2, \ldots, X_{i-1}$, for $i = 2, \ldots, p$, i.e., the*

conditional distribution of (Y_k, X_i) *given* $X_1, X_2, \ldots, X_{i-1}$ *can be written as,*

$$\begin{aligned}
&F_{Y_k,X_i|X_1,\ldots,X_{i-1}}(y, x_i|X_1, \ldots, X_{i-1}) \\
&= C_{k,i|1,2,\ldots,i-1}(F_{Y_k|X_1,\ldots,X_{i-1}}(y|X_1, \ldots, X_{i-1}), F_{X_i|X_1,\ldots,X_{i-1}}(x_i|X_1, \ldots, X_{i-1})) \\
&= C_{k,i|1,2,\ldots,i-1}(F_{Y_k|X_1,\ldots,X_{i-1}}(y|X_1, \ldots, X_{i-1}), F_{X_i}(x_i)),
\end{aligned}$$

where the last equality follows from the independence among X_i*'s.*

Intuitively, the simplifying assumption is saying that the values of $(X_1, \ldots, X_{i-1})$ only affect the conditional distribution of $(Y, X_i|X_1, \ldots, X_{i-1})$ through the conditional marginal $F_{Y_k|X_1,\ldots,X_{i-1}}(y|X_1, \ldots, X_{i-1})$, but not the conditional copula $C_{k,i|1,2,\ldots,i-1}(\cdot, \cdot)$. Now we state Theorem 3.7.1.

Theorem 3.7.1. *Under the above framework and the simplifying assumption, we have*

$$f_i(X_1, X_2, \ldots, X_p) = f_{i1}(X_1) + f_{i2}(X_2) + \ldots + f_{ip}(X_p), i = 1, 2, \ldots, d,$$

i.e., $f_i(X_1, X_2, \ldots, X_p)$*'s take the additive form.*

Based on Theorem 3.7.1, if the observed variables $\{Y_i\}_{i=1}^d$ are summation of functions of latent factors $f_i(X_1, X_2, \ldots, X_p)$'s and independent noise ξ_i's, then the simplifying assumption implies that the general function $f_i(\cdot)$'s actually take the additive form, which means that the newly constructed copulas taken by $\{Y_i\}_{i=1}^d$ are in the linear factor model framework. In this sense, Krupskii and Joe (2013) partially coincides with copulas constructed by additive linear factor models.

Roughly speaking, Theorem 3.7.1 establishes a weak equivalence of the nonlinear factor copula construction technique in Krupskii and Joe (2013) with the linear copula construction technique under additive model framework. Also, as is pointed out in Oh and Patton (2015), most of the widely used copulas (e.g., Gumbel copula, Clayton copula etc.) attain nonlinear stochastic representation. This motivates us to seek a general nonlinear copula construction technique.

3.7.2 Bivariate *p.d.f.* of the max-copula

As mentioned in Section 3.3.2.1 of the main text, the pairwise *c.d.f.* of (U_1, U_2) can be written as

$$
\begin{aligned}
&C_{12}(u_1, u_2) = \\
&= C_{Gaussian}(u_1^c, u_2^c, \rho) \cdot C_{Gumbel}(u_1^{1-c}, u_2^{1-c}, \alpha) \\
&= \Phi(\Phi^{-1}(u_1^c), \Phi^{-1}(u_2^c), \rho) \cdot \exp\left[- \left(\sum_{i=1}^{2} (-\log u_i^{1-c})^\alpha \right)^{1/\alpha} \right] \\
&= \Phi(\Phi^{-1}(u_1^c), \Phi^{-1}(u_2^c), \rho) \cdot \exp\left[-(1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right].
\end{aligned}
$$

By taking partial derivatives w.r.t. u_1 and u_2, the pairwise *p.d.f.* of (U_1, U_2) can be written as

$$
\begin{aligned}
&f_{12}(u_1, u_2) = \\
&\frac{1}{\sqrt{1-\rho^2}} \frac{c^2}{u_1^{1-c} u_2^{1-c}} \exp\left[- \frac{\rho^2 \left(\Phi^{-1}(u_1^c)\right)^2 + \rho^2 \left(\Phi^{-1}(u_2^c)\right)^2 - 2\rho \Phi^{-1}(u_1^c)\Phi^{-1}(u_2^c)}{2(1-\rho^2)} - (1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right] \\
&+ \frac{(1-c)c}{u_1^{1-c} u_2} \Phi\left(\frac{\Phi^{-1}(u_2^c) - \rho \Phi^{-1}(u_1^c)}{\sqrt{1-\rho^2}} \right) \exp\left(-(1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right) (-\log u_2)^{\alpha-1} \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha - 1} \\
&+ \frac{(1-c)c}{u_2^{1-c} u_1} \Phi\left(\frac{\Phi^{-1}(u_1^c) - \rho \Phi^{-1}(u_2^c)}{\sqrt{1-\rho^2}} \right) \exp\left(-(1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right) (-\log u_1)^{\alpha-1} \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha - 1} \\
&+ \Phi\left(\Phi^{-1}(u_1^c), \Phi^{-1}(u_2^c), \rho \right) \exp\left(-(1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} \right) \frac{1-c}{u_1 u_2} (\log u_1 \log u_2)^{\alpha-1} \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha - 2} \\
&\cdot \left[(1-c) \left(\sum_{i=1}^{2} (-\log u_i)^\alpha \right)^{1/\alpha} + \alpha - 1 \right].
\end{aligned}
$$

3.7.3 Simulation performance

3.7.3.1 Block-factor max-copula result

In this section we conduct numerical investigations on the performance of CMLE in the block-factor SDM model. We conduct a simulation experiment for a large block-factor SDM model with $d = 108$ and $p = 9$, i.e., the observations are of dimension 108 with 12 subjects in each of

the 9 groups. We set $c = 0.5, (\beta_1, \ldots, \beta_9) = (0.8, 0.8, 0.8, 1, 1, 1, 1.2, 1.2, 1.2)$, and $(\gamma_1, \ldots, \gamma_9) = (0.6, 0.6, 0.6, 0.8, 0.8, 0.8, 1, 1, 1)$ for the factor Gaussian copula, and set $(\alpha_0, \alpha_1, \ldots, \alpha_9) = (1.5, 1.7, 1.7, 1.7, 1.9, 1.9, 1.9, 2.1, 2.1, 2.1)$ for the one-level hierarchical Gumbel copula. Under the current setting, the Gaussian copula has within group correlations of 0.5, 0.62, 0.71 and between group correlations of $0.32, 0.35, 0.37, 0.38, 0.40, 0.42$; the Gumbel copula has within group upper tail dependence coefficients of 0.50, 0.56, 0.61 and common between group upper tail dependence coefficient of 0.41.

Due to computational complexity, 500 simulations have been conducted for each $T = (100, 500, 1000)$ and the result is summarized in Table 3.6. As can be clearly seen, although there are 29 parameters in total, the CMLE still performs well and gives consistent estimators, where the CMLE gives decent accuracy when $T = 500$.

	c	β_1	β_2	β_3	β_4	β_5	β_6	β_7
T=100	0.499 (0.224)	0.935 (0.428)	0.904 (0.395)	0.919 (0.356)	1.128 (0.373)	1.126 (0.346)	1.125 (0.357)	1.323 (0.374)
T=500	0.486 (0.131)	0.799 (0.128)	0.801 (0.133)	0.815 (0.083)	1.003 (0.118)	1.023 (0.120)	0.997 (0.123)	1.216 (0.183)
T=1000	0.493 (0.087)	0.787 (0.056)	0.795 (0.055)	0.802 (0.054)	1.005 (0.085)	0.997 (0.082	0.998 (0.074)	1.208 (0.104)
	β_8	β_9	γ_1	γ_2	γ_3	γ_4	γ_5	
T=100	1.344 (0.345)	1.260 (0.364)	0.707 (0.411)	0.696 (0.387)	0.713 (0.436)	0.912 (0.415)	0.888 (0.357)	
T=500	1.252 (0.173)	1.241 (0.164)	0.668 (0.138)	0.681 (0.157)	0.670 (0.137)	0.856 (0.162)	0.849 (0.123)	
T=1000	1.215 (0.108)	1.197 (0.095)	0.642 (0.082)	0.627 (0.076)	0.634 (0.086)	0.826 (0.077)	0.832 (0.088)	
	γ_6	γ_7	γ_8	γ_9	α_0	α_1	α_2	
T=100	0.897 (0.417)	1.006 (0.381)	1.006 (0.379)	1.072 (0.401)	1.437 (0.190)	1.809 (0.467)	1.840 (0.491)	
T=500	0.838 (0.140)	1.024 (0.202)	1.020 (0.149)	1.018 (0.142)	1.503 (0.085)	1.692 (0.143)	1.683 (0.148)	
T=1000	0.820 (0.073)	1.025 (0.096)	1.036 (0.083)	1.023 (0.080)	1.517 (0.044)	1.699 (0.078)	1.702 (0.080)	
	α_3	α_4	α_5	α_6	α_7	α_8	α_9	
T=100	1.758 (0.406)	1.944 (0.401)	1.954 (0.414)	1.992 (0.470)	2.179 (0.448)	2.154 (0.399)	2.128 (0.391)	
T=500	1.695 (0.150)	1.914 (0.164)	1.891 (0.172)	1.932 (0.180)	2.106 (0.161)	2.097 (0.165)	2.107 (0.177)	
T=1000	1.700 (0.084)	1.911 (0.100)	1.911 (0.084)	1.909 (0.075)	2.106 (0.099)	2.099 (0.096)	2.112 (0.103)	

Table 3.6: *Semiparametric CMLE for data generated by the block-factor SDM model with* $d = 108$ *and* $p = 9$. *The standard deviations of estimators are in brackets.*

3.7.3.2 Performance of bootstrap estimated variance

As mentioned in the main text, although the classical plug-in estimator for asymptotic covariance is available, due to analytical complexity, we instead use a parametric bootstrap method to estimate the variance for each of the estimators. The basic procedures are as follows. Based on the estimated $\hat{\lambda}$ for the dynamic component, $\hat{F}_j$'s for the nonparametric marginal component and $\hat{\theta}$ for the parametric max-copula component, we simulate the multivariate time series of length T by B times. For each of the B parametric bootstrap sample, we obtain a bootstrapped CMLE and we use the sample variance of the B bootstrapped CMLE's as an estimator for variance of the true CMLE.

We test the performance of the parametric bootstrap for the single-factor SDM model and block-factor SDM model in Section 3.4.2 of the main text. We set $B = 100$ and set all the parameters to be the same as the ones in Section 3.4.2. The performance of the 95% C.I. for true parameters constructed using the above parametric bootstrap method is summarized in Table 3.7 and Table 3.8.

	c	α	β_1	β_2	β_3	β_4
T=1000	0.91	0.98	0.93	0.93	0.93	0.90
T=2000	0.90	0.95	0.95	0.93	0.94	0.91
T=5000	0.96	0.98	0.95	0.94	0.94	0.94

Table 3.7: *Performance of 95 % parametric bootstrap C.I. for the single-factor SDM model with* $d = 4$.

	c	β_1	β_2	β_3	β_4	γ_1	γ_2
$T = 1000$	0.93	0.95	0.97	0.94	0.98	0.97	0.98
$T = 2000$	0.94	0.96	0.94	0.96	0.97	0.97	0.95
$T = 5000$	0.97	0.97	0.94	0.95	0.98	0.96	0.95
	γ_3	γ_4	α_0	α_1	α_2	α_3	α_4
$T = 1000$	0.96	0.97	0.96	0.93	0.98	0.97	0.97
$T = 2000$	0.98	0.97	0.97	0.98	0.93	0.97	0.99
$T = 5000$	0.95	0.98	0.93	0.95	0.94	0.97	0.95

Table 3.8: *Performance of 95 % parametric bootstrap C.I. for the block-factor SDM model with* $d = 20$ *and* $p = 4$.

As is clearly shown in Tables 3.7 and 3.8, the performance of parametric bootstrap C.I. is improving with the increase of sample size T and can achieve satisfactory results for all T, which

verifies that CMLE is asymptotically normally distributed.

3.7.4 Real data application

3.7.4.1 Back testing on Value at Risk

The expected and actual number of violations, as well as p-values of the binomial tests for all seven different portfolio weights are reported in Table 3.9. As can be seen, the estimated VaR by the max-copula performs well, which indicates that the single-factor SDM model captures the joint dynamics of the three stocks accurately.

Portfolio	Expected VaR Level	Actual VaR Level	Expected Violation	Actual Violation	p-value
$w_1 = .33$	0.950	0.952	164.75	158	0.63
$w_2 = .33$	0.990	0.991	32.95	28	0.43
$w_3 = .33$	0.995	0.995	16.48	17	0.81
$w_1 = .2$	0.950	0.952	164.75	158	0.63
$w_2 = .3$	0.990	0.992	32.95	27	0.34
$w_3 = .5$	0.995	0.997	16.48	10	0.14
$w_1 = .2$	0.950	0.951	164.75	161	0.81
$w_2 = .5$	0.990	0.992	32.95	27	0.34
$w_3 = .3$	0.995	0.995	16.48	18	0.71
$w_1 = .3$	0.950	0.954	164.75	150	0.25
$w_2 = .2$	0.990	0.992	32.95	26	0.25
$w_3 = .5$	0.995	0.996	16.48	13	0.46
$w_1 = .3$	0.950	0.950	164.75	164	1
$w_2 = .5$	0.990	0.990	32.95	32	1
$w_3 = .2$	0.995	0.994	16.48	21	0.26
$w_1 = .5$	0.950	0.953	164.75	156	0.52
$w_2 = .2$	0.990	0.989	32.95	36	0.60
$w_3 = .3$	0.995	0.995	16.48	16	1
$w_1 = .5$	0.950	0.951	164.75	161	0.81
$w_2 = .3$	0.990	0.989	32.95	35	0.66
$w_3 = .2$	0.995	0.995	16.48	18	0.71

Table 3.9: *Performance of conditional daily VaR based on the single-factor SDM model, with VaR level 0.95, 0.99 and 0.995.*

3.7.4.2 Optimal portfolio construction on Dow Jones Industrial Average

The 30 companies in DJI30 index come from six industrial sectors, according to the Standard Industrial Classification (SIC) system. The SIC system assigns each stock a three-digit code and

the first number of the code represents the industrial sector that the stock is in. The detailed group information of DJI30 in terms of SIC code is summarized in Table 3.10.

Company	AAPL	AXP	BA	CAT	CSCO	CVX	DD	DIS	GE	GS
SIC	357	614	372	351	367	291	287	799	351	621
Company	HD	IBM	INTC	JNJ	JPM	KO	MCD	MMM	MRK	MSFT
SIC	521	737	367	283	602	208	581	384	283	737
Company	NKE	PFE	PG	TRV	UNH	UTX	V	VZ	WMT	XOM
SIC	302	283	284	633	632	358	738	481	533	291

Table 3.10: *SIC code for the 30 component stocks of Dow Jones Industrial Average Index*

The summary statistics of the portfolio returns before the minimal point in the out-of-sample test set for DJI30 and optimal portfolios constructed by each of the four copulas can be found in Table 3.11, where the max-copula gives the second best Sharpe ratio after the Gumbel copula.

	TR	AR	Mean	SD	VaR95	VaR99	ES95	ES99	Sharpe Ratio
Max-copula	-26.67	-24.76	-0.48	2.81	5.09	8.97	7.94	10.23	-0.170
Mixture	-28.40	-26.37	-0.51	2.84	4.68	9.42	8.10	10.41	-0.179
Gaussian	-31.51	-29.26	-0.56	2.94	5.84	9.45	8.52	11.99	-0.192
Gumbel	-24.45	-22.70	-0.44	3.04	5.51	9.14	8.13	9.63	-0.143
DJI30	-57.92	-53.78	-1.03	4.07	6.70	12.47	11.37	18.22	-0.254

Table 3.11: *Summary statistics of the returns of different portfolios from December 21, 2007 to February 26, 2009. "TR" stands for total return, "AR" stands for annualized return and "SR" stands for Sharpe ratio. All numbers, except SR, are in percentage.*

The summary statistics of the portfolio returns after the minimal point in the out-of-sample test set for DJI30 and optimal portfolios constructed by each of the four copulas can be found in Table 3.12, where the max-copula gives the best Sharpe ratio and is followed by the mixture copula.

	TR	AR	Mean	SD	VaR95	VaR99	ES95	ES99	Sharpe Ratio
Max-copula	42.12	15.10	0.29	1.77	2.47	4.32	3.79	5.14	0.164
Mixture	39.78	14.27	0.27	1.77	2.36	4.90	3.93	5.46	0.155
Gaussian	38.50	13.81	0.27	1.84	2.41	5.18	4.01	5.71	0.145
Gumbel	32.08	11.50	0.22	1.96	3.50	5.09	4.42	5.99	0.113
DJI30	53.56	19.21	0.37	2.59	3.92	5.61	5.79	9.71	0.143

Table 3.12: *Summary statistics of the returns of different portfolios from February 27, 2009 to December 30, 2011. "TR" stands for total return, "AR" stands for annualized return and "SR" stands for Sharpe ratio. All numbers, except SR, are in percentage.*

3.7.5 Technical proof

3.7.5.1 Proof of theorems in the Appendix

Proof of Theorem 3.7.1: We only prove the case for $p = 2$, i.e., the two-factor case. The proof for $p > 2$ follows easily by using the same argument and induction. By the simplifying assumption in Definition 2, the conditional copula of (Y_k, X_2) given X_1 does not depend on the value of X_1, for $k = 1, \ldots, d$. The proof can proceed with any of the Y_k's, so in the following, we ignore the index k and use Y to denote a generic Y_k, where $Y = f(X_1, X_2) + \xi$.

Let Y' be an independent copy of Y, i.e., $Y' = f(X_1', X_2')+\xi'$, where (X_1', X_2', ξ') are independent copies of (X_1, X_2, ξ). By an abuse of notation, we use Y_1 to denote the conditional random variable of Y given $X_1 = z_1$ and use Y_2 to denote the conditional random variable of Y' given $X_1' = z_2$. Due to independence among X_1, X_2, and ξ, we have $Y_1 = f(z_1, X_2) + \xi$ and $Y_2 = f(z_2, X_2') + \xi'$. By the simplifying assumption, we know that the conditional copula of (Y, X_2) given X_1 does not depend on the value of X_1, so we have

$$(F_1(Y_1), F_{X_2}(X_2)) \stackrel{d}{=} (F_2(Y_2), F_{X_2}(X_2')),$$

$$\text{hence } (F_1(Y_1), X_2) \stackrel{d}{=} (F_2(Y_2), X_2'), \text{ since } X_2 \text{ is a continuous random variable,}$$

where F_{X_2} denotes the *c.d.f.* of X_2, F_1 and F_2 denote the *c.d.f.* of Y_1 and Y_2, i.e., the conditional *c.d.f.* of Y given $X_1 = z_1$ and $X_1 = z_2$ respectively. Notice that F_1 and F_2 depend on the values of z_1 and z_2 respectively. Also, F_1 and F_2 are strictly monotone increasing functions since ξ is continuous random variable with *p.d.f.* $f_\xi(x) > 0$, for all $x \in \mathbb{R}$.

Take the conditional distribution on X_2 and X_2' respectively, we have

$$(F_1(Y_1)|X_2 = x) \stackrel{d}{=} (F_2(Y_2)|X_2' = x), \qquad \text{for all } x \in \text{Range}(X_2).$$

By independence we have $F_1(f(z_1, x) + \xi) \stackrel{d}{=} F_2(f(z_2, x) + \xi')$ for all $x \in \text{Range}(X_2)$. We proceed the proof by introducing Lemma 3.7.1.

Lemma 3.7.1. *If* $F_1(f(z_1, x)+\xi) \stackrel{d}{=} F_2(f(z_2, x)+\xi')$ *for all* $x \in$ *Range*(X_2), *then* $F_1(f(z_1, x)+m) = F_2(f(z_2, x) + m)$ *for all* $x \in$ *Range*(X_2) *and* $m \in \mathbb{R}$.

Proof. By equivalence in distribution, we have $P(F_1(f(z_1,x)+\xi) \le K) = P(F_2(f(z_2,x)+\xi') \le K)$ for all $K \in [0,1]$, i.e., $P(\xi \le F_1^{-1}(K) - f(z_1,x)) = P(\xi' \le F_2^{-1}(K) - f(z_2,x))$, so we have

$$F_1^{-1}(K) - f(z_1,x) = F_2^{-1}(K) - f(z_2,x) \text{ for all } K \in [0,1].$$

Let $K = F_1(f(z_1,x)+m)$, where $m \in \mathbb{R}$, we have $f(z_1,x)+m-f(z_1,x) = F_2^{-1}(F_1(f(z_1,x)+m)) - f(z_2,x)$, thus $F_1(f(z_1,x)+m) = F_2(f(z_2,x)+m)$ for all $x \in \text{Range}(X_2)$ and $m \in \mathbb{R}$. □

By Lemma 3.7.1, the simplifying assumption implies that there exists a strictly monotonic increasing function $g(\cdot) = F_2^{-1} \circ F_1(\cdot)$ such that $g(f(z_1,x)+m) = f(z_2,x)+m$ for all $x \in \text{Range}(X_2)$ and $m \in \mathbb{R}$, where z_1 and z_2 are arbitrary fixed numbers in $\text{Range}(X_1)$. Note that $g(\cdot)$ as a function depends on the values of z_1 and z_2.

Let $g'(\cdot)$ denote the derivative of $g(\cdot)$, we have at point $f(z_1,x)+m$,

$$g'(f(z_1,x)+m) = \lim_{h\to 0} \frac{g(f(z_1,x)+m+h) - g(f(z_1,x)+m)}{h} = 1,$$

where the last equality follows from the fact that $g(f(z_1,x)+m) = f(z_2,x)+m$ for all $x \in \text{Range}(X_2)$ and $m \in \mathbb{R}$. By integration, we have $g(\cdot) = \cdot + b(z_1,z_2)$, where $b(z_1,z_2)$ is a function of (z_1,z_2). Thus we have $f(z_2,x) = f(z_1,x) + b(z_1,z_2)$ for all $z_1, z_2 \in \text{Range}(X_1)$ and $x \in \text{Range}(X_2)$.

Take the partial derivative w.r.t. x, we have

$$\frac{\partial f(z_1,x)}{\partial x} = \frac{\partial f(z_2,x)}{\partial x} \quad \text{for all } z_1, z_2 \in \text{Range}(X_1),$$

i.e., the partial derivative $\frac{\partial f(z,x)}{\partial x}$ is only a function of x, say $h(x)$.

Based on the above arguments, we can fix a $z_1 \in \text{Range}(X_1)$ and a $x_1 \in \text{Range}(X_2)$ such that for all $z \in \text{Range}(X_1)$ and $x \in \text{Range}(X_2)$ we have

$$\begin{aligned} f(z,x) &= f(z_1,x) + b(z_1,z) = f(z_1,x) - f(z_1,x_1) + f(z_1,x_1) + b(z_1,z) \\ &= \int_{x_1}^{x} \frac{\partial f(z_1,s)}{\partial x} ds + f(z_1,x_1) + b(z_1,z) = f(z_1,x_1) + \int_{x_1}^{x} h(s)ds + b(z_1,z), \end{aligned}$$

so $f(X_1,X_2) = f_1(X_1) + f_2(X_2)$, i.e., $f(X_1,X_2)$ takes the additive form. □

3.7.5.2 Proof of theorems in main text

Proof of Theorem 3.2.1: We only prove the case for λ^U_{max}, the proof for λ^L_{max} is similar. By definition, suppose $(U_1, U_2) \sim \mathbf{C}$, then

$$\lambda^U_{max} = \lim_{u\to 1} P(U_2 > u | U_1 > u) = \lim_{u\to 1} \frac{1 - P(U_1 < u) - P(U_2 < u) + P(U_1 < u, U_2 < u)}{P(U_1 > u)}$$
$$= \lim_{u\to 1} \frac{1 - 2u + \mathbf{C}_1(u^c, u^c) \cdot \mathbf{C}_2(u^{1-c}, u^{1-c})}{1-u} = c\lambda^U_1 + (1-c)\lambda^U_2.$$

where the last equality comes from L'Hospital rule and the fact that

$$\lambda^U_1 = \lim_{u\to 1} \frac{1 - 2u + \mathbf{C}_1(u,u)}{1-u},$$
$$\lambda^U_2 = \lim_{u\to 1} \frac{1 - 2u + \mathbf{C}_2(u,u)}{1-u}.$$

□

Proof of Theorem 3.3.1 : The proof of Theorem 3.3.1 follows the same line as the proof of Proposition 3.1 in Chen and Fan (2006a) and therefore is omitted. The proof details can be obtained from the authors upon request. □

Proof of Theorem 3.3.2: The proof of Theorem 3.3.2 follows the same line as the proof of Proposition 3.2 in Chen and Fan (2006a) and therefore is omitted. The proof details can be obtained from the authors upon request. □

Chapter 4

Semiparametric Multivariate D-vine Copula Model for Multivariate Time Series

4.1 Motivation

Modeling multivariate time series is the fundamental task for many statistical applications, see Patton (2012) for applications in economics, Erhardt et al. (2015) for applications in climate monitoring and Smith (2015) for applications in energy markets. Roughly speaking, there are two types of dependence for multivariate time series. The first type is the temporal dependence for each of its component univariate time series. The second type is the cross-sectional dependence across all of its component univariate time series. To capture the behavior of multivariate time series, it is essential for the statistical model to have an accurate modeling of both the cross-sectional and the temporal dependence. This requires the time series model to be highly flexible such that it is capable of modeling various dependence structures, including nonlinear dependence, tail dependence and asymmetric dependence.

The construction of multivariate distributions is technically difficult and most of the early multivariate modeling is restricted within the Gaussian/elliptical family. Thanks to Sklar (1959)'s theorem, which states that multivariate dependency can be separated into a copula and individual

marginal distributions, the "time of the copula" has emerged for the construction of multivariate distributions. Copulas are d-dimensional distribution functions on $[0,1]^d$ with uniform margins. By Sklar (1959)'s theorem, any multivariate distribution $\mathbf{F}$ can be separated into its marginals $(F_1, \ldots, F_d)$ and its copula C, where the copula captures all the scale-free dependence in the multivariate distribution. In particular, suppose there is a random vector $\mathbf{Y} \in \mathbb{R}^d$ such that $\mathbf{Y}$ follows $\mathbf{F}$, then we have $\mathbf{F}(\mathbf{y}) = C(F_1(y_1), \ldots, F_d(y_d))$, where $\mathbf{y} = (y_1, \ldots, y_d)'$ is the realization of $\mathbf{Y}$. If all the marginals of $\mathbf{F}$ are absolutely continuous, the copula C is unique.

Various copulas have been proposed in the literature, see Joe (2014) and Nelsen (1999) for a summary. Copula-based models for multivariate distributions are widely used in a variety of applications, see Frees and Valdez (1998) in actuarial science and insurance, Cherubini et al. (2004) in finance, and Genest and Favre (2007) in hydrology.

Most applications of copulas under time series setting focus on the modeling of the cross-sectional dependence of multivariate time series, see the semiparametric copula-based multivariate dynamic (SCOMDY) models in Patton (2006) and Chen and Fan (2006a). Under the SCOMDY framework, conventional univariate time series models, e.g. ARMA and GARCH (Engle, 1982, Bollerslev, 1986), are used to capture the temporal dependence, in particular conditional mean and variance, of each component univariate time series. A parametric copula and nonparametric marginal distributions are used to specify the cross-sectional dependence across the innovations of all the component univariate time series.

There are also applications where copulas are used for modeling the temporal dependence of a univariate time series. Darsow et al. (1992) provide a necessary and sufficient condition for a copula-based time series to be a Markov process. Joe (1997), Chen and Fan (2006b), Domma et al. (2009), Ibragimov (2009) and Beare (2010) consider copula-based stationary Markov chains, where copulas and flexible marginal distributions are used to specify the transitional probability of the Markov chains. Most copula-based time series models focus on first-order Markov chains, since there are many flexible and tractable bivariate copulas. For higher-order Markov chains, the result is rare and mostly is based on elliptical copulas such as Gaussian copula and t copula due to their tractability.

To extend the copula-based univariate time series model to higher-order Markov chains, a framework to generate flexible yet tractable multivariate copulas is required. A promising direction

is to use the recently developed vine-copula (see Aas et al. (2009)), which is a framework that generates flexible multivariate copulas based on iterative pairwise construction of bivariate copulas. The D-vine copula, which is a special structured vine-copula, is of particular interest due to its simplicity and natural interpretation under time series setting. Smith et al. (2010) and Shi and Yang (2017) use D-vine for modeling longitudinal data. According to Aas et al. (2009), the density of a univariate time series $\mathbf{Y} = \{Y_t\}_{t=1}^T \in \mathbb{R}^T$ based on D-vine is given by its marginal distribution $F(\cdot)$ and $T(T-1)/2$ bivariate copulas $\{\{c_{s,t}\}_{s=1}^{t-1}\}_{t=1}^T$,

$$\begin{aligned} f_D(\mathbf{y};\boldsymbol{\beta}) &= \prod_{t=1}^{T} f(y_t|y_{t-1},\ldots,y_1) \\ &= \prod_{t=1}^{T} f(y_t) \prod_{s=1}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1});\beta_{s,t}), \end{aligned}$$

where $f(\cdot)$ is the pdf of $F(\cdot)$, $F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1})$ and $F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1})$ are conditional distributions of Y_s and Y_t given variables $(Y_{s+1},\ldots,Y_{t-1})$, and can be calculated recursively based on the marginal $F(\cdot)$ and the bivariate copulas $\{c_{s,t}\}$ by the algorithm in Aas et al. (2009). The parameter of the bivariate copula $c_{s,t}$ is denoted by $\beta_{s,t}$.

An example of a D-vine for $T = 5$ observations is exhibited in Figure 4.1. The nodes in tree 1 represent ordered marginals $\{F(Y_t)\}_{t=1}^T$ and the edges in each tree becomes the nodes in the next tree. From left to right, the sth edge in tree $t-s$ corresponds to the (conditional) bivariate copula $c_{s,t}$ that is used in $f_D(\mathbf{y},\boldsymbol{\beta})$ to specify the conditional joint distribution of (Y_s, Y_t) given variables $(Y_{s+1},\ldots,Y_{t-1})$. The edges of the entire D-vine indicate the bivariate copulas $\{\{c_{s,t}\}_{s=1}^{t-1}\}_{t=1}^T$ that contribute to the pair copula constructions. The key feature of the D-vine is that the nodes of each tree only connect adjacent nodes, which makes it simple to understand and naturally interpretable for time series.

As mentioned in Smith (2015), despite that copulas have been employed in either cross-sectional or temporal dependence modeling of time series, there is few multivariate time series model that uses copulas to simultaneously characterize both the cross-sectional and temporal dependence. One exception is Smith (2015) itself, where the author stacks the multivariate time series into a univariate time series and designs copula-based dependence structures for the resulted univariate time series. The proposed model in Smith (2015) is flexible and can capture nonlinear and asymmetric

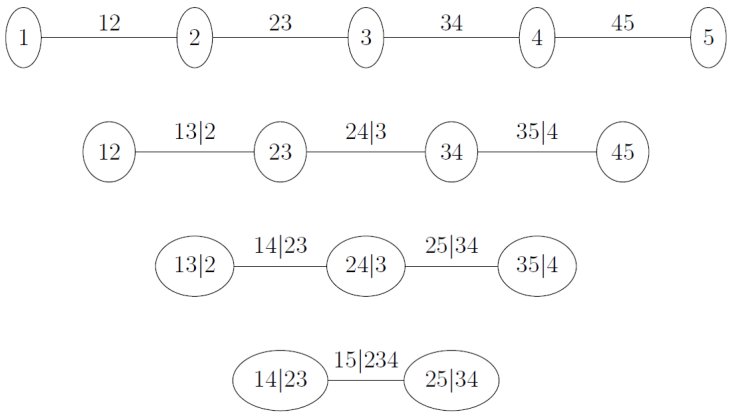

Figure 4.1: *A 5-dimension D-vine.*

tail dependence, and it outperforms the conventional Vector AR model in various applications. One potential drawback is that the model is complicated and may not be easily implemented by practitioners.

In this paper, we aim to design a simple, intuitive and flexible multivariate time series model that enables the simultaneous copula-based modeling for both temporal and cross-sectional dependence. Specifically, based on the D-vine copula, we first design a univariate D-vine time series model (uDvine) for the modeling of each component univariate time series. The uDvine generalizes the existing first-order copula-based Markov chain to an arbitrary-order Markov chain and is capable of modeling various temporal dependence structures due to the use of D-vine copula. Using the same idea of SCOMDY model, we further design a multivariate D-vine time series model (mDvine) based on the uDvine, where a parametric copula is employed to specify the cross-sectional dependence among all the uDvine-based component univariate time series. The proposed mDvine is interpretable, tractable and offers flexible dependence structures for the modeling of multivariate time series.

In Section 4.2, we present a detailed description of uDvine and mDvine and investigate their probabilistic properties. In Section 4.3, a sequential model selection procedure and a two-stage maximum likelihood approach (MLE) are proposed for model inference and estimation, with their statistical properties derived. Empirical studies are carried out in Section 4.4 to investigate the flexibility of mDvine and to demonstrate the performance of the MLE. The results of mDvine in two real data applications, one on Australian electricity price and one on Ireland wind speed, are

shown in Section 4.5, where significant improvement over conventional time series models has been observed. We conclude the paper in Section 4.6. The Appendix contains the proof of the theorems and other technical materials.

4.2 The D-vine Based Time Series Model

4.2.1 Univariate D-vine time series model (uDvine)

In this section, we introduce the univariate D-vine time series model (uDvine), which generalizes the popular existing first-order copula-based Markov chains to an arbitrary-order Markov chains using D-vine copula. As is shown later, the proposed uDvine is intuitive and is capable of modeling various temporal dependence structures due to the use of D-vine copula.

4.2.1.1 Model specification of uDvine

As mentioned above, the density of a univariate time series $\mathbf{Y} = \{Y_t\}_{t=1}^T \in \mathbb{R}^T$ based on a D-vine copula is given by its marginal distribution $F(\cdot)$ and $T(T-1)/2$ bivariate copulas $\{\{c_{s,t}\}_{s=1}^{t-1}\}_{t=1}^T$,

$$\begin{aligned} f_D(\mathbf{y};\boldsymbol{\beta}) &= \prod_{t=1}^{T} f(y_t|y_{t-1},\ldots,y_1) \\ &= \prod_{t=1}^{T} f(y_t) \prod_{s=1}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1});\beta_{s,t}), \end{aligned}$$

where $f(\cdot)$ is the pdf of $F(\cdot)$, $F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1})$ and $F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1})$ are conditional distributions of Y_s and Y_t given variables $(Y_{s+1},\ldots,Y_{t-1})$, and can be calculated recursively based on the marginal $F(\cdot)$ and the bivariate copulas $\{c_{s,t}\}$ by the algorithm in Aas et al. (2009).

A necessary condition for the stationarity of $\{Y_t\}_{t=1}^T$ is that bivariate copulas in the same tree must be exactly the same, i.e. $c_{s,t} = c_{s',t'}$ if $t-s = t'-s'$, we call this the homogeneity condition of the D-vine copula. Thus, under stationarity assumption of $\mathbf{Y}$, to fully specify $f_D(\mathbf{y},\boldsymbol{\beta})$, we need to specify all edges from tree 1 to tree $T-1$, which requires $T-1$ bivariate copulas. This is unrealistic when T is large and keeps growing.

The natural solution is that we 'truncate' the D-vine after a certain tree (say tree p) and set

all copulas beyond pth tree, $\{c_{s,t}, t-s>p\}$, to be independent copulas, where $p \ll T$. We call the univariate D-vine time series model truncated at level p the uDvine(p) model. As is shown later in Proposition 4.2.1, uDvine(p) is a p-order homogeneous Markov chain.

To fully specify uDvine(p), we need the marginal distribution function $F(\cdot)$ and p bivariate copulas for the p trees. In this paper, we only assume F is absolutely continuous and do not impose any parametric assumption on F, which makes the uDvine a semiparametric time series model. The joint distribution of $\{Y_t\}_{t=1}^T$ from uDvine(p) can be written as

$$\begin{aligned} f_D(\mathbf{y};\boldsymbol{\beta}) &= \prod_{t=1}^{T} f(y_t|y_{t-1},\ldots,y_1) = \prod_{t=1}^{T} f(y_t|y_{t-1},\ldots,y_{1\vee(t-p)}) \\ &= \prod_{t=1}^{T} f(y_t) \prod_{s=1\vee(t-p)}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\cdots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\cdots,y_{t-1});\beta_{s,t}), \end{aligned}$$

where $c_{s,t}$ is the bivariate copula in tree $t-s$ having parameters $\beta_{s,t}$, $F_{s|(s+1):(t-1)}$ and $F_{t|(s+1):(t-1)}$ are the conditional distribution functions of Y_s and Y_t given $(Y_{s+1},\cdots,Y_{t-1})$. By the homogeneity of uDvine(p), we have $F_{s|(s+1):(t-1)} = F_{t|(s+1):(t-1)}$ for all eligible (s,t). Also, we have $F_{s|(s+1):(t-1)} = F_{s'|(s'+1):(t'-1)}$, $F_{t|(s+1):(t-1)} = F_{t'|(s'+1):(t'-1)}$ and $c_{s,t} = c_{s',t'}$ for all eligible (s,t,s',t') such that $t'-s' = t-s$. We denote $\boldsymbol{\beta} = \{\beta_{s,t}\}$ as the collection of all parameters for the bivariate copulas and denote $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$.

By Aas et al. (2009), for $t > p$, the conditional density function of Y_t takes the form

$$\begin{aligned} f(y_t|\mathcal{F}_{t-1}) &= f(y_t|y_{t-1},y_{t-2},\cdots,y_{t-p}) \\ &= f(y_t)\cdot \prod_{s=(t-p)}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\cdots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\cdots,y_{t-1});\beta_{s,t}), \end{aligned}$$

which is a function of $f(y_t)$, $\{F(y_{t-k})\}_{k=0}^p$ and $\boldsymbol{\beta}$. For simplicity of notation, we denote

$$\begin{aligned} &w(F(y_t),F(y_{t-1}),\cdots,F(y_{t-p});\boldsymbol{\beta}) \\ &= \prod_{s=(t-p)}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\cdots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\cdots,y_{t-1});\beta_{s,t}). \end{aligned} \tag{4.1}$$

where $w(u_1,u_2,\cdots,u_{p+1};\boldsymbol{\beta})$ can be derived analytically based on the algorithms in Aas et al. (2009). Together, we have $f(y_t|y_{t-1},y_{t-2},\cdots,y_{t-p}) = f(y_t)\cdot w(F(y_t),F(y_{t-1}),\cdots,F(y_{t-p});\boldsymbol{\beta})$.

Similarly, for $t > p$, the conditional cumulative distribution function of Y_t given $\mathcal{F}_{t-1}$ is also a function of $\{F(y_{t-k})\}_{k=0}^{p}$ and $\boldsymbol{\beta}$. For simplicity of notation, we denote

$$F(y_t|\mathcal{F}_{t-1}) = F(y_t|y_{t-1}, \cdots, y_{t-p}) = g(F(y_t), F(y_{t-1}), \cdots, F(y_{t-p}); \boldsymbol{\beta}), \tag{4.2}$$

where $g(u_1, u_2, \cdots, u_{p+1}; \boldsymbol{\beta})$ can be derived analytically based on the algorithms in Aas et al. (2009).

The uDvine(p) is a general time series model and incorporates many commonly used time series model as special cases. All the first-order copula-based Markov chains in the literature, e.g. Chen and Fan (2006b), are essentially uDvine(1). In fact, all the stationary first-order Markov chains with absolute continuous marginals in $\mathbb{R}$, e.g. AR(1) models and ARCH(1) models in Engle (1982), are special cases of uDvine(1). Another important special case of uDvine(p) is a stationary AR(p) process with Gaussian innovations (see Aas et al. (2009) for the detail).

4.2.1.2 A uDvine(2) example

In this section, we give the detailed formulas of distributions for uDvine(2). We assume that $\{Y_t\}_{t=1}^{T}$ follows uDvine(2) with marginal distribution $F(\cdot)$, first tree bivariate copula C_1 with parameters β_1 and second tree bivariate copula C_2 with parameters β_2. Since $p = 2$, for a given t, we know that s can be $t-1$ or $t-2$.

For $s = t-1$, we have

$$F_{s|(s+1):(t-1)}(y_s|y_{s+1}, \cdots, y_{t-1}) = F(y_s) \text{ and } F_{t|(s+1):(t-1)}(y_t|y_{s+1}, \cdots, y_{t-1}) = F(y_t).$$

For $s = t-2$, we have

$$F_{s|(s+1):(t-1)}(y_s|y_{s+1}, \cdots, y_{t-1}) = F(y_s|y_{s+1}) = \frac{\partial}{\partial 2} C_1(F(y_s), F(y_{s+1}); \beta_1),$$
$$F_{t|(s+1):(t-1)}(y_t|y_{s+1}, \cdots, y_{t-1}) = F(y_t|y_{t-1}) = \frac{\partial}{\partial 2} C_1(F(y_t), F(y_{t-1}); \beta_1).$$

For $t > 2$, the conditional density function is

$$\begin{aligned} &f(y_t|y_{t-1}, y_{t-2}) = c_2(F(y_t|y_{t-1}), F(y_{t-2}|y_{t-1}); \beta_2) \cdot c_1(F(y_t), F(y_{t-1}); \beta_1) \cdot f(y_t) \\ =&c_2\left(\frac{\partial}{\partial 2} C_1(F(y_t), F(y_{t-1}); \beta_1), \frac{\partial}{\partial 2} C_1(F(y_{t-2}), F(y_{t-1}); \beta_1); \beta_2\right) \cdot c_1(F(y_t), F(y_{t-1}); \beta_1) \cdot f(y_t), \end{aligned}$$

thus $w(u_1, u_2, u_3; \boldsymbol{\beta}) = c_2\left(\frac{\partial}{\partial 2}C_1(u_1, u_2; \beta_1), \frac{\partial}{\partial 2}C_1(u_3, u_2; \beta_1); \beta_2\right) \cdot c_1(u_1, u_2; \beta_1)$.

For $t > 2$, the conditional cumulative distribution function is

$$\begin{aligned} F(y_t|y_{t-1}, y_{t-2}) &= \frac{\partial}{\partial 2}C_2(F(y_t|y_{t-1}), F(y_{t-2}|y_{t-1}); \beta_2) \\ =&\frac{\partial}{\partial 2}C_2\left(\frac{\partial}{\partial 2}C_1(F(y_t), F(y_{t-1}); \beta_1), \frac{\partial}{\partial 2}C_1(F(y_{t-2}), F(y_{t-1}); \beta_1); \beta_2\right), \end{aligned}$$

thus $g(u_1, u_2, u_3; \boldsymbol{\beta}) = \frac{\partial}{\partial 2}C_2\left(\frac{\partial}{\partial 2}C_1(u_1, u_2; \beta_1), \frac{\partial}{\partial 2}C_1(u_3, u_2; \beta_1); \beta_2\right)$.

4.2.1.3 Stationarity and ergodicity of uDvine

In this section, we show that under certain conditions, the univariate time series $\{Y_t\}$ generated by uDvine(p) is stationary and ergodic.

Proposition 4.2.1. *Under the homogeneity condition of D-vine, the univariate time series $\{Y_t\}$ generated by uDvine(p) is a p-order homogeneous Markov chain.*

By Proposition 4.2.1, if we define $X_t = (F(Y_t), F(Y_{t-1}), \ldots, F(Y_{t-p+1}))$, the new process $\{X_t\}$ is a first-order homogeneous Markov chain with state space $(0, 1)^p$. Since the marginal distribution $F(\cdot)$ of uDvine is absolutely continuous, we know that $F(Y_t)$ marginally follows the uniform distribution on $(0, 1)$. As is noted in Chen and Fan (2006b), the stationarity and ergodicity of $\{Y_t\}$ and $\{F(Y_t)\}$ are equivalent due to the absolute continuity of the marginal distribution $F(\cdot)$. Theorem 4.2.1 gives sufficient conditions for the ergodicity of $\{X_t\}$ and thus that of $\{Y_t\}$.

Theorem 4.2.1. *Under Assumption 4.7.1 and 4.7.2 in Section 4.7.1 in the Appendix, $\{X_t\}$ is Harris recurrent and geometrically ergodic, so is $\{Y_t\}$, which follows uDvine(p).*

A direct result of Theorem 4.2.1 is the β-mixing property of uDvine(p), which is later used for the analysis of MLE properties.

Corollary 4.2.1. *If Theorem 4.2.1 holds, uDvine(p) is β-mixing with an exponential decaying rate.*

4.2.2 Multivariate D-vine time series model (mDvine)

The ultimate goal of this paper is to develop a flexible statistical model for multivariate time series. The proposed uDvine offers flexible modeling for the temporal dependence of the component

univariate time series. To fully specify the multivariate time series model, we need to design the cross-sectional dependence across the component univariate time series. Following the SCOMDY framework in Chen and Fan (2006a), we propose the multivariate D-vine time series model (mDvine).

Denote a d-dimensional multivariate time series by $\{\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{td})\}_{t=1}^T$. We say that $\{\mathbf{Y}_t\}_{t=1}^T$ follows an mDvine if marginally all its component univariate time series follow uDvine(p_i), for $i = 1, \ldots, d$, and the conditional joint distribution $\mathbf{F}(\cdot)$ of $\mathbf{Y}_t$ given the past information $\mathcal{F}_{t-1} = \sigma(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots)$ can be written as

$$\mathbf{F}(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{F}(y_{t1}, \ldots, y_{td}|\mathcal{F}_{t-1}) = C_J(F_1(y_{t1}|\mathcal{F}_{t-1}^1), \ldots, F_d(y_{td}|\mathcal{F}_{t-1}^d)), \tag{4.3}$$

where $C_J(\cdot)$ is a d-dimensional copula designed to capture the cross-sectional dependence, $\mathcal{F}_{t-1}^i = \sigma(Y_{t-1,i}, Y_{t-2,i}, \ldots)$ is the sigma field generated by the ith component univariate time series and $F_i(\cdot|\mathcal{F}_{t-1}^i)$ are the conditional distribution of Y_{ti} given its own history.

As mentioned before, since uDvine(p_i) is a p_i-order Markov chain, we have that $F_i(y_{ti}|\mathcal{F}_{t-1}^i) = F_i(y_{ti}|y_{t-1,i}, \ldots, y_{t-p,i})$. Given the marginal distribution $F_i(\cdot)$ and the bivariate copula parameter $\boldsymbol{\beta}_i$ of the ith uDvine(p_i), $F_i(y_{ti}|y_{t-1,i}, \ldots, y_{t-p,i})$ is a function of $\{F_i(y_{t-k,i})\}_{k=0}^p$ and $\boldsymbol{\beta}_i$ such that

$$F_i(y_{ti}|\mathcal{F}_{t-1}^i) = F_i(y_{ti}|y_{t-1,i}, \ldots, y_{t-p,i}) = g_i(F_i(y_{ti}), F_i(y_{t-1,i}), \ldots, F_i(y_{t-p,i}); \boldsymbol{\beta}_i), \tag{4.4}$$

where $g_i(u_1, \ldots, u_{p+1}; \boldsymbol{\beta}_i)$ can be derived analytically based on the algorithms in Aas et al. (2009) as demonstrated in Section 4.2.1.2.

Notice here there are two implicit assumptions of mDvine: A1. the conditional marginal distribution of the ith component univariate time series Y_{ti} given $\mathcal{F}_{t-1}$ only depends on its own history $\mathcal{F}_{t-1}^i$; A2. the copula C_J for the conditional joint distribution of $\mathbf{Y}_t$ given $\mathcal{F}_{t-1}$ does not depend on $\mathcal{F}_{t-1}$. Both assumptions are standard in the literature of multivariate time series modeling, e.g. see Chen and Fan (2006a), Dias and Embrechts (2010), Oh and Patton (2013) and Oh and Patton (2015) for more examples.

The mDvine employs copulas for modeling both the temporal and cross-sectional dependence of the multivariate time series. The use of semiparametric uDvine introduces extra flexibility in

the modeling of the marginal behavior of each component univariate time series and the copula $C_J(\cdot)$ in SCOMDY framework enables more sophisticated cross-sectional dependence structures. As is seen later in real data analysis, the simultaneous copula-based modeling for temporal and cross-sectional dependence provided by mDvine helps it generate more sophisticated dependence structures such as nonlinear dependence and tail dependence and helps it offer a promising and more favorable performance compared to the conventional multivariate time series models in many cases.

4.3 Two-stage Semiparametric Maximum Likelihood Estimation

As is pointed out by Aas et al. (2009), the inference for D-vine consists of two parts: (a) the choice of bivariate copula types and (b) the estimation of the copula parameters. The same tasks apply to mDvine. In Section 4.3.1, we talk about the model selection of mDvine, specifically, we give a sequential model selection procedure for its component uDvines. In Section 4.3.2, we propose a two-stage MLE for the estimation of all the parameters of a given mDvine.

4.3.1 Selection of bivariate copulas for uDvine

To use mDvine as a statistical model, we need to select the order p and the bivariate copulas $c_{s,t}$ for each of its component uDvine and we also need to select the cross-sectional copula $C_J(\cdot)$. The selection of $C_J(\cdot)$ can be solved by Akaike information criterion (AIC) or Bayesian information criterion (BIC), which is the standard procedure in the literature. Here we focus on the model selection of the component uDvine.

For a uDvine, we need to select its order p and the bivariate copulas $c_{s,t}$ for its p trees. Given a set of candidate copulas (say m copulas) and a given order p, the number of possible uDvines is m^p, which can be quite large even for moderate m and p. For computational feasibility, we adopt the tree-by-tree sequential selection procedure as described in Shi and Yang (2017).

The basic procedure is as follows. We start with the first tree, selecting the appropriate copula from a given set of candidates and estimating its parameters. Fixing the copula and its parameters in the first tree, we then select the optimal copula and estimate its dependence parameters for the second tree. We continue selecting copulas and estimating parameters for the next tree of a

higher order while holding the selected copulas and their parameters fixed in all previous trees. If an independence copula is selected for a certain tree, we then truncate the uDvine, i.e. assume conditional independence in all higher order trees (see, for example, Brechmann et al. (2012)). The commonly used model selection method BIC is employed for the copula selection of each tree.

As is shown in the simulation study, this sequential model selection procedure for uDvine is computationally efficient and can select the optimal models accurately. Similar findings have been reported in Shi and Yang (2017). More properties of the sequential estimation procedures for D-vine can be found in Brechmann et al. (2012) and Haff (2013).

4.3.2 Two-stage MLE for mDvine

After fixing the parametric form of mDvine, there are three components that need to be estimated: (a) the true marginal distributions $F_1^0(\cdot), F_2^0(\cdot), \ldots, F_d^0(\cdot)$ of the d component uDvines, (b) the true bivariate copula parameters $\boldsymbol{\beta}_1^0, \ldots, \boldsymbol{\beta}_d^0$ of the d component uDvines, (c) the true parameter γ^0 of the cross-sectional copula $C_J(\cdot)$.

The number of parameters to be estimated are at least $\sum_{i=1}^d p_i + d$ if we assume all the bivariate copulas of uDvines are single-parameter copulas. The full likelihood based estimation can be computationally expensive especially if dimension d is big. For feasible estimation purpose, we instead use the standard two-stage maximum likelihood estimator (MLE) in the copula literature. Throughout this section, we assume that the parametric form (i.e. the bivariate copula types for each uDvine(p_i) and the cross-sectional copula type for $C_J(\cdot)$) of mDvine is known and we present the properties of the two-stage MLE under the correct model specification.

Given observations $\{\mathbf{y}_t = \{y_{ti}\}_{i=1}^d\}_{t=1}^T$, the conditional likelihood function of $\mathbf{y}_t$ given the history can be written as

$$f(\mathbf{y}_t|\mathcal{F}_{t-1}) = c_J(F_1(y_{t1}|\mathcal{F}_{t-1}^1), \ldots, F_d(y_{td}|\mathcal{F}_{t-1}^d); \gamma) \prod_{i=1}^d f_i(y_{ti}|\mathcal{F}_{t-1}^i), \tag{4.5}$$

where the conditional marginal distributions $F_i(y_{ti}|\mathcal{F}_{t-1}^i)$ and $f_i(y_{ti}|\mathcal{F}_{t-1}^i)$ can be derived from the marginal distribution $F_i(\cdot)$ and bivariate copula parameter $\boldsymbol{\beta}_i$ of the ith uDvine.

For simplicity of notation, without loss of generality, we assume that the order of all uDvines

to be p. Based on (4.5), the conditional log-likelihood function is

$$\begin{aligned} L(F_1,\ldots,F_d;\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_d;\gamma\big|\{\mathbf{y}_t\}_{t=1}^T) &= \sum_{t=p+1}^{T}\log f(\mathbf{y}_t|\mathcal{F}_{t-1}) \\ &= \sum_{t=p+1}^{T}\log c_J(F_1(y_{t1}|\mathcal{F}_{t-1}^1),\ldots,F_d(y_{td}|\mathcal{F}_{t-1}^d);\gamma)+\sum_{i=1}^{d}\sum_{t=p+1}^{T}\log f_i(y_{ti}|\mathcal{F}_{t-1}^i). \end{aligned} \tag{4.6}$$

In the first stage, the marginal distribution $F_i^0(\cdot)$ and bivariate copula parameter $\boldsymbol{\beta}_i^0$ are estimated based on each component univariate time series $\{y_{ti}\}_{t=1}^T$. The marginal distributions $F_i^0(\cdot)$ is estimated by the rescaled empirical distribution functions $\hat{F}_i(\cdot)$ where $\hat{F}_i(\cdot)=\frac{1}{T+1}\sum_{t=1}^{T}I(y_{ti}\le\cdot)$. Given the estimated marginal distribution $\hat{F}_i(\cdot)$, the MLE $\hat{\boldsymbol{\beta}}_i$ for $\boldsymbol{\beta}_i^0$ can be calculated by maximizing

$$L_{1i}(\boldsymbol{\beta}_i)=\sum_{t=p+1}^{T}\log f_i(y_{ti}|\mathcal{F}_{t-1}^i)=\sum_{t=p+1}^{T}\left[\log f_i(y_{ti})+\log w_i(\hat{F}_i(y_{ti}),\cdots,\hat{F}_i(y_{t-p,i});\boldsymbol{\beta}_i)\right], \tag{4.7}$$

where the last equality follows from (4.1). The starting point of the optimization can be obtained by the sequential estimation procedure described in Haff (2013).

In the second stage, given the first stage estimation of $\{\hat{F}_i(\cdot)\}_{i=1}^d$ and $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$, the MLE $\hat{\gamma}$ for the parameter γ^0 can be calculated by maximizing

$$\begin{aligned} L_2(\gamma) &= \sum_{t=p+1}^{T}\log c_J(\hat{F}_1(y_{t1}|\mathcal{F}_{t-1}^1),\ldots,\hat{F}_d(y_{td}|\mathcal{F}_{t-1}^d);\gamma) \\ &= \sum_{t=p+1}^{T}\log c_J(g_1(\hat{F}_1(y_{t1}),\ldots,\hat{F}_1(y_{t-p,1});\hat{\boldsymbol{\beta}}_1),\ldots,g_d(\hat{F}_d(y_{td}),\ldots,\hat{F}_d(y_{t-p,d});\hat{\boldsymbol{\beta}}_d);\gamma), \end{aligned} \tag{4.8}$$

where the last equality follows from (4.2).

4.3.3 Consistency and normality

Both the first stage MLE of the bivariate copula parameters $\{\boldsymbol{\beta}_i^0\}_{i=1}^d$ and the second stage MLE of the cross-sectional copula parameter γ^0 are essentially the so-called semiparametric two-stage estimator in Newey and McFadden (1994), where the estimator is obtained based on estimating equations that consist of some estimated infinite-dimensional functions. In our case, the estimated infinite-dimensional functions for $\{\beta_i^0\}_{i=1}^d$ are the corresponding rescaled empirical distributions

$\{\hat{F}_i(\cdot)\}_{i=1}^d$ and the ones for γ^0 are all the first stage estimators, i.e. $\{\hat{F}_i(\cdot)\}_{i=1}^d$ and $\{\hat{\beta}_i\}_{i=1}^d$.

The semiparametric two-stage estimator is not uncommon in econometric literature. Newey and McFadden (1994) offer a general treatment on its consistency and normality. In this paper we mainly follow the theoretical result in Chen and Fan (2006b) and Chen and Fan (2006a), where the authors provide the consistency and normality result for univariate first-order Markov chains based on a bivariate copula and for multivariate time series model under SCOMDY framework. Here, we extend the result to uDvine, which is an arbitrary-order Markov chain based on D-vine copulas.

4.3.3.1 First stage MLE

In the first stage, given the estimated marginal distribution $\hat{F}_i(\cdot)$, each $\hat{\boldsymbol{\beta}}_i$ is calculated by maximizing the log-likelihood function (4.7). Since uDvine(p) is a generalization of the bivariate copula based first-order Markov chain in Chen and Fan (2006b), it is natural to expect the theoretical properties of the first stage MLE for uDvines are the same as the ones in Chen and Fan (2006b).

Theorem 4.3.1. *Assume conditions C1-C5 in Chen and Fan (2006b) hold for the ith uDvine, we have* $\|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0\| = o_p(1)$, *i.e. the first stage MLE estimator* $\hat{\boldsymbol{\beta}}_i$ *for* $\boldsymbol{\beta}_i^0$ *is consistent.*

Before stating the result for asymptotic normality, we first introduce some notations for the ease of presentation. Denote $l_i(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i) = \log w_i(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i)$, $l_{i,\boldsymbol{\beta}}(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i) = \partial l_i(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i)/\partial\boldsymbol{\beta}_i$, $l_{i,\boldsymbol{\beta},\boldsymbol{\beta}}(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i) = \partial^2 l_i(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i)/\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i'$ and $l_{i,\boldsymbol{\beta},k}(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i) = \partial^2 l_i(u_1,\ldots,u_{p+1};\boldsymbol{\beta}_i)/\partial\boldsymbol{\beta}_i\partial u_k$, for $k = 1,2,\ldots,p+1$.

Further denote $U_{ti} = F_i^0(Y_{ti})$, $B_i = -E^0(l_{i,\boldsymbol{\beta},\boldsymbol{\beta}}(U_{ti},U_{t-1,i},\ldots,U_{t-p,i};\boldsymbol{\beta}_i^0))$ and

$$A_T^i = \frac{1}{T-p}\sum_{t=p+1}^{T}\left[l_{i,\boldsymbol{\beta}}(U_{ti},U_{t-1,i},\ldots,U_{t-p,i};\boldsymbol{\beta}_i^0) + \sum_{k=0}^{p} W_k^i(U_{t-k,i})\right],$$

where $W_k^i(x) = E^0\left(l_{i,\boldsymbol{\beta},k+1}(U_{ti},\ldots,U_{t-p,i};\boldsymbol{\beta}_i^0)(I(x \le U_{t-k,i}) - U_{t-k,i})\right)$. Define $\Sigma_i = \lim_{T\to\infty} Var^0(\sqrt{T}A_T^i)$.

Theorem 4.3.2. *Assume conditions A1-A6 in Chen and Fan (2006b) hold for the ith uDvine, we have: (1)* $\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0 = B_i^{-1}A_T^i + o_p(1/\sqrt{T})$; *(2)* $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) \to N(0, B_i^{-1}\Sigma_i B_i^{-1})$ *in distribution.*

As is mentioned in Chen and Fan (2006b), the appearance of extra $W_k^i(U_{t-k,i})$ terms in A_T^i is due to the nonparametric estimation of marginal distribution $F_i^0(\cdot)$ and if $F_i^0(\cdot)$ is known, the W_k^i terms will disappear.

4.3.3.2 Second stage MLE

In the second stage, given $\{\hat{F}_i(\cdot)\}_{i=1}^d$ and $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$, $\hat{\gamma}$ can be obtained by maximizing the log-likelihood function (4.8). Compared to the first stage MLE, the second stage MLE $\hat{\gamma}$ is obtained based on a log-likelihood function that depends on both the estimated infinite-dimensional functions $\{\hat{F}_i(\cdot)\}_{i=1}^d$ and extra finite-dimensional estimators $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$. The presence of the extra $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$ is the main difference between the setting of $\hat{\gamma}$ and the setting of the first stage estimators $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$. However, the consistency and normality result still holds, with an extra term in asymptotic covariance due to the presence of $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$.

Chen and Fan (2006a) have shown the consistency and normality of such second stage MLE under the SCOMDY framework for a different setting, where the component univariate time series are semiparametric GARCH model. Note that mDvine is also under the SCOMDY framework with semiparametric uDvine as the component univariate time series. Thus, it is not unexpected to see that the second stage MLE for mDvine is also consistent and asymptotically normal.

Theorem 4.3.3. *Under assumptions D and C stated in Chen and Fan (2006a), we have* $\|\hat{\gamma}-\gamma^0\| = o_p(1)$, *i.e. the second stage MLE estimator* $\hat{\gamma}$ *for* γ^0 *is consistent.*

Given the true marginal distributions $\{F_i^0(\cdot)\}_{i=1}^d$ and true bivariate copula parameters $\{\boldsymbol{\beta}_i^0\}_{i=1}^d$, we denote

$$F_i(Y_{ti}|\mathcal{F}_{t-1}^i) = g_i(F_i^0(Y_{ti}), F_i^0(Y_{t-1,i}), \ldots, F_i^0(Y_{t-p,i}); \boldsymbol{\beta}_i^0) = V_{ti},$$

where the first equality follows from (4.4) and $\{(V_{t1}, \ldots, V_{td})\}_{t=1}^T$ can be thought as the unobserved *i.i.d.* copula process generated by the cross-sectional copula $c_J(v_1, \ldots, v_d; \gamma^0)$. Denote $g_{i,\boldsymbol{\beta}}(u_1, \ldots, u_{p+1}; \boldsymbol{\beta}_i) = \partial g_i(u_1, \ldots, u_{p+1}; \boldsymbol{\beta}_i)/\partial \boldsymbol{\beta}_i$ and $g_{i,k}(u_1, \ldots, u_{p+1}; \boldsymbol{\beta}_i) = \partial g_i(u_1, \ldots, u_{p+1}; \boldsymbol{\beta}_i)/\partial u_k$ for $k = 1, \ldots, p+1$.

We denote $h(v_1, \ldots, v_d; \gamma) = \log c_J(v_1, \ldots, v_d; \gamma)$, $h_\gamma(v_1, \ldots, v_d; \gamma) = \partial h(v_1, \ldots, v_d; \gamma)/\partial \gamma$, $h_{\gamma,\gamma}(v_1, \ldots, v_d; \gamma) = \partial^2 h(v_1, \ldots, v_d; \gamma)/\partial \gamma \partial \gamma'$ and $h_{\gamma,i}(v_1, \ldots, v_d; \gamma) = \partial^2 h(v_1, \ldots, v_d; \gamma)/\partial \gamma \partial v_i$ for $i = 1, \ldots, d$. Denote $U_{ti} = F_i^0(Y_{ti})$ and

$$A_T^* = \frac{1}{T-p} \sum_{t=p+1}^{T} \left[h_\gamma(V_{t1}, \ldots, V_{td}; \gamma^0) + \sum_{i=1}^{d} Q_{\gamma i}(U_{ti}) \right] + \sum_{i=1}^{d} B_{\boldsymbol{\beta}}^i B_i^{-1} A_T^i,$$

where $Q_{\gamma i}(x) = E^0 \left(h_{\gamma,i}(V_{t1}, \ldots, V_{td}, \gamma^0) \sum_{k=0}^{p} g_{i,k+1}(U_{ti}, \ldots, U_{t-p,i}; \boldsymbol{\beta}_i^0)(I(x \le U_{t-k,i}) - U_{t-k,i}) \right)$,

$B^i_{\boldsymbol{\beta}} = E^0\left(h_{\gamma,i}(V_{t1},\ldots,V_{td},\gamma^0)g_{i,\boldsymbol{\beta}}(U_{ti},\ldots,U_{t-p,i};\boldsymbol{\beta}^0_i)'\right)$ and $B_i^{-1}A_T^i$ are the same as the ones in Theorem 4.3.2. We also denote $B^* = -E^0(h_{\gamma,\gamma}(V_{t1},\ldots,V_{td};\gamma_0))$ and $\Sigma^* = \lim_{T\to\infty} Var^0(\sqrt{T}A_T^*)$.

Theorem 4.3.4. *Under assumptions D and N stated in Chen and Fan (2006a), we have: (1)* $\hat{\gamma} - \gamma_i^0 = B^{*-1}A_T^* + o_p(1/\sqrt{T})$*; (2)* $\sqrt{n}(\hat{\gamma} - \gamma^0) \to N(0, B^{*-1}\Sigma^* B^{*-1})$ *in distribution.*

Notice here the asymptotic result for the second stage MLE is similar to the one for the first stage MLE. The extra $Q_{\gamma i}$ terms are introduced by the nonparametric estimation of the marginal distributions $\{F_i^0(\cdot)\}_{i=1}^d$ and the extra $B_i^{-1}A_T^i$ terms are introduced by the estimation of bivariate copula parameters $\{\boldsymbol{\beta}_i^0\}_{i=1}^d$. As is observed in Newey and McFadden (1994), the estimation of $\boldsymbol{\beta}_i^0$ does not influence the asymptotic covariance of $\hat{\gamma}$ if $B^i_{\boldsymbol{\beta}} = 0$.

There is no closed form solution for the asymptotic covariance for the first-stage and second-stage MLEs. Though the standard plug-in estimator can be constructed, it will be quite complicated to implement. A practical solution for the estimation of the asymptotic covariance can be based on a parametric bootstrap, e.g. see Zhao and Zhang (2017) for more detail.

4.4 Simulation Study

4.4.1 Performance of tree-by-tree sequential selection on uDvine

In this section, we investigate the performance of the tree-by-tree sequential selection procedure described in Section 4.3.1. Specifically, we conduct experiments on three uDvine(2)s with different parameter settings. To specify a uDvine(2), we need to specify its marginal distribution and the two bivariate copulas in tree 1 and tree 2. The marginal distributions of all uDvine(2)s are set to be $N(0,1)$.

For the first uDvine(2), we set tree 1 to be Gaussian($\rho^1 = 0.7$) copula and tree 2 to be Gumbel($\alpha^1 = 1.25$) copula. For the second uDvine(2), we set tree 1 to be $t_{\nu^2=3}(\rho^2 = 0.7)$ copula and tree 2 to be Clayton($\theta^2 = 0.5$) copula. For the third uDvine(2), we set tree 1 to be Gaussian($\rho^{31} = 0.7$) copula and tree 2 to be Gaussian($\rho^{32} = 0.3$) copula. All the parameters of the copulas are set to make the Kendall's tau of tree 1 to be 0.5 and that of tree 2 to be 0.2.

We set the candidate pool of bivariate copulas to be (Gaussian, t, Clayton, Gumbel, Frank, Joe), which are the most widely used copulas in practice. For each uDvine(2), we investigate

the performance of the sequential selection procedure under sample size of $T = 1000, 2000$ and 5000. For each sample size T, we repeat the numerical experiment 100 times. We report the percentage of correctly selected uDvine order and the percentage of correctly selected copulas for each tree of uDvine. The result is summarized in Table 4.1. As can be seen, the sequential selection procedure performs well in both order selection and copula family selection. Also, the performance is improving with the increase of sample size T.

T	order $p = 2$	tree 1 (Gaussian)	tree 2 (Gumbel)
1000	0.99	0.99	0.88
2000	0.98	0.97	0.97
5000	1.00	0.99	1.00
T	order $p = 2$	tree 1 (t_3)	tree 2 (Clayton)
1000	0.98	0.98	0.97
2000	1.00	1.00	1.00
5000	0.99	1.00	1.00
T	order $p = 2$	tree 1 (Gaussian)	tree 2 (Gaussian)
1000	0.99	0.99	0.92
2000	1.00	1.00	0.98
5000	1.00	0.99	1.00

Table 4.1: *Performance of tree-by-tree sequential selection procedure for uDvine.*

4.4.2 Performance of MLE on mDvine

In this section, we investigate the performance of the two-stage MLE on a three-dimensional mDvine consisting of the three uDvine(2) models in Section 4.4.1. To fully specify the mDvine, we set the cross-sectional copula $C_J(\cdot)$ to be Gaussian with $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.2, 0.5, 0.8)$. We assume that the parametric form (i.e. the bivariate copula types for each uDvine(2) and the cross-sectional copula type) of the mDvine is known.

We study the performance of the two-stage MLE under sample size $T = 1000, 2000$ and 5000. For each sample size T, we repeat the experiment 500 times. The result is summarized in Table 4.2. As can be seen, the two-stage MLE is consistent and the accuracy of MLE is improving with T growing.

T	$\rho^1 = 0.7$	$\alpha^1 = 1.25$	$\rho^2 = 0.7$	$\nu^2 = 3$	$\theta^2 = 0.5$
1000	0.699(0.030)	1.250(0.035)	0.694(0.034)	3.374(0.760)	0.482(0.088)
2000	0.700(0.024)	1.248(0.024)	0.700(0.022)	3.146(0.558)	0.489(0.068)
5000	0.700(0.016)	1.247(0.015)	0.699(0.016)	3.090(0.299)	0.495(0.041)
T	$\rho^{31} = 0.7$	$\rho^{32} = 0.3$	$\rho_{12} = 0.2$	$\rho_{13} = 0.5$	$\rho_{23} = 0.8$
1000	0.692(0.026)	0.300(0.032)	0.202(0.032)	0.498(0.027)	0.795(0.012)
2000	0.699(0.021)	0.296(0.019)	0.198(0.024)	0.498(0.018)	0.796(0.010)
5000	0.700(0.013)	0.301(0.012)	0.201(0.013)	0.499(0.011)	0.799(0.005)

Table 4.2: *Performance of two-stage MLE for three-dimensional mDvine. The sample standard deviations of the MLE are in brackets.*

4.5 Real Data Applications

4.5.1 Australian electricity price data

In this section, we compare the performance of mDvine with the standard vector autoregressive model (VAR) on the Australian National Electricity Market (NEM) price dataset. The NEM interconnects five regional market jurisdictions of Australia – New South Wales (NSW), Victoria (VIC), Queensland (QLD), Tasmania (TAS) and South Australia (SA). Western Australia and the Northern Territory are not connected to the NEM. A map of the relative locations of the regions can be found in Figure 4.2. Out of the five regions, NSW, VIC and QLD are the major electricity markets with average daily demands of $\mathrm{N}_d = 8235.03$, $\mathrm{V}_d = 5476.74$ and $\mathrm{Q}_d = 5913.03$ MW, while TAS and SA are significantly smaller markets with demands of $\mathrm{T}_d = 1120.56$ and $\mathrm{S}_d = 1441.43$ MW respectively.

The dataset contains six-year observations of daily maximum electricity price of the five regions from 2009-01-01 to 2015-12-31. The day of week effect is removed by a linear regression with seven dummy variables. A standard STL decomposition (see Cleveland et al. (1990) for more detail) is employed to remove the remaining trend and seasonality components of the time series. We train the mDvine and VAR using four-year data from 2009-01-01 to 2013-12-31 and hold out the rest two-year data as test set.

For all five time series, uDvine(2) is selected according to the tree-by-tree sequential selection procedure. The candidate pool for the bivariate copulas consists of 40 different bivariate copulas that are implemented in the R package VineCopula (more detail can be found in Schepsmeier et al.

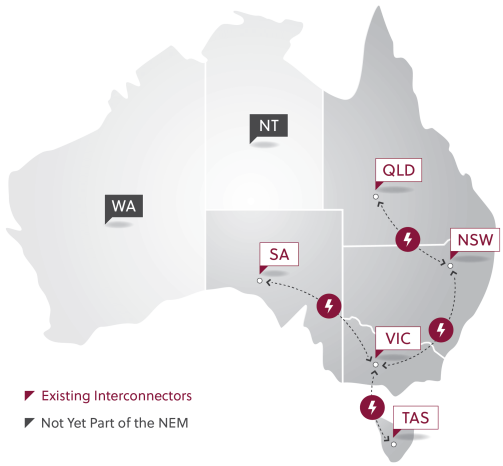

Figure 4.2: *The relative locations of the five regions in the Australian National Electricity Market.*

	NSW	VIC	QLD	TAS	SA
NSW	1	0.676	0.453	0.324	0.426
VIC		1	0.355	0.391	0.579
QLD			1	0.179	0.215
TAS				1	0.237
SA					1

Table 4.3: *Estimated correlation matrix of the cross-sectional t-copula.*

(2017)). For NSW, both tree 1 and tree 2 are selected as t-copula. For VIC, tree 1 is t-copula and tree 2 is Gumbel copula. For QLD, both tree 1 and tree 2 are selected as t-copula. For TAS, both trees are selected as BB8 copula. For SA, tree 1 is t-copula and tree 2 is Gumbel copula. The cross-sectional copula for mDvine is set to be t-copula[1], where the estimated degree of freedom is 10.62 and estimated correlation matrix is reported in Table 4.3.

As is shown in Figure 4.2, there are high voltage interconnectors between NSW and VIC, NSW and QLD, VIC and SA, and VIC and TAS. This pattern matches the estimated parameters of the cross-sectional t-copula, where the correlations of the four pairs are respectively 0.676, 0.453, 0.579 and 0.391, which are the highest among all the pair correlations. According to AIC, a VAR(1) model is selected.

We test the model performance on the one-day ahead prediction of each component univariate time series (NSW, VIC, QLD, TAS, SA) and on the one-day ahead prediction of the difference be-

[1]Both Gaussian copula and t-copula are fitted, and t-copula provides a better BIC.

	NSW	VIC	QLD	TAS	SA
Dvine	0.189	0.198	0.373	0.246	0.431
VAR	0.204	0.211	0.402	0.262	0.449
Percentage	70.36%	67.70%	66.99%	67.70%	66.57%

Table 4.4: *Mean CRPS for mDvine and VAR, and the percentage that mDvine is better than VAR on one-day ahead prediction for each component univariate time series.*

tween pairs of time series (VIC-NSW, QLD-NSW, TAS-NSW, SA-NSW, QLD-VIC, TAS-VIC, SA-VIC, TAS-QLD, SA-QLD, SA-TAS) and on the one-day ahead prediction of the demand-weighted mean of all five time series. On day t, denote the price for the five regions as NSW_t, VIC_t, QLD_t, TAS_t and SA_t, the demand-weighted mean is defined to be the demand-normalized price mean of the five regions $(\text{N}_d{\cdot}\text{NSW}_t{+}\text{V}_d{\cdot}\text{VIC}_t{+}\text{Q}_d{\cdot}\text{QLD}_t{+}\text{T}_d{\cdot}\text{TAS}_t{+}\text{S}_d{\cdot}\text{SA}_t)/(\text{N}_d{+}\text{V}_d{+}\text{Q}_d{+}\text{T}_d{+}\text{S}_d)$. The demand-weighted mean can be used as a potential price-index of the Australian National Electricity Market.

For each day in the test set, we generate the prediction based on 1000 bootstrapped sample from the trained mDvine or VAR in the training set. The detailed algorithm for generating bootstrapped samples from the fitted mDvine can be found in Section 4.7.3 in the Appendix. For the evaluation of the prediction accuracy, we mainly use two measurements, CRPS and QRPS, from Gneiting and Raftery (2007), where CRPS is a measurement for overall prediction accuracy and QRPS is a measurement for a specific quantile (e.g. 95% quantile) prediction accuracy. Smaller CRPS and QRPS indicate better prediction accuracy.

The mean CRPS for NSW, VIC, QLD, TAS and SA achieved by mDvine and VAR are reported in Table 4.4. We also report the percentage of days where CRPS of mDvine is smaller than CRPS of VAR. As can be seen, in terms of CRPS, mDvine outperforms VAR in every time series around two thirds of the time and always gives a better overall performance.

We report the CRPS of one-day ahead prediction of the difference between two time series in Table 4.5. It is clearly shown that mDvine outperforms VAR in modeling every pair of time series difference. We present the prediction result for the weighted sum of all five time series in Table 4.6. We report the CRPS and QRPS for the 95% quantile. Again, mDvine delivers a better performance than VAR in all metrics. We also provide the coverage rate of the one-day ahead 95% confidence interval and 95% Value at Risk (V@R) constructed by mDvine and VAR, along with the

CRPS	VIC-NSW	QLD-NSW	TAS-NSW	SA-NSW	QLD-VIC
Dvine	0.168	0.363	0.277	0.434	0.396
VAR	0.215	0.405	0.306	0.445	0.428
Percentage	78.37%	70.51%	72.05%	66.29%	69.94%
CRPS	TAS-VIC	SA-VIC	TAS-QLD	SA-QLD	SA-TAS
Dvine	0.242	0.375	0.445	0.602	0.455
VAR	0.281	0.401	0.481	0.619	0.477
Percentage	75.14%	70.37%	69.80%	61.94%	66.85%

Table 4.5: *Mean CRPS for mDvine and VAR, and the percentage that mDvine is better than VAR on one-day ahead prediction of the difference between two time series.*

	CRPS	QRPS	V@R 95%	C.I. 95%
Dvine	0.184	0.049	93.68% (0.121)	93.96% (0.197)
VAR	0.191	0.051	89.75% (0)	86.66% (0)

Table 4.6: *Mean CRPS/QRPS for mDvine and VAR, and the coverage rate of 95% Value at Risk and 95% confidence interval on one-day ahead prediction for the weighted sum of five time series. The p-value of the corresponding binomial test is reported in the brackets.*

corresponding p-values for the binomial test. As can be seen, mDvine gives a coverage rate that is very close to the target rate (95%) and passes both the binomial test for C.I. and V@R while VAR does not provide a satisfactory performance.

4.5.2 Ireland spatial-temporal wind data

In this section, we investigate the performance of mDvine in the context of spatial-temporal data. The data set contains the daily average wind speed of 12 stations across Ireland from 1961-01-01 to 1966-12-31. The locations of the 12 stations are plotted in Figure 4.3.

The twelve series are adjusted to remove the day-to-day effect, see Pebesma (2004) for more detail. The longitude and latitude of each station is known, thus we can calculate a distance matrix among all the 12 stations, which is reported in Table 4.9 in Section 4.7.4 in the Appendix. We use the four-year data from 1961-01-01 to 1964-12-31 as the training set and the rest of the data as the test set. For comparison, we also fit a spatial Gaussian model. The spatial Gaussian model can be seen as a special case of the mDvine, where all the marginal distributions are set to be Gaussian distributions, and all the bivariate copulas and the cross-sectional copula are set to be Gaussian copulas. See Erhardt et al. (2015) for more detail of the spatial Gaussian model. Both the mDvine and spatial Gaussian model are fitted on the 12 stations.

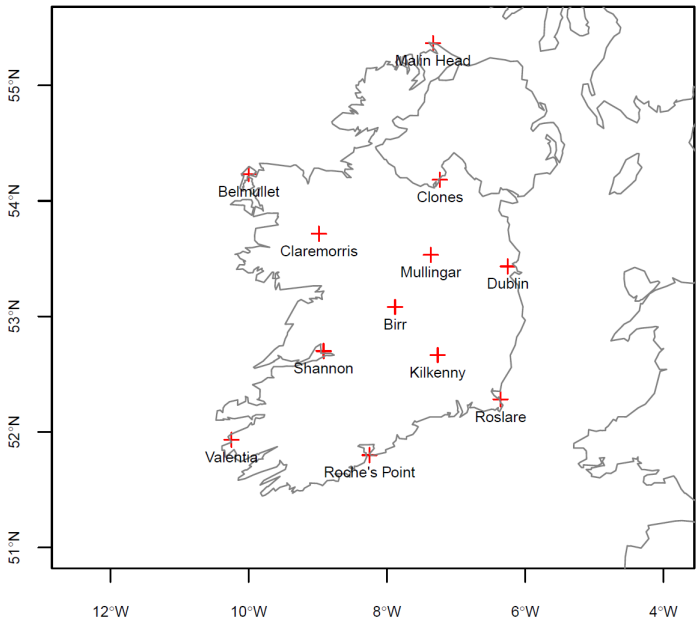

Figure 4.3: *The relative locations of Australian National Electricity Market.*

To demonstrate the usefulness of spatial correlation and mDvine's ability of accurately capturing the spatial correlation, we perform two types of out-of-sample validation. We first randomly pick up four stations (Claremorris, Dublin, Valentia, Shannon) as the test stations. For type I validation, we only use the selected four stations' own history to compute their one-day ahead prediction. For type II validation, we use the four stations' own history and the other eight stations' contemporaneous observations to perform the one-day ahead prediction for the selected four stations. The type II validation is inspired by a similar validation approach in Erhardt et al. (2015) for a spatial-temporal data of daily mean temperatures for different locations in Germany. As before, for each day in the test set, we generate the prediction based on 1000 bootstrapped sample from the trained mDvine or spatial Gaussian model in the training set. The detailed algorithm for generating bootstrapped samples from the fitted mDvine can be found in Section 4.7.3 in the Appendix.

For mDvine, by the sequential selection procedure, uDvine(1) has been selected for 10 stations and uDvine(2) has been selected for 2 stations. We use the Gaussian copula as the cross-sectional copula, where the correlation structure is specified by the Matérn covariance class based on the distance matrix in Table 4.9. For the spatial Gaussian model, the AR(p) model is used for modeling the marginal time series where p is selected by AIC for each component univariate time series. A multivariate Gaussian distribution is used to model the cross-sectional error terms across the com-

Type I	Claremorris	Dublin	Valentia	Shannon
Dvine	1.206	1.242	1.369	1.308
AR	1.208	1.250	1.370	1.317
Percentage	50.90%	53.66%	49.93%	51.72%
Type II	Claremorris	Dublin	Valentia	Shannon
Dvine	0.448	0.563	0.739	0.541
AR	0.839	0.918	0.968	0.925
Percentage	73.38%	70.35%	58.34%	73.10%

Table 4.7: *CRPS of one-day ahead predictions for four stations, type I and type II.*

	CRPS	QRPS	V@R 95%	C.I. 95%
Dvine	0.615	0.292	93.93% (0.20)	93.24% (0.033)
AR	1.398	0.695	77.79% (0)	66.21% (0)

Table 4.8: *CRPS and QRPS for the mean wind speed of four stations, type II*

ponent univariate AR(p)s, where the covariance structure is also specified by the Matérn covariance class.

For type I validation, we report the CRPS for each of the four stations in the upper section of Table 4.7. For type II validation, we report the CRPS in the lower section of Table 4.7. Comparing the upper and lower section, we can see that the prediction accuracy of both spatial Gaussian model and mDvine are improved after including the contemporaneous spatial observations of the other eight stations. However, mDvine utilizes the spatial information more efficiently than spatial Gaussian model, since the improvement of mDvine from Type I to Type II is much more significant than that of spatial Gaussian model. This indicates mDvine's promising ability in the accurate modeling of both the marginal time series behavior of each station and the joint spatial behavior among all stations.

For type II validation, we also report the CRPS and QRPS for the 95% quantile for the mean wind speed of the four stations in Table 4.8. Again, mDvine delivers a better performance than spatial Gaussian in all metrics. We also provide the coverage rate of the one-day ahead 95% confidence interval and 95% Value at Risk (V@R), along with the corresponding p-values for the binomial test. Again, mDvine gives a coverage rate that is very close to the target rate (95%) and passes both the binomial test for C.I. and V@R while VAR does not provide a satisfactory performance.

4.6 Conclusion and Future Research

In this paper, we have proposed and studied the semiparametric mDvine, which provide a novel multivariate time series model that enables the simultaneous copula-based modeling for temporal and cross-sectional dependence of a multivariate time series. The semiparametric uDvine extends the first-order copula-based Markov chain to an arbitrary-order Markov chain. By the semiparametric assumption, the uDvine provides flexible models for marginal behavior of the time series. Due to the use of D-vine copula, the uDvine is capable of generating sophisticated temporal dependence structures such as nonlinear dependence, asymmetric dependence and tail dependence. Using the same idea of SCOMDY framework, we design the mDvine, which offers a tractable and flexible model for multivariate time series. A two-stage MLE has been proposed for the estimation of the mDvine. The consistency and asymptotic normality of the MLE have been established and affirmed by extensive numerical experiments. Real data applications on the Australian electricity market and the Ireland wind speeds have demonstrated mDvine's promising ability for modeling multivariate time series, where significant improvement over conventional time series models has been observed.

There are several interesting extensions of the mDvine. A natural extension is to use time-varying instead of constant cross-sectional copulas. For example, a certain autoregressive evolution scheme can be designed for the parameters of copula C_J such that the conditional cross-sectional dependence is related to past information.

4.7 Appendix

4.7.1 Proof on stationarity and ergodicity of uDvine

In this section, $\{Y_t\}_{t=1}^T$ denotes a univariate time series following uDvine(p) satisfying homogeneity condition.

Proof of Theorem 4.2.1: We first prove that $\{Y_t\}_{t=1}^T$ is a p-order Markov chain. By Aas et al.

(2009) we have

$$f(y_t|y_{t-1},\ldots,y_1) = f(y_t)\prod_{s=1}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1});\beta_{s,t}),$$

where $c_{s,t}(\cdot)$ denotes the bivariate copula in tree $t-s$. By the definition of uDvine(p), all bivariate copulas $c_{s,t}(\cdot)$ beyond pth tree are independent copulas. Thus, we have for $t > p$

$$f(y_t|y_{t-1},\ldots,y_1) = f(y_t)\prod_{s=t-p}^{t-1} c_{s,t}(F_{s|(s+1):(t-1)}(y_s|y_{s+1},\ldots,y_{t-1}), F_{t|(s+1):(t-1)}(y_t|y_{s+1},\ldots,y_{t-1});\beta_{s,t}),$$

i.e. the conditional distribution of Y_t given the history only depends on the past p observations $(Y_{t-1},\ldots,Y_{t-p})$, thus $\{Y_t\}$ is a p-order Markov chain.

The homogeneity of the p-order Markov chain $\{Y_t\}$ follows from the homogeneity of uDvine(p), i.e. uDvine(p) has the same copula in each tree. □

Since the marginal distribution $F(\cdot)$ of uDvine is absolutely continuous, we know that the probability integral transformed process $\{F(Y_t)\}_{t=1}^T$ marginally follows a uniform distribution on (0,1). As is noted by Chen and Fan (2006b), the stationary and ergodic property of uDvine $\{Y_t\}$ is the same as the one of $\{F(Y_t)\}$ due to the absolute continuity of $F(\cdot)$.

In the following, we define $X_t = (F(Y_t), F(Y_{t-1}),\ldots,F(Y_{t-p+1}))$. Note that by Proposition 4.2.1, $\{X_t\}$ is a first-order homogeneous Markov chain in $(0,1)^p$. The result of Theorem 4.2.1 depends on two assumptions stated below.

Assumption 4.7.1. *There exists a $k \in \mathbb{N}^+$ such that the pdf of k-step transition probability $p(X_t = x, X_{t+k} = y) > 0$, for all $x, y \in (0,1)^p$.*

Assumption 4.7.2. *Denote the m-step transition probability of $\{X_t\}$ as $P^m(x,y)$. There exists a function $V(x) = 1 + W(x)$, where $W : x \in (0,1)^p \to [0,\infty)$ such that $W(x) \to \infty$ as $\min(x) \wedge (1 - \max(x)) \to 0$ (i.e. as x approaches the boundary of $(0,1)^p$). For some $m \in \mathbb{N}^+$, $\int_{(0,1)^p} W(y)P^m(x,dy)$ is bounded on compact sets of $(0,1)^p$ and that*

$$\limsup_{\min(x)\wedge(1-\max(x))\to 0} \frac{\int_{(0,1)^p} W(y)P^m(x,dy)}{W(x)} < 1.$$

Remark: For the verification of Assumption 4.7.1, a possible choice of k can be $p+1$, where $X_t = (F(Y_t), F(Y_{t-1}), \ldots, F(Y_{t-p+1}))$ and $X_{t+p+1} = (F(Y_{t+p+1}), F(Y_{t+p}), \ldots, F(Y_{t+1}))$. For Assumption 4.7.2, possible choices of the drift function $V(x)$ can be $1+|\log(\min(x))|^\gamma+|\log(1-\max(x))|^\gamma$ or $1+\frac{1}{\min(x)^\gamma(1-\max(x))^\gamma}$, where $\gamma > 0$.

Proof of Theorem 4.2.1: To facilitate the proof, we first give Lemma 4.7.1. Denote the p-dimensional copula of $(Y_t, Y_{t-1}, \ldots, Y_{t-p+1})$ generated by uDvine(p), i.e. the joint distribution of $(F(Y_t), \ldots, F(Y_{t-p+1}))$, as $\mathbf{C}_p(\cdot)$, by the homogeneity condition of uDvine, we know that $\mathbf{C}_p(\cdot)$ does not depend t. We have

Lemma 4.7.1. *$\mathbf{C}_p(\cdot)$ is an invariant probability measure for $\{X_t\}$.*

Proof. We need to prove $P(X_t \in A) = \mathbf{C}_p(A)$ for all $t \geq 1$ if X_1 follows $\mathbf{C}_p$. We only need to prove the equality for the rectangle sets $A = A_1 \times \cdots \times A_p$, where $A_i \in \mathcal{B}$ and $\mathcal{B}$ is the Borel set of $\mathbb{R}$. The proof follows by induction and the homogeneity property of uDvine(p). In the following, $f(\cdot)$ denotes a generic joint or conditional pdf. Denote $U_t = F(Y_t)$, we have $X_t = (U_t, \ldots, U_{t-p+1})$. Assume that X_{t-1} follows $\mathbf{C}_p(\cdot)$,

$$
\begin{aligned}
P(X_t \in A) &= E(P(X_t \in A | X_{t-1})) \\
&= \int_{A_2 \times \cdots \times A_p \times \mathbb{R}} P(u_t \in A_1 | u_{t-1}, \ldots, u_{t-p}) c_p(u_{t-1}, \ldots, u_{t-p}) du_{t-p} \ldots du_{t-1} \\
&= \int_{A_1 \times A_2 \times \cdots \times A_p \times \mathbb{R}} f(u_t | u_{t-1}, \ldots, u_{t-p}) \cdot c_p(y_{t-1}, \ldots, y_{t-p}) dy_{t-p} \ldots dy_{t-1} dy_t \\
&= \int_{A_1 \times A_2 \times \cdots \times A_p \times \mathbb{R}} f(u_t, u_{t-1}, \ldots, u_{t-p}) / f(u_{t-1}, \ldots, u_{t-p}) \cdot c_p(u_{t-1}, \ldots, u_{t-p}) du_{t-p} \ldots du_{t-1} du_t \\
&= \int_{A_1 \times A_2 \times \cdots \times A_p \times \mathbb{R}} f(u_t, u_{t-1}, \ldots, u_{t-p}) du_{t-p} \ldots du_{t-1} du_t \\
&= \int_{A_1 \times A_2 \times \cdots \times A_p} c_p(u_t, u_{t-1}, \ldots, u_{t-p+1}) du_{t-p+1} \ldots du_{t-1} du_t = \mathbf{C}_p(A),
\end{aligned}
\tag{4.9}
$$

where second equality comes from the fact that X_{t-1} follows $\mathbf{C}_p(\cdot)$, the fourth and the second to last equality follow from the homogeneity of uDvine(p), i.e. the joint distribution of $(U_t, \ldots, U_{t-p+1})$ is always $\mathbf{C}_p(\cdot)$ regardless of t. □

Based on Theorem 4.2.1 and Lemma 4.7.1, $\{X_t\}$ is a well-behaved first-order homogeneous Markov chain with state space $(0,1)^p$ and has $\mathbf{C}_p(\cdot)$ as its invariant probability measure. By Assumption 4.7.1, Theorem 2.1.1 and Theorem 1.2.2 in Hansen (2000), it is easy to prove that $\{X_t\}$ is an λ-irreducible, aperiodic, positive recurrent Markov chain, where λ is the Lebesgue measure on $\mathbb{R}^p$.

By Assumption 4.7.2, there exists an m such that

$$\alpha = \limsup_{\min(x)\wedge(1-\max(x))\to 0} \frac{\int_{(0,1)^p} W(y)P^m(x,dy)}{W(x)} < 1,$$

thus we can find a $\beta < 1$ such that $\alpha < \beta$ and a suitable compact set K such that

$$\begin{aligned}\int_{(0,1)^p} V(y)P^m(x,dy) &= 1 - \frac{\int_{(0,1)^p} W(y)P^m(x,dy)}{W(x)} + \frac{\int_{(0,1)^p} W(y)P^m(x,dy)}{W(x)}V(x) \\ &\leq \beta V(x) + b1_K(x),\end{aligned}$$

where $b = \sup_{x\in K} \int_{(0,1)^p} V(y)P^m(x,dy)$ and the last inequality follows from the fact that $\beta > \alpha$ and $V(x) \to \infty$ as x approaches the boundary of $(0,1)^p$. The rest of the proof follows the same logic as Theorem 2.1.6 in Hansen (2000). □

Proof of Corollary 4.2.1: The corollary follows directly from Theorem 3.7 in Bradley (2005). □

4.7.2 Proof on consistency and asymptotic normality of MLE

Proof of Theorem 4.3.1: The proof follows the same line as the proof of proposition 4.2 in Chen and Fan (2006b) and therefore is omitted. □

Proof of Theorem 4.3.2: The proof follows the same line as the proof of proposition 4.3 in Chen and Fan (2006b) and therefore is omitted. □

Proof of Theorem 4.3.3: The proof follows the same line as the proof of proposition 3.1 in Chen and Fan (2006a) and therefore is omitted. □

Proof of Theorem 4.3.4: The proof follows the same line as the proof of proposition 3.2 in Chen and Fan (2006a). An additional Taylor expansion of the score function $\frac{\partial L_2(\gamma)}{\partial \gamma}$ of the 2nd

stage log-likelihood function $L_2(\gamma)$ in (4.8) w.r.t. $\{\boldsymbol{\beta}_k\}_{k=1}^d$ around the true value $\{\boldsymbol{\beta}_k^0\}_{k=1}^d$ and the influence function representation of $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0 = B_k^{-1} A_T^k + o_p(1/\sqrt{T})$ as stated in Theorem 4.3.2 are used to obtain the final result. □

4.7.3 Algorithm for generating predictions from mDvine

Based on the training set, the two-stage MLE is used to estimate all the marginal distributions $\{\hat{F}_i(\cdot)\}_{i=1}^d$, the bivariate copula parameters $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^d$ and the cross-sectional copula parameter $\hat{\gamma}$.

Algorithm A: Assume that we are on day $t-1$, based on the history $\mathcal{F}_{t-1}$, we want to make predictions for day t. The prediction is based on 1000 bootstrapped sample. For b from 1 to 1000, we repeat

- Step 1: Simulate $(V_{b,1}, \ldots, V_{b,d})$ from copula $C_J(\hat{\gamma})$.
- Step 2: Generate one-day ahead sample $(Y_{b,1}, \ldots, Y_{b,d})$ by $Y_{b,i} = \hat{F}_i^{-1}(V_{b,i}|\mathcal{F}_{t-1}^i)$, where $\hat{F}_i(y_{ti}|\mathcal{F}_{t-1}^i) = g_i(\hat{F}_i(y_{ti}), \hat{F}_i(y_{t-1,i}), \ldots, \hat{F}_i(y_{t-p_i,i}); \hat{\boldsymbol{\beta}}_i)$, for $i = 1, \ldots, d$.

Algorithm B: Assume that we are on day t, based on the complete history $\mathcal{F}_{t-1}$ and the contemporaneous observations of first $d_1 < d$ component univariate time series $(y_{t1}, y_{t2}, \ldots, y_{td_1})$, we want to make predictions for the other component time series on day t. The prediction is based on 1000 bootstrapped sample. For b from 1 to 1000, we repeat

- Step 1: Simulate $(V_{b,d_1+1}, \ldots, V_{b,d})$ from the conditional distribution of $(V_{d_1+1}, \ldots, V_d)$ given $(\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_{d_1})$, where $\hat{V}_i = \hat{F}_i(y_{ti}|\mathcal{F}_{t-1}^i) = g_i(\hat{F}_i(y_{ti}), \hat{F}_i(y_{t-1,i}), \ldots, \hat{F}_i(y_{t-p_i,i}); \hat{\boldsymbol{\beta}}_i)$ for $i = 1, \ldots, d_1$.
- Step 2: Generate one-day ahead sample $(Y_{b,d_1+1}, \ldots, Y_{b,d})$ by $Y_{b,i} = \hat{F}_i^{-1}(V_{b,i}|\mathcal{F}_{t-1}^i)$, where $\hat{F}_i(y_{ti}|\mathcal{F}_{t-1}^i) = g_i(\hat{F}_i(y_{ti}), \hat{F}_i(y_{t-1,i}), \ldots, \hat{F}_i(y_{t-p_i,i}); \hat{\boldsymbol{\beta}}_i)$ for $i = d_1+1, \ldots, d$.

⋆ Note that the conditional distribution of $(V_{d_1+1}, \ldots, V_d)$ given $(V_1, \ldots, V_{d_1})$ needs to be derived from the cross-sectional copula $C_J(\cdot)$. For copulas such as elliptical copulas, a closed form solution for the conditional distribution is available.

4.7.4 Distance matrix for Ireland wind data

	Roche's Point	Valentia	Roslare	Kilkenny	Shannon	Birr
Roche's Point	0.00	138.56	140.54	117.53	110.01	144.96
Valentia	138.56	0.00	269.50	219.25	124.67	205.40
Roslare	140.54	269.50	0.00	75.16	179.96	136.37
Kilkenny	117.53	219.25	75.16	0.00	111.64	62.24
Shannon	110.01	124.67	179.96	111.64	0.00	81.59
Birr	144.96	205.40	136.37	62.24	81.59	0.00
Dublin	226.66	317.79	128.28	109.21	196.49	115.75
Claremorris	218.96	216.03	237.81	163.76	113.23	101.59
Mullingar	201.93	263.85	154.89	96.68	139.18	60.78
Clones	273.87	321.82	219.49	168.81	199.40	129.76
Belmullet	295.15	256.51	325.88	251.72	185.19	189.63
Malin Head	401.57	427.95	349.24	300.56	314.40	256.67
	Dublin	Claremorris	Mullingar	Clones	Belmullet	Malin Head
Roche's Point	226.66	218.96	201.93	273.87	295.15	401.57
Valentia	317.79	216.03	263.85	321.82	256.51	427.95
Roslare	128.28	237.81	154.89	219.49	325.88	349.24
Kilkenny	109.21	163.76	96.68	168.81	251.72	300.56
Shannon	196.49	113.23	139.18	199.40	185.19	314.40
Birr	115.75	101.59	60.78	129.76	189.63	256.67
Dublin	0.00	183.77	74.96	105.66	262.42	226.41
Claremorris	183.77	0.00	108.89	126.09	88.07	212.44
Mullingar	74.96	108.89	0.00	72.88	189.87	204.09
Clones	105.66	126.09	72.88	0.00	180.59	131.88
Belmullet	262.42	88.07	189.87	180.59	0.00	212.88
Malin Head	226.41	212.44	204.09	131.88	212.88	0.00

Table 4.9: *Distance matrix of the 12 stations (in kilometer).*

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Almeida, C., Czado, C., and Manner, H. (2012). Modeling high dimensional time-varying dependence using d-vine scar models. *Working paper*, page http://arxiv.org/pdf/1202.2008v1.pdf.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bali, T. and Weinbaum, D. (2007). A conditional extreme value volatility estimator based on high-frequency returns. *Journal of Economic Dynamics and Control*, 31(2):361–397.

Balkema, A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2:792–804.

Basrak, B., Davis, R. A., and Mikosch, T. (2002). A characterization of multivariate regular variation. *Annals of Applied Probability*, 12:908–920.

Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080.

Bücher, A. and Segers, J. (2016). On the maximum likelihood estimator for the generalized extreme-value distribution. *arXiv*.

Bücher, A. and Segers, J. (2017). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli*.

Beare, B. (2010). Copulas and temporal dependence. *Econometrica*, 78(1):395–410.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proc. Amer. Math. Soc.*, 12:788–792.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Boutahar, M., Dufrénot, G., and Péguin-Feissolle, A. (2008). A simple fractionally integrated model with a time-varying long memory parameter d_t. *Computational Economics*, 31:225–241.

Bradley, R. (2005). Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys*, 2:107–144.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Cai, Z. and Wang, X. (2014). Selection of mixed copula model via penalized likelihood. *Journal of American Statistical Association*, 109(506):788–801.

Chan, K. and Tong, H. (1994). A note on noisy chaos. *Journal of Royal Statistical Society - Series B*, 56(2):301–311.

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 188(1):44–52.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1–2):125–154.

Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley.

Cleveland, R., Cleveland, W., McRae, J., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer.

Coval, J., Jurek, J., and Stafford, E. (2009). The economics of structured finance. *Journal of Economic Perspectives*, 23(1):3–25.

Creal, D., Koopman, S., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Cui, Q. and Zhang, Z. (2016). Max-linear competing factor models. *Journal of Business and Economic Statistics*, page Forthcoming.

Darsow, W., Nguyen, B., and Olsen, E. (1992). Copulas and markov processes. *Illinois Journal of Mathematics*, 36:600–642.

Davison, A. and Smith, R. (1990). Models for exceedances over high thresholds. *Journal of Royal Statistical Society - Series B*, 52(3):393–442.

Dias, A. and Embrechts, P. (2010). Modeling exchange rate dependence dynamics at different time horizons. *Journal of International Money and Finance*, 29:1687–1705.

Diebold, F., Schuermann, T., and Stroughair, J. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In Refenes, A., Moody, J., and Burgess, A., editors, *Decision Technologies for Computational Finance*, volume 2 of *Advances in Computational Management Science*, chapter 1, pages 3–12. Springer US.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1):420–436.

Domma, F., Giordano, S., and Perri, P. F. (2009). Statistical modeling of temporal dependence in financial data via a copula function. *Communications in Statistics - Simulation and Computation*, 38(4):703–728.

Drees, H., Segers, J., and Warchol, M. (2015). Statistics for tail processes of Markov chains. *Extremes*, 18(3).

Durante, F. and Sempi, C. (2015). *Principles of copula theory*. CRC Press.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3:30–41.

Engle, R. and Russell, J. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66:1127–1162.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50:987–1007.

Erhardt, T., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71:323–332.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, pages 3–56.

Francq, C. and Zakoian, J. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637.

Frees, E. and Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Gao, X. and Song, P. (2010). Composite likelihood Bayesian Information Criteria for model selection in high-dimensional data. *Journal of American Statistical Association*, 105(492):1531–1540.

Genest, C. and Favre, A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368.

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C., Remillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In Aigner, D. and Goldberger, A., editors, *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of American Statistical Association*, 102(477):359–378.

Haff, I. H. (2013). Parameter estimation for pair-copula constructions. *Bernoulli*, 19(2):462–491.

Hall, S., Swamy, P., and Tavlas, G. (2016). Time-varying coefficient models: A proposal for selecting the coefficient driver sets. *Macroeconomic Dynamics*, pages 1–17.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730.

Hansen, N. (2000). Convergence speed of Markov chains with emphasis on discrete approximations of the Langevin diffusion and the Metropolis-Hastings algorithm. *Master's Thesis*, pages 1–102.

Harris, R. and Mazibas, M. (2013). Dynamic hedge fund portfolio construction: A semi-parametric approach. *Journal of Banking & Finance*, 37(1):139–149.

Harvey, A. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, Cambridge, UK.

Harvey, A. and Chakravarty, T. (2008). Beta-t-(E)GARCH. *Discussion Paper, University of Cambridge*.

Hull, J. and White, A. (2004). Valuation of a cdo and an nth to default cds without monte carlo simulation. *Journal of Derivatives*, (12):8–23.

Ibragimov, R. (2009). Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(3):819–846.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC.

Kelly, B. (2014). The dynamic power law model. *Extremes*, 17(4):557–583.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *The Review of Financial Studies*, 27(10):2841–2871.

Khoudraji, A. (1995). *Contributions a l' etude des copules et a la modelisation des valeurs extremes bivariees*. PhD thesis, Universite Laval Quebec.

Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Laurini, F. and Tawn, J. A. (2009). Regular variation and extremal dependence of garch residuals with application to market risk measures. *Econometric Reviews*, 28:146–169.

Leadbetter, M., Lindgren, G., and Rootzen, H. (1983). *Extremes and related properties of random sequences and processes*. Springer.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99:2234–2250.

Lindsay, B. (1988). Composite likelihood methods. *Comtemporary Mathematics*, (80):221–239.

Lundbergh, S., Terasvirta, T., and Van Dijk, D. (2003). Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics*, 21(1):104–128.

Makelainen, T., Schmidt, K., and Styan, G. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Annals of Statistics*, 9:758–767.

Malinowski, A., Schlather, M., and Zhang, Z. (2015). Marked point process adjusted tail dependence analysis for high frequency financial data. *Statistics and Its Interface*, 8:109–122.

Massacci, D. (2016). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science*, pages 1–18.

McNeil, A. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4):271 – 300.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Min, A. and Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783.

Murray, J., Dunson, D., Lawrence, C., and Lucas, J. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108:656–665.

Nelsen, R. (1999). *An Introduction to Copulas*. Springer.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *The Handbook of Econometrics*, volume 4, chapter 36. North Holland.

Oh, D. and Patton, A. (2013). Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502):689–700.

Oh, D. and Patton, A. (2015). Modelling dependence in high dimensions with factor copulas. *Journal of Business and Economic Statistics*, page Forthcoming.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. (2012). Copula methods for forecasting multivariate time series. In *Handbook of economic forecasting*. Springer Verlag.

Pebesma, E. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7):683–691.

Picklands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer.

Salmon, F. (2012). The formula that killed wall street. *Significance*, 9(1):16–20.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T. (2017). Vinecopula: Statistical inference of vine copulas. https://CRAN.R-project.org/package=VineCopula. R package version 2.1.3.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Shi, P. and Yang, L. (2017). Pair copula constructions for insurance experience rating. *Journal of American Statistical Association*.

Sibuya, M. (1959). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, (8):229–231.

Smith, M. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, 31:815–833.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.

Smith, R. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. and Goodman, D. (2000). Bayesian risk analysis, extremes and integrated risk management. *Risk Books*.

Stock, J. and Watson, M. (2011). Dynamic factor models. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

van Oordt, M. and Zhou, C. (2016). Systematic tail risk. *Journal of Financial and Quantitative Analysis*, 51(2):685–705.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, (21):5–42.

Zhang, X. and Schwaab, B. (2017). Tail risk in government bond markets and ECB asset purchases. *Working paper*.

Zhao, Z. and Zhang, Z. (2017). Semi-parametric dynamic max-copula model for multivariate time series. *Journal of Royal Statistical Society - Series B*, Accepted.