UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Econometric Analysis of Interval-Valued Time Series

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Yun Luo

June 2019

Dissertation Committee:

Dr. Gloria González-Rivera, Chairperson
Dr. Tae-Hwy Lee
Dr. Aman Ullah

The Dissertation of Yun Luo is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First of all, I would like to sincerely thank my advisor Professor Gloria González-Rivera for her guidance, encouragement, patience, and understanding. Her constant guidance and encouragement helped me gain a deeper understanding of econometrics and develop this dissertation. More importantly, she shaped the way I think in many perspectives towards being a better researcher. I still remember her advice of building toy examples to effectively identify the mechanisms in the complex research problems. Her patience and understanding also allowed me to do the research at my own pace, which made me more concentrated and productive. I would not have been here without her help and support.

I am also indebted to the other committee members: Professor Tae-Hwy Lee and Professor Aman Ullah, for their insightful advice and valuable inspirations at various stages of my PhD studies. I would also like to thank Professor Ruoyao Shi, who gave me very helpful suggestions in preparing for the job market; Professor Joseph Cummins, who helped broaden my interests into microeconometrics and were always there to discuss research; Professor Bree Lang and Professor Matthew Lang, for their useful comments in my job market paper from the applied economists' perspectives. My thanks also go to all the professors who have taught me at University of California, Riverside. A special thanks to Gary Kuzas for his kindly support.

I am thankful to the fellow graduate students in my cohort, who made my five years at UCR one of the most memorable time in my life. Particularly, I want to express my thanks to Jianghao Chu for being a good roomate and a supportive friend. Many inspirational discussions we had at lunchs and dinners helped me develop creative ideas.

My deepest gratitude goes to my grandparents and parents for their love, care, and firmly support. They always stand by my side during the good and bad times. Finally, I would like to thank my beloved wife for her continuous support of my PhD education, for her sacrifice in the stressful long-distance relationship, and for her accompany and encouragement during many dark days and nights.

This dissertation is dedicated to my parents and my wife for all their support.

ABSTRACT OF THE DISSERTATION

The Econometric Analysis of Interval-Valued Time Series

by

Yun Luo

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2019
Dr. Gloria González-Rivera, Chairperson

This dissertation covers three topics in modeling and forecasting interval-valued time series.

In Chapter 1, we propose a model for interval-valued time series (ITS) that aims to generate valid point-valued forecasts. We dispense with the positive constraint on the range by estimating a bivariate system of the center/log-range. However, a forecast based on this system needs to be transformed to the original units of center/range, which requires bias correction. We examine the out-of-sample forecast performance of naive transformed forecasts (biased), parametric bias-corrected forecasts, and semiparametric correction methods like smearing correction and bootstrap forecasts. Monte Carlo simulations show that the biased correction methods do not generate forecasts that are uniformly superior. We apply these methods to the daily low/high intervals of the SP500 index and Google prices.

In Chapter 2, we go beyond point forecasts to construct the probabilistic forecasts for interval-valued time series. We estimate a bivariate system of the center/log-range, which may not be normally distributed. Implementing analytical or bootstrap methods, we

directly transform prediction regions for center/log-range into those for center/range and upper/lower bounds systems. We propose new metrics to evaluate the regions performance. Monte Carlo simulations show bootstrap methods being preferred even in Gaussian systems. We build prediction regions for daily SP500 low/high return intervals, and apply them to develop a trading strategy.

In Chapter 3, we develop an alternative model directly on the ITS (upper/lower bounds system). The model specifies the conditional joint distribution of the upper and lower bounds of the interval to be a mixture of truncated bivariate normal distribution. This specification guarantees that the natural order of the interval (upper bound not smaller than lower bound) is preserved. The model also captures the potential conditional heteroscedasticity and non-Gaussian features in ITS. We propose an EM algorithm for model estimation. We establish the consistency of the maximum likelihood estimator. Monte Carlo simulations show the new EM algorithm has good convergence properties. We apply the model to the interval-valued IBM daily stock returns and it exhibits superior performance over competing methods.

Contents

List of Figures

List of Tables

Chapter 1

Point Forecast for Interval-valued Time Series

1.1 Introduction

Data sets in interval format are common in many disciplines. See Blanco-Fernández and Winker (2016) for different data generation mechanisms of interval data. In economics, we have many examples. For instance, in stock markets, it is standard to provide the daily interval of low/high asset prices. In bond markets, traders report bid/ask intervals. In energy markets, the US Energy Information Administration provides min/max retail prices of electricity at the state level. The US Department of Agriculture also provides daily low and high prices on agricultural commodities and livestock. In earth sciences, temperatures are also recorded in the min/max format. It should be noted that in many instances, the interval format is the only available format to the researcher. There are several reasons to

prefer interval records. For instance, in stock markets it is customary to report the daily closing price, which is just one-point measurement, while we observe plenty of price points over the trading day. Other records, like an average temperature or like those provided by USDA as daily weighted average prices on commodities and livestock, are not very informative to market participants. In other instances in which the data is sensitive to privacy concerns such as income reporting, the records must be aggregated, e.g. income intervals.

Our interest is in interval-valued times series defined as a collection of interval realizations ordered over time, i.e., $\{(y_{l,t}, y_{u,t})\}$ for $t = 1, ...T$, where $y_{l,t}$ is the lower bound and $y_{u,t}$ is the upper bound of the interval at time t, such that $y_{l,t} \leq y_{u,t}$ for all t. An equivalent representation is given by considering the center of the interval $C_t = (y_{l,t}+y_{u,t})/2$ and the range $R_t = y_{u,t} - y_{l,t} \geq 0$, i.e., $\{(C_t, R_t)\}$ for $t = 1, ...T$. Most of the econometric analysis in this area has focused on model estimation and inference, and though it is possible to construct point forecasts based on a given model or algorithm.

When dealing with lower/upper bounds systems, one needs to incorporate the constraint $y_{l,t} \leq y_{u,t}$ into the estimation. González-Rivera and Lin (2013) propose a two-step estimator (Maximum Likelihood (ML) and Least Squares (LS)) and a modified two-step estimator (ML and Minimum Distance) based on assuming a truncated bivariate normal density of the errors of the lower/upper bounds system. The estimation of the system is complex but it is possible to construct a direct bivariate density forecast for the upper/lower bounds, if the truncated bivariate normal density is the right assumption. Alternatively, dealing with the center/range system, one needs to incorporate the constraint $R_t \geq 0$. Lima Neto and De Carvalho (2010) impose non-negative constraints on the parameters of

the range equation, which are unnecessarily too restrictive and complicate the estimation of the system. Tu and Wang (2016) overcome the restriction $R_t \geq 0$ by log-transforming the range, and estimating the center/log-range system without imposing any distributional assumptions. However, forecasting the center/range or lower/upper bounds will be more complicated. First, for point forecasts, one needs the inverse transformation, i.e. $R_t = \exp[\log R_t]$, which itself introduces non-trivial econometric issues. Secondly, for a density forecast, a joint distributional assumption for the center and range or for the upper and lower bounds is required.

We will start with the estimation of a dynamic model for the system center/log-range. We specify a VAR system to be estimated by quasi-maximum likelihood, assuming a bivariate Gaussian density, that guarantees the consistency of the estimators. Tu and Wang (2016) used the estimator of Yao and Zhao (2013) that relies on kernel estimates of the likelihood. This estimator is computationally more demanding than QMLE and depends on the choice of tuning parameters. However, their empirical results suggest that both estimators are very similar. Since our purpose is to forecast the center/range system, a naive inverse transformation of the log-range, i.e. $R_t = \exp[\log R_t]$, will generate biased estimates of the conditional mean of the range and of its point forecasts (Granger and Newbold, 1976). We analyze three different bias correction approaches. The first is a parametric factor correction as in Guerrero (1993) that it is very easy to implement and works well when log-range is normally or almost normally distributed. The other two are semiparametric corrections that do not rely on any particular distributional distribution of the errors. The smearing estimator proposed by Duan (1983), also used by Tu and Wang

(2016), estimates the unknown error distribution by the empirical CDF of the residuals and take the desired expectation with respect to the estimated error distribution. Duan's estimator was designed to obtain the fitted values of the range (in-sample) and we extend it to obtain the range forecasts at any forecasting horizon. The bootstrap estimator (Pascual *et al.*, 2005) is another semiparametric correction that, as the smearing estimator, does not require a distributional assumption and has the advantage of incorporating parameter uncertainty in the construction of the forecasts.

We perform several Monte Carlo simulations to assess the out-of-sample performance of a naive point forecast compared to bias-corrected point forecasts. It is interesting to note that biased-corrected forecasts are not uniformly superior to the naive forecast, which in some instances is the preferred choice when faced with a Mean-Absolute-Error loss function.

We apply these methods to the time series of the daily low/high price intervals of the SP500 and Google (GOOG) stock. We deviate from the standard practice in financial econometrics in that we work directly with prices and not with the end-of-day returns. These intervals are more informative than just a daily one-point measurement as they encompass all prices during the day. There are commonalities between the analysis of the center/range system of prices and the standard analysis of end-of-the-day returns and their volatility. As returns, the first-differenced center has large kurtosis, which is more pronounced in the individual stock GOOG than in the index SP500. The log-range, which is close to be normally distributed, is a proxy for volatility as proposed by Parkinson (1980) and Alizadeh *et al.* (2002). It shows a strong autocorrelation as that of an autoregressive

process, which is similar to the patterns found in ARCH processes and stochastic volatility. We also find that that there is Granger-causality from the first-differenced center to the log-range such that positive and large changes in the center will predict narrower ranges, which is similar to the so-called leverage effect. However, an important difference pertains to the construction of the forecasts. In standard ARCH and stochastic volatility processes, the forecast of the return is mostly zero and together with a forecast of the conditional volatility and some conditional distribution of the return, it is possible to generate a density forecast of future returns. In the interval approach, we forecast directly the future low/high *prices* and construct prediction regions of the center and range prices at any desired horizon that do not require parametric distributional assumptions. We find that there is a substantial overlap between the one-day-ahead interval forecast and realized intervals up to 54%.

The organization of the chapter is as follows. In section 2, we explain the estimation of the system center/log-range. In section 3, we present the bias-corrected point forecasts for the center/range system. In section 4, we analyze several Monte Carlo simulations. In section 5, we analyze the time series of the SP500 and GOOG low/high prices, and in section 6, we conclude.

1.2 The Center/Log-Range System

For the interval-valued time series $\{(y_{l,t}, y_{u,t})\}$ for $t = 1, ...T$, we consider the equivalent representation, center/range, $\{(C_t, R_t)\}$ for $t = 1, ...T$ with $R_t \geq 0$. Our objective is to build a model-based point forecast for either representation. To that end, we start with the estimation of a dynamic bivariate system. We propose a linear VAR(p) for the

center/log-range system, from which we construct forecasts for $\{(C_t, \log R_t)\}$. We transform the forecasts to the original measures $\{(C_t, R_t)\}$ or $\{(y_{l,t}, y_{u,t})\}$ and explore the properties of the point forecasts and approximated prediction regions. Let us call $y_{c,t} \equiv C_t$ and $y_{r,t} \equiv \log R_t$. Consider the bivariate VAR(p) system

$$y_{c,t} = \alpha_1 + \sum_{i=1}^{p} \beta_{11}^{(i)} y_{c,t-i} + \sum_{i=1}^{p} \beta_{12}^{(i)} y_{r,t-i} + \varepsilon_{c,t} \tag{1.1}$$

$$y_{r,t} = \alpha_2 + \sum_{i=1}^{p} \beta_{21}^{(i)} y_{c,t-i} + \sum_{i=1}^{p} \beta_{22}^{(i)} y_{r,t-i} + \varepsilon_{r,t} \tag{1.2}$$

where the components of the error vector $(\varepsilon_{1,t}, \varepsilon_{2,t})'$ are white noise processes, possibly contemporaneous correlated, with covariance matrix Ω. The estimation of the system proceeds by least squares. The OLS estimator is consistent under mild assumptions and it will also be a full information maximum likelihood (ML) estimator when bivariate normality of the errors is the true density. Otherwise, assuming bivariate normality, a quasi-maximum likelihood (QML) estimator will be equivalent to a LS estimator. Let $\theta \equiv (\alpha_1, \alpha_2, \beta_{11}^{(1)}, ..., \beta_{12}^{(1)}, ..., \beta_{21}^{(1)}, ..., \beta_{22}^{(1)}, ...)$ be the parameter vector to estimate. Following White (1982), the asymptotic distribution of the Gaussian QML estimator is $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, A^{-1}BA^{-1})$ where matrix A is the (minus) expectation of the Hessian and matrix B is the expectation of the outer product of the score of a Gaussian log-likelihood function.

The QML environment will be the most common estimation approach because bivariate normality of $(\varepsilon_{1,t}, \varepsilon_{2,t})'$ will be difficult to entertain. To guarantee bivariate normality of the system, the conditional densities as well as the marginal densities must also

be normal functions. For financial data, there is evidence that the log-range $y_{r,t}$ (as a proxy for volatility) is near-normal (Alizadeh, Brandt, and Diebold, 2002). The center $y_{c,t}$ is less likely to be normally distributed because the prevalence of fat tails at least in financial data at a relative high frequency, e.g. daily financial returns. In the empirical section, we will test the assumption of bivariate normality as a starting step to construct density forecasts of the full system. Notice that, even if bivariate normality is the correct assumption, the system (1.1-1.2) will generate density forecasts for the center and log-range, but ultimately we will be interested in density forecasts of the center/range system.

Given an information set up to time T, the optimal h-step forecast is the conditional mean i.e. $\hat{y}_{c,T+h|T}$ and $\hat{y}_{r,T+h|T}$ whenever the loss function is symmetric. The point forecasts of the center and the log-range are

$$\hat{y}_{c,T+h|T} = \alpha_1 + \sum_{i=1}^{p} \beta_{11}^{(i)} \hat{y}_{c,T+h-i|T} + \sum_{i=1}^{p} \beta_{12}^{(i)} \hat{y}_{r,T+h-i|T} \tag{1.3}$$

$$\hat{y}_{r,T+h|T} = \alpha_2 + \sum_{i=1}^{p} \beta_{21}^{(i)} \hat{y}_{c,T+h-i|T} + \sum_{i=1}^{p} \beta_{22}^{(i)} \hat{y}_{r,T+h-i|T} \tag{1.4}$$

where $\hat{y}_{c,T+h-i|T} = y_{c,T+h-i}$ and $\hat{y}_{r,T+h-i|T} = y_{r,T+h-i}$ for $i \geq h$. The corresponding forecast errors are $e_{c,T+h|T} = y_{c,T+h} - \hat{y}_{c,T+h|T}$ and $e_{r,T+h|T} = y_{r,T+h} - \hat{y}_{r,T+h|T}$. The Mean Squared Prediction Error (MSPE) or, in this case, the variance-covariance matrix W_h of the forecast error vector $(e_{c,T+h|T}, e_{r,T+h|T})'$ is $W_h = \Omega + \sum_{i=1}^{h-1} \Psi_i \Omega \Psi_i'$ where the coefficient matrix Ψ_i come from the MA(∞) representation of the VAR(p) system. In practice, we plug in consistent parameter estimates, i.e. $\hat{\theta}, \hat{\Omega}$, and $\hat{\Psi}_i$, in the VAR(p) to obtain the estimated point forecasts and their estimated variance-covariance matrix. Parameter uncertainty will

contribute to the MSPE but it will be negligible when the sample size T is large relative to the number of estimated parameters.

1.3 Forecasting the Center/Range System

Taking advantage of the forecasts produced by the VAR(p) for the center/log-range system, we wish to construct the forecasts for the center/range system. The main problem lies in the equation of the log-range. It is easy to see that by taking the exponential transformation of the conditional mean of the log-range, we will obtain a downward biased forecast. By the Jensen's inequality, we have that $\exp(E(y_{r,T+h|T})) < E_T(\exp(y_{r,T+h}))$. In the following sections, we explore different bias correction techniques for point forecasts and different approaches to construct prediction regions for the center/range system.

We start with the simplest forecast. Let $y^*_{r,T+h|T}$ be the forecast for the range of the interval. The biased **naive forecast** of the range is just the exponential transformation of the forecast of the log-range, i.e.,

$$y^*_{r,T+h|T} = \exp(E_T(y_{r,T+h})) = \exp(\hat{y}_{r,T+h|T}) \tag{1.5}$$

1.3.1 Point Forecasts: Factor correction

An approximate bias correction factor can be obtained based on a Taylor' s expansion of $y_{r,T+h}$ around the conditional mean $E_T(\exp(y_{r,T+h}))$. By adapting Guerrero (1993)

to the bivariate VAR(p), the biased-corrected forecast of the range is

$$y^*_{r,T+h|T} \simeq \exp(\hat{y}_{r,T+h|T}) \exp(\frac{W_{h,22}}{2}) \tag{1.6}$$

where $W_{h,22}$ is the lower term of the diagonal of the variance-covariance matrix of the forecast errors W_h. This approximation holds for any distributional assumption of the log-range. However, if we assume that the log-range is normally distributed the expression (1.6) is exact (for details, see the Appendix). For large deviations from normality e.g., skewness or fat tails, we expect that this bias-corrected forecast will not perform well.

1.3.2 Point Forecasts: Smearing correction

Consider the forecast of the log-range (1.4). We write the random variable $y_{r,T+h}$ as the sum of its optimal forecast plus the corresponding forecast error, i.e. $y_{r,T+h} = \hat{y}_{r,T+h|T} + e_{r,T+h|T}$. Then, the conditional mean of the range, i.e. $E_T(\exp(y_{r,T+h}))$, is

$$y^*_{r,T+h|T} = E_T(\exp(y_{r,T+h})) = \exp(\hat{y}_{r,T+h|T}) E_T(\exp(e_{r,T+h|T})) \tag{1.7}$$

that is an unbiased forecast of the range. If we know the distribution function of the forecast error, we can calculate the correction factor $E_T(\exp(e_{r,T+h|T}))$. In the absence of such a knowledge, we introduce the smearing correction based on a nonparametric estimator of the unknown distribution given by the empirical cdf of the forecast error. The smearing estimator was introduced by Duan (1983) in the context of a regression model (in-sample estimation) where the smearing estimator was based on the empirical distribution of the

regression residuals. We adapt this estimator to the forecast errors of the log-range equation of the VAR(p) system. The forecast error $e_{r,T+h|T}$ is a linear function of the unknown innovations from time $T+1$ to $T+h$. Then, we will need the multivariate density function of these innovations, i.e. $f_T(\varepsilon_{r,T+h}, \varepsilon_{r,T+h-1}, \varepsilon_{c,T+h-1}, \varepsilon_{r,T+h-2}, \varepsilon_{c,T+h-2}, ..., \varepsilon_{r,T+1}, \varepsilon_{c,T+1})$. Under the assumption that the vector $(\varepsilon_{c,t}, \varepsilon_{r,t})$ in the VAR(p) system (1.1)-(1.2) is i.i.d, the multivariate density simplifies to a product of identical bivariate densities and a marginal, i.e

$$g_T(\varepsilon_{r,T+h}) f_T(\varepsilon_{r,T+h-1}, \varepsilon_{c,T+h-1}) f_T(\varepsilon_{r,T+h-2}, \varepsilon_{c,T+h-2}) \times .... \times f_T(\varepsilon_{r,T+1}, \varepsilon_{c,T+1}) \quad (1.8)$$

The empirical counterparts to these functions will be the basis for the implementation of the smearing estimator. As an example, consider a VAR(1) system and $h = 2$. Then, the two-step-ahead forecast of the range is

$$\begin{aligned} E_T[\exp(y_{r,T+2})] &= E_T[\exp(\alpha_2 + \beta_{21} y_{c,T+1} + \beta_{22} y_{r,T+1} + \varepsilon_{r,T+2})] = \quad (1.9) \\ &= \exp(C_T) E_T[\exp(\beta_{21}\varepsilon_{c,T+1} + \beta_{22}\varepsilon_{r,T+1} + \varepsilon_{r,T+2})] \end{aligned}$$

where $C_T = \alpha_2 + \beta_{21}(\alpha_1 + \beta_{11} y_{c,T} + \beta_{12} y_{r,T}) + \beta_{22}(\alpha_2 + \beta_{21} y_{c,T} + \beta_{22} y_{r,T})$. Considering (1.8), the conditional expectation of the transformed forecast error will be estimated by the smearing estimator

$$\hat{E}_T[\exp(\beta_{21}\varepsilon_{c,T+1} + \beta_{22}\varepsilon_{r,T+1} + \varepsilon_{r,T+2})] = \Big[\frac{1}{T}\sum_{i=1}^{T} \exp(\varepsilon_{r,i})\Big]\Big[\frac{1}{T}\sum_{j=1}^{T} \exp(\beta_{21}\varepsilon_{c,j} + \beta_{22}\varepsilon_{r,j})\Big] \quad (1.10)$$

1.3.3 Point Forecasts: Bootstrap approach

The three previous approaches to construct the forecast of the range (naive (1.5), factor correction (1.6), and smearing correction (1.7)) depend on the parameter estimates of the VAR(p) system. When the estimation sample is not very large, the effect of parameter estimation on the forecast and on the correction factors is not negligible. In order to account for parameter uncertainty, we extend the bootstrap procedure proposed by Pascual *et al.* (2005) to a VAR system.

The procedure consists of the following steps:

Step 1. Estimate the VAR(p) system (1.1)-(1.2). Obtain the residual vector $(\hat{\varepsilon}_{c,t}, \hat{\varepsilon}_{r,t})'$. Center the residuals, i.e. $\hat{\varepsilon}_t - \bar{\varepsilon}_t$ where $\bar{\varepsilon}_t = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\varepsilon}_t$. Rescale the residuals using the factor $[\frac{T-p}{T-p-d}]^{1/2}$, where d is the number of parameters to estimate. Denote the empirical distribution of the centered and rescaled residuals as $\hat{F}_{\hat{\varepsilon}}$.

Step 2. From $\hat{F}_{\hat{\varepsilon}}$, draw pairwise residuals $(\hat{\varepsilon}_{c,t}, \hat{\varepsilon}_{r,t})'$ with replacement. Together with the parameter estimates from Step 1, generate bootstrap series $\{y^b_{c,1}, ..., y^b_{c,T}\}$ and $\{y^b_{r,1}, ..., y^b_{r,T}\}$ as follows,

$$y^b_{c,t} = \hat{\alpha}_1 + \sum_{i=1}^{p}\hat{\beta}^{(i)}_{11}y^b_{c,t-i} + \sum_{i=1}^{p}\hat{\beta}^{(i)}_{12}y^b_{r,t-i} + \hat{\varepsilon}_{c,t}$$
$$y^b_{r,t} = \hat{\alpha}_2 + \sum_{i=1}^{p}\hat{\beta}^{(i)}_{21}y^b_{c,t-i} + \sum_{i=1}^{p}\hat{\beta}^{(i)}_{22}y^b_{r,t-i} + \hat{\varepsilon}_{r,t}$$

Fix p initial values of $y_{c,t}$ and $y_{r,t}$, that is, $y^b_{c,t} = y_{c,t}$ and $y^b_{r,t} = y_{r,t}$ for $t = 1, ..., p$. For the bootstrap replicate, estimate a VAR(p) to obtain bootstrap parameter estimates $\hat{\alpha}^b = \begin{bmatrix} \hat{\alpha}^b_1 \\ \\ \hat{\alpha}^b_2 \end{bmatrix}$

and $\hat{\beta}^b = \begin{bmatrix} \hat{\beta}_{11}^{b(i)} & \hat{\beta}_{12}^{b(i)} \\ \hat{\beta}_{21}^{b(i)} & \hat{\beta}_{22}^{b(i)} \end{bmatrix}_{i=1,\ldots,p}$.

Step 3. Construct the bootstrap h-step ahead future values of the vector $(y_{c,T+h}, y_{r,T+h})'$ as follows,

$$\hat{y}_{c,T+h}^b = \hat{\alpha}_1^b + \sum_{i=1}^{p} \hat{\beta}_{11}^{b(i)} \hat{y}_{c,T+h-i}^b + \sum_{i=1}^{p} \hat{\beta}_{12}^{b(i)} \hat{y}_{r,T+h-i}^b + \hat{\varepsilon}_{c,T+h}$$
$$\hat{y}_{r,T+h}^b = \hat{\alpha}_2^b + \sum_{i=1}^{p} \hat{\beta}_{21}^{b(i)} \hat{y}_{c,T+h-i}^b + \sum_{i=1}^{p} \hat{\beta}_{22}^{b(i)} \hat{y}_{r,T+h-i}^b + \hat{\varepsilon}_{r,T+h}$$

where $\hat{y}_{c,T+h-i}^b = y_{c,T+h-i}$, and $\hat{y}_{r,T+h-i}^b = y_{r,T+h-i}$ for $i \geq h$, and $(\hat{\varepsilon}_{c,T+h}, \hat{\varepsilon}_{r,T+h})'$ are pairwise random draws with replacement from $\hat{F}_{\hat{\epsilon}}$. Notice that the last p values of the original data are fixed in this step.

Step 4. Repeat steps 2 and 3 B times. We obtain a bootstrap conditional distribution of the vector $(\hat{y}_{c,T+h}^b, \hat{y}_{r,T+h}^b)'$ and, by taking the exponential of log-range, we finally obtain a bootstrap conditional distribution of $(\hat{y}_{c,T+h}^b, \exp(\hat{y}_{r,T+h}^b))'$ that we denote as F_y^b .

Step 5. The h-step ahead bootstrap forecasts of the center and range of the interval are obtained by averaging over the number of bootstrap replicates,

$$y_{c,T+h|T}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{y}_{c,T+h,i}^b \tag{1.11}$$

$$y_{r,T+h|T}^* = \frac{1}{B} \sum_{i=1}^{B} \exp(\hat{y}_{r,T+h,i}^b) \tag{1.12}$$

1.4 Monte Carlo Simulations

We perform extensive Monte Carlo simulations to assess the performance of the point forecasts for the center/range system obtained by implementing the different approaches surveyed in section 3.

We consider a VAR(1) for the center/log-range system (1.1)-(1.2) with the following parameter values: $(\alpha_1, \alpha_2)' = (0, 0)'$, $(\beta_{11} \;\; \beta_{12} | \; \beta_{21} \;\; \beta_{22})' = (0.4 \;\; 0.1 | \; 0.2 \;\; 0.4)'$ (mid-persistent dynamics), and $(\beta_{11} \;\; \beta_{12} | \; \beta_{21} \;\; \beta_{22})' = (0.5 \;\; 0.1 | \; 0.2 \;\; 0.8)'$ (high-persistent dynamics). The dominant roots of the respective characteristic equations are 0.54 and 0.86 in the mid-persistent and high-persistent cases. The variance-covariance matrix of the error term Ω will be determined according to the assumed marginal distributions of the center and the range, as we will see in the following sections. The VAR(1) center/log-range system is estimated recursively by least squares. We consider three forecast horizons, i.e. $h = 1, 2, 3$. We construct point forecasts for the center and range according to the methods in section 3. The point forecasts are evaluated in an out-of-sample exercise with $P = 100$ observations. The number of Monte Carlo replications is 500 and the number of bootstrap samples is $B = 2000$.

We consider several probability density functions for the center and the range and log-range. For the center, we assume (i) normal density and (ii) Student-t with 7 degrees of freedom. For the range, we assume (i) exponential density with $\lambda = 1$, and (ii) log-normal density so that the log-range is normally distributed. Note that the bivariate density of the system center/range will not be normal for any combination of the assumed marginal densities of the center and the range. We simulate data from the four possible

combinations of the center/range marginal densities. We also need to assume some degree of contemporaneous dependence between the two equations of the VAR(1). When the center and the log-range are simulated under the normal assumption, we assume that the variance-covariance matrix of the errors of the VAR has a non-zero covariance. In the remaining three cases, we generate contemporaneous dependence between the center and log-range by using a bivariate normal copula function (correlation coefficient equal to 0.5) with the assumed marginal densities, i.e. (normal center, exponential range), (Student-t center, exponential range), and (Student-t center, normal log-range).

We evaluate the point forecasts according to four loss functions: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Distance Error (MDE), and Average of Coverage and Efficiency Rates (ACE). We convert the center and range forecasts into the lower and upper bounds forecasts of the interval, i.e., $(y^*_{c,T+h|T} - (1/2)y^*_{r,T+h|T},\ \ y^*_{c,T+h|T} + (1/2)y^*_{r,T+h|T}) = (y_{l,T+h|T}, y_{u,T+h|T})$ because this is the final object of interest.

The RMSE of the upper bound (RMSEU) and of the lower bound (RMSEL) forecasts are

$$RMSEU = \sqrt{\frac{1}{P-h+1}\sum_{t=T}^{T+P-h}(y_{u,t+h} - y_{u,t+h|t})^2}\ ; \tag{1.13}$$

$$RMSEL = \sqrt{\frac{1}{P-h+1}\sum_{t=T}^{T+P-h}(y_{l,t+h} - y_{l,t+h|t})^2}$$

The MAE of the upper bound (MAEU) and of the lower bound (MAEL) forecasts are

$$MAEU = \frac{1}{P-h+1} \sum_{t=T}^{T+P-h} |y_{u,t+h} - y_{u,t+h|t}| \quad ; \tag{1.14}$$
$$MAEL = \frac{1}{P-h+1} \sum_{t=T}^{T+P-h} |y_{l,t+h} - y_{l,t+h|t}|$$

These two loss functions, RMSE and MAE, consider the performance of each bound forecast separately. If we wish to asses the interval forecast as a whole and not just each of the bounds, we need either distance measures or measures of coverage between the actual and the forecast interval.

The MDE is defined as

$$MDE = \frac{1}{P-h+1} \sum_{t=T}^{T+P-h} D(y_{t+h}, y_{t+h|t}) \tag{1.15}$$

where $D(y_{t+h}, y_{t+h|t}) = \frac{1}{\sqrt{2}}[(y_{u,t+h} - y_{u,t+h|t})^2 + (y_{l,t+h} - y_{l,t+h|t})^2]^{\frac{1}{2}}$.

The Coverage Rate (CR) and the Efficiency Rate (ER) are defined as

$$CR = \frac{1}{P-h+1} \sum_{t=T}^{T+P-h} \frac{w(y_{t+h} \cap y_{t+h|t})}{w(y_{t+h})} \quad ; \quad ER = \frac{1}{P-h+1} \sum_{t=T}^{T+P-h} \frac{w(y_{t+h} \cap y_{t+h|t})}{w(y_{t+h|t})} \quad ; \tag{1.16}$$

where $w(X)$ is the width of interval X and "$\cap$" the intersection of two intervals. The ACE is the simple average of CR and ER, i.e., $ACE = (CR + ER)/2$.

Next, we compare and comment on the simulation results for the following cases with more details:

1.4.1 Center and Log-Range are Normally distributed

Since the log-range is normally distributed, the Guerrrero biased-corrected forecast of the range is exact (see expression 1.6) and this forecast is the conditional mean of the range variable at time $T+h$. The forecasts based on the smearing and bootstrap approaches will also deliver the conditional mean of the future center and range variables. Consequently, we expect Guerrero, smearing, and bootstrap approaches outperform (lower loss) the naive forecast when the loss function is quadratic. The naive forecast is under-biased and, thus iit will not be the minimizer of the expected quadratic loss function. In Table 1.1, we report the values of the RMSE loss functions associated with the one-, two-, and three-step ahead forecasts of the upper and lower bounds of the intervals. We estimate the mid-persistence VAR(1) with variance-covariance matrix of the error term $\Omega = (2 \;\; 0.5| \; 0.5 \;\; 1)'$. We run two estimation samples $T = 100$ and $T = 500$ and the average loss is calculated over an out-of-sample period of 100 observations, i.e. $P = 100$. The RMSE losses delivered by Guerrero, Smearing and Bootstrap approaches are practically identical with Guerrero having a very tiny advantage. The loss delivered by the naive approach is slightly higher than the other approaches; it is between 1 and 4 % larger than the loss attached to Guerrero's bias-corrected forecast. As expected, when $T = 500$ the losses are smaller than those when $T = 100$. This is because the contribution of parameter uncertainty is less important as the estimation sample becomes larger.

	Naïve	Guerrero	Smearing	Bootstrap	Nve./Guer.
$T = 100$ $P = 100$	RMSEU				
1-step	2.3938	2.3666	2.3701	2.3675	1.012
2-step	2.6612	2.6292	2.6356	2.6307	1.012
3-step	2.7306	2.6954	2.7034	2.6979	1.013
	RMSEL				
1-step	1.9306	1.8895	1.8930	1.8909	1.022
2-step	1.9928	1.9363	1.9433	1.9385	1.029
3-step	1.9971	1.9333	1.9414	1.9355	1.033
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Guer.
$T = 500$ $P = 100$	RMSEU				
1-step	2.3787	2.3370	2.3366	2.3376	1.012
2-step	2.6441	2.5987	2.5987	2.5996	1.017
3-step	2.7141	2.6682	2.6684	2.6691	1.017
	RMSEL				
1-step	1.9230	1.8666	1.8674	1.8673	1.030
2-step	1.9864	1.9198	1.9209	1.9204	1.035
3-step	1.9932	1.9245	1.9261	1.9254	1.036

Table 1.1: RMSE losses. Mid-persistence VAR(1)

In Table 1.2, we run similar experiment but for the high-persistence VAR(1). We observe that when $T = 100$ and $P = 100$, the naive method delivers substantially lower losses than Guerrero, smearing, and bootstrap methods. The naive loss can be up to 25% lower than that of Guerrero. Though these results are puzzling, they are two possible explanations. First, this is a small sample estimation problem. When the sample is increased to $T = 500$ so that parameter estimates are more precise, the naive loss is similar to that of the other approaches, though still 5-8% lower than that of Guerrero's.

	Naïve	Guerrero	Smearing	Bootstrap	Naïve/Guerrero	
$T = 100$ $P = 100$	RMSEU				$P = 100$	$P = 1000$
1-step	17.3680	19.5536	20.2202	19.3771	0.888	0.973
2-step	20.5549	25.5747	27.6231	25.5726	0.804	0.947
3-step	21.5866	28.6626	32.2994	29.4509	0.753	0.938
	RMSEL					
1-step	16.9687	19.1570	19.8272	18.9868	0.886	0.973
2-step	19.9160	24.9597	27.0185	24.9715	0.798	0.946
3-step	20.8181	27.9295	31.5851	28.7459	0.745	0.937
	Naïve	Guerrero	Smearing	Bootstrap	Naïve/Guerrero	
$T = 500$ $P = 100$	RMSEU				$P = 100$	$P = 2000$
1-step	15.8332	16.6583	16.7766	16.6272	0.950	1.009
2-step	18.5702	19.9974	20.2574	20.0655	0.929	1.007
3-step	19.7148	21.5187	21.9005	21.8275	0.916	1.006
	RMSEL					
1-step	15.4485	16.2687	16.3876	16.2399	0.949	1.009
2-step	17.9623	19.3955	19.6583	19.4670	0.926	1.007
3-step	18.9771	20.7992	21.1865	21.1167	0.912	1.006

Table 1.2: RMSE losses. High-persistence VAR(1)

The second reason is related to the small size of the evaluation sample P. For simplicity consider $h = 1$ and compare point-wise the forecast errors associated with the naive forecast (1.5) and the Guerrero bias-corrected forecast (1.6) for the bivariate VAR (1.1)-(1.2). The one-step-ahead forecast error for the range associated with the naive approach is $e^N_{r,T+1|T} = \exp(\hat{y}_{r,T+1|T})(\exp(\varepsilon_{r,T+1}) - 1)$. The one-step- ahead forecast error for the range based on the Guerrero method is $e^G_{r,T+1|T} = \exp(\hat{y}_{r,T+1|t})(\exp(\varepsilon_{r,T+1}) - \exp(\sigma^2_r/2))$ where

σ_r^2 is the variance of $\varepsilon_{r,T+1}$. Then,

$$
\begin{aligned}
&|e^G_{r,T+1|T}| > |e^N_{r,T+1|T}| && \text{for} && \varepsilon_{r,T+1} \leq 0 \\
&|e^G_{r,T+1|T}| > or < |e^N_{r,T+1|T}| && \text{for} && 0 \leq \varepsilon_{r,T+1} \leq \sigma_r^2/2 \\
&|e^G_{r,T+1|T}| \ll |e^N_{r,T+1|T}| && \text{for} && \sigma_r^2/2 < \varepsilon_{r,T+1}
\end{aligned}
$$

Since $\varepsilon_{r,T+1}$ is symmetrically distributed around zero, we expect that $|e^G_{r,T+1|T}| > |e^N_{r,T+1|T}|$ will happen for more than 50% of the observations in the evaluation sample. In addition, the distribution of the forecast errors is skewed to the right because of the exponential transformation. It is only when $\sigma_r^2/2 < \varepsilon_{r,T+1}$, that the Guerrero forecast errors become smaller than those of the naive approach. How often will encounter these cases will depend on the magnitude of σ_r^2 and, when $h > 1$, on the magnitude of the parameters of the model as well. In a small sample environment, e.g. $P = 100$, the RMSE is likely to be dominated by instances like $|e^G_{r,T+1|T}| > |e^N_{r,T+1|T}|$; we will need a very large sample for the full skewed distribution of the forecast errors to show up. In Table 1.2, we show that for $T = 100$ and $P = 1000$, the ratio naive/Guerrero RMSEs increases from about 0.80 to about 0.95. When we increase both the estimation sample and the evaluation sample sizes, i.e. $T = 500$ and $P = 2000$, the RMSE loss values attached to the Naive approach are larger than those attached to Guerrero's corrected forecast, as we expected. In summary, in small samples the simplest naive approach delivers a good point forecast of the interval evaluated according to a RMSE loss function.

In Table 1.3, we report the values of the MAE, MDE, and ACE loss functions for the mid-persistence VAR(1) and the small sample case, i.e. $T = 100$ and $P = 100$. When

considering the MAE function, the h-step forecast minimizer is the conditional median of the variable at time $T+h$. Recall that after estimating the VAR system, we obtain the h-step ahead forecast of the log-range, i.e. $\hat{y}_{r,T+h|T}$, which is the conditional mean of $y_{r,T+h}$. Since the log-range is normally distributed, this conditional mean is also the conditional median of $y_{r,T+h}$. The naive approach applies a monotonic transformation to $\hat{y}_{r,T+h|T}$ so that the quantile is preserved and the naive forecast $\exp(\hat{y}_{r,T+h|t})$ is also the conditional median of the range at time $T+h$. Consequently, we expect the naive forecasts for center and range to deliver smaller MAE losses than Guerrero, smearing and bootstrap approaches. However, on translating the forecasts of center/range into the forecasts of the lower/upper bounds of the interval, this may not be the case because the median does not have the additivity property. In Table 1.3, we observe that the four approaches are practically equivalent and they deliver similar losses.

	Naïve	Guerrero	Smearing	Bootstrap	Nve./Guer.
$T = 100$ $P = 100$	MAEU				
1-step	1.6333	1.6576	1.6611	1.6589	0.985
2-step	1.8157	1.8441	1.8510	1.8475	0.985
3-step	1.8605	1.8891	1.8979	1.8945	0.985
	MAEL				
1-step	1.3160	1.3190	1.3227	1.3216	0.998
2-step	1.3647	1.3494	1.3566	1.3531	1.011
3-step	1.3704	1.3451	1.3546	1.3490	1.019
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Guer.
$T = 100$ $P = 100$	MDE				
1-step	1.5610	1.5995	1.6029	1.6022	0.976
2-step	1.6866	1.7335	1.7406	1.7389	0.973
3-step	1.7150	1.7626	1.7722	1.7701	0.973
	Naïve	Guerrero	Smearing	Bootstrap	Boot./Nve.
$T = 100$ $P = 100$	ACE				
1-step	0.3232	0.3865	0.3850	0.3880	1.200
2-step	0.3017	0.3926	0.3906	0.3951	1.309
3-step	0.2975	0.3987	0.3970	0.4021	1.352

Table 1.3: MAE and MDE losses and ACE. Mid-persistence VAR(1)

We evaluate the joint performance of the lower/upper bounds forecasts with the MDE and ACE functions. For the MDE, all four approaches deliver similar losses with the naive approach showing a slight better performance. However, when we assess the coverage and efficiency of the interval forecast, Guerrero, smearing and bootstrap forecasts are much superior to the naive forecast even in this small sample setting. The bootstrap approach delivers between 20 and 35% improvement in ACE over the naive forecast.

1.4.2 Center is Student-t and Log-Range is Normally distributed

We deviate from bivariate normality of the center/log-range system by introducing fat tail behavior in the center. We assume that the conditional density of the center is Student-t with 7 degrees of freedom while the log-range remains normally distributed. We estimate the same mid-persistence VAR(1) with variance-covariance matrix of the error term $\Omega = (1.4 \;\; 0.6|\; 0.6 \;\; 1)'$ and an estimation sample of $T = 100$. The average losses are calculated over an out-of-sample period of 100 observations, i.e. $P = 100$. In Table 1.4, we report the values of the RMSE, MAE, MDE, and ACE functions. Regarding RMSE, the Guerrero's bias-corrected forecast, the smearing and the bootstrap forecasts deliver similar losses. Guerrero's bias correction is not exact any longer as it does not consider the kurtosis in the center. Nevertheless, Guerrero's bias-corrected forecast and the bootstrap forecast have similar performance and they are preferred to the naive approach. Regarding MAE, the naive approach seems to have a slight advantage over the other three methods. Regarding MDE, the naive forecast delivers the smallest losses overall but, when the ACE function is considered, the naive forecast does not provide as much efficiency and interval coverage as the bootstrap forecast. The bootstrap approach delivers between 18 and 32% improvement in ACE over the naive forecast.

	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$	RMSEU				
1-step	2.5108	2.4583	2.4661	2.4302	1.033
2-step	2.8117	2.7796	2.7925	2.7874	1.009
3-step	2.8785	2.8439	2.8554	2.8555	1.008
	RMSEL				
1-step	1.8946	1.8229	1.8309	1.7953	1.055
2-step	1.9968	1.9311	1.9448	1.9395	1.029
3-step	2.0080	1.9363	1.9494	1.9482	1.031

	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$	MAEU				
1-step	1.4971	1.5308	1.5362	1.5310	0.978
2-step	1.6818	1.7309	1.7478	1.7474	0.962
3-step	1.7285	1.7775	1.7993	1.8019	0.959
	MAEL				
1-step	1.0636	1.0675	1.0741	1.0690	0.995
2-step	1.1176	1.1047	1.1232	1.1209	0.997
3-step	1.1263	1.1031	1.1272	1.1266	0.999

	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$	MDE				
1-step	1.3643	1.4059	1.4120	1.4078	0.969
2-step	1.4949	1.5528	1.5710	1.5712	0.951
3-step	1.5251	1.5850	1.6086	1.6118	0.946

	Naïve	Guerrero	Smearing	Bootstrap	Boot./Nve.
$T = 100$ $P = 100$	ACE				
1-step	0.3905	0.4582	0.4572	0.4602	1.178
2-step	0.3665	0.4647	0.4639	0.4682	1.277
3-step	0.3607	0.4711	0.4704	0.4752	1.317

Table 1.4: RMSE, MAE and MDE losses and ACE. Mid-persistence VAR(1)

1.4.3 Center is Student-t and Range is Exponential (λ) distributed

We deviate from bivariate normality of the center/log-range system by introducing fat tail behavior in the center and skewness in the log-range. We assume that the conditional density of the center is Student-t with 7 degrees of freedom and that of the range is exponentially distributed, i.e. $\exp(\lambda = 1)$. On applying the log transformation to the exponentially distributed range, the asymmetry in the log-range still persists. We estimate the same mid-persistence VAR(1) with variance-covariance matrix of the error term $\Omega = (1.4 \;\; 0.77 | \; 0.77 \;\; 1.69)'$ and an estimation sample of $T = 100$. The average losses are calculated over an out-of-sample period of 100 observations, i.e. $P = 100$. In Table 1.5, we report the values of the RMSE, MAE, MDE, and ACE functions. Since the departure from bivariate normality is more acute, we expect the naive and Guerrero's bias-corrected forecasts to be bad approximations to the conditional means of the center and range at time $T + h$ and consequently, they will not be able to minimize the RMSEs. This is what we observe. The Guerrero's bias-corrected forecast delivers between 3 and 7% higher losses than the bootstrap forecast while the smearing and bootstrap approaches provide the smallest RMSE losses. Regarding MAE, Guerrero's bias-corrected forecast is also the worst performer. Naive, smearing, and bootstrap forecasts seem to be equivalent. Regarding MDE, the naive forecast delivers the smallest losses but, when the ACE function is considered, the naive forecast does not provide as much efficiency and interval coverage as the other three methods. The bootstrap approach delivers between 20 and 38% improvement in ACE over the naive forecast.

	Naïve	Guerrero	Smearing	Bootstrap	Guer./Boot.
$T = 100$ $P = 100$	RMSEU				
1-step	2.3527	2.3980	2.3069	2.3082	1.039
2-step	2.6584	2.7013	2.6149	2.6162	1.033
3-step	2.7334	2.7624	2.6861	2.6859	1.028
	RMSEL				
1-step	1.6873	1.7083	1.5974	1.5985	1.069
2-step	1.7635	1.7722	1.6607	1.6616	1.067
3-step	1.7703	1.7620	1.6591	1.6598	1.062
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$	MAEU				
1-step	1.5946	1.7181	1.6274	1.6264	0.980
2-step	1.7896	1.9736	1.8509	1.8544	0.965
3-step	1.8387	2.0367	1.9099	1.9141	0.960
	MAEL				
1-step	1.1129	1.1971	1.0972	1.0963	1.015
2-step	1.1680	1.2754	1.1440	1.1445	1.020
3-step	1.1766	1.2830	1.1441	1.1463	1.026
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$	MDE				
1-step	1.4521	1.5891	1.4817	1.4821	0.980
2-step	1.5892	1.8007	1.6525	1.6573	0.959
3-step	1.6204	1.8536	1.6954	1.7030	0.951
	Naïve	Guerrero	Smearing	Bootstrap	Boot./Nve.
$T = 100$ $P = 100$	ACE				
1-step	0.4039	0.5118	0.4820	0.4860	1.203
2-step	0.3799	0.5360	0.4979	0.5037	1.326
3-step	0.3742	0.5493	0.5076	0.5153	1.377

Table 1.5: RMSE, MAE and MDE losses and ACE. Mid-persistence VAR(1)

1.4.4 Center is Normal and Range is Exponential (λ) distributed

We deviate from bivariate normality of the center/log-range system by introducing only skewness in the log-range. We assume that the conditional density of the center is normal and that of the range is exponentially distributed, i.e. $\exp(\lambda = 1)$. We estimate the same mid-persistence VAR(1) with variance-covariance matrix of the error term $\Omega = (2 \;\; 0.92| \; 0.92 \;\; 1.69)'$ and an estimation sample of $T = 100$. The average losses are calculated over an out-of-sample period of 100 observations, i.e. $P = 100$. In Table 1.6, we report the values of the RMSE, MAE, MDE, and ACE functions. Regarding RMSE, Guerrero's correction ignores the skewness and the bias-corrected forecast as well as the naive forecast generate the largest losses. The smearing and bootstrap approaches provide the smallest RMSE losses. Regarding MAE, Guerrero's bias-corrected forecast is the worst performer. Naive, smearing, and bootstrap forecasts seem to be equivalent. Regarding MDE, the naive forecast delivers the smallest losses but, when the ACE function is considered, the naive forecast does not provide as much efficiency and interval coverage as the other three methods. The bootstrap approach delivers between 22 and 43% improvement in ACE over the naive forecast.

	Naïve	Guerrero	Smearing	Bootstrap	Guer./Boot.
$T = 100$ $P = 100$			RMSEU		
1-step	2.7333	2.7649	2.6777	2.6756	1.033
2-step	3.0916	3.1110	3.0369	3.0323	1.026
3-step	3.1861	3.1929	3.1311	3.1248	1.022
			RMSEL		
1-step	1.9832	1.9954	1.8848	1.8836	1.059
2-step	2.0745	2.0599	1.9587	1.9546	1.054
3-step	2.0811	2.0477	1.9599	1.9537	1.048
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$			MAEU		
1-step	1.8624	1.9811	1.8929	1.8915	0.985
2-step	2.0878	2.2579	2.1457	2.1447	0.973
3-step	2.1495	2.3292	2.2161	2.2138	0.971
			MAEL		
1-step	1.3214	1.3995	1.3019	1.3008	1.016
2-step	1.3844	1.4724	1.3519	1.3484	1.027
3-step	1.3942	1.4699	1.3491	1.3442	1.037
	Naïve	Guerrero	Smearing	Bootstrap	Nve./Boot.
$T = 100$ $P = 100$			MDE		
1-step	1.7032	1.8400	1.7323	1.7316	0.984
2-step	1.8608	2.0677	1.9231	1.9230	0.968
3-step	1.8998	2.1235	1.9727	1.9727	0.963
	Naïve	Guerrero	Smearing	Bootstrap	Boot./Nve.
$T = 100$ $P = 100$			ACE		
1-step	0.3508	0.4569	0.4269	0.4294	1.224
2-step	0.3284	0.4843	0.4423	0.4469	1.361
3-step	0.3221	0.5004	0.4539	0.4600	1.428

Table 1.6: RMSE, MAE and MDE losses and ACE. Mid-persistence VAR(1)

1.4.5 Summary

1. When the departure from bivariate normality of the center/log-range system is large, we recommend implementing a bootstrap forecast or a smearing forecast if the loss function of the forecaster is RMSE. If log-range is normal or approximately normal, the Guerrero's biased-corrected forecast is also equivalent to the bootstrap forecast. If the overall system is closely bivariate normal, it should be noted that the naive forecast can be a good performer when the estimation sample is small.

2. If the loss function of the forecaster is MAE, the Guerrero's biased-corrected forecast should not be implemented; the naive, smearing, and bootstrap forecasts are preferred.

3. If the loss function of the forecaster is MDE, the naive forecast performs well.

4. If the loss function is ACE, a bootstrap forecast provides a large coverage of the realized interval with coverage rates between 40 % and 52%.

1.5 Empirical Application

We collect the daily intervals of low/high prices (in $) of the SP500 index and of the Google (GOOG) stock from January 2, 2009 to January 25, 2017 for a total of 2030 daily observations. The center time series is non-stationary and we work with the first differences of the center. In Table 1.7, we provide the descriptive statistics of the first-differenced center, range, and log-range. For the SP500, the first-differenced center shows fat tails with a coefficient of kurtosis of 6.29 and it is slightly skewed to the left. The range is skewed with a long right tail but the log-range is almost symmetric and has a kurtosis of about 3. For

GOOG, the first-differenced center is skewed to the right and very fat-tailed with a kurtosis of 35.81. The range is also heavily skewed to the right and the log-range, though slightly skewed to the right, has a kurtosis of about 3. We can appreciate these characteristics in Figure 1.1 where we plot the first-differenced center and range of the daily intervals and their unconditional bivariate density function. In both cases, SP500 and GOOG, we are dealing with a fat-tail density for the first-differenced center, more pronounced in GOOG than in the SP500 index, and a log-range that seems to be normally distributed. These characteristics are similar to those in the simulation case presented in section 4.2.

	SP500 Index		
	D(Center)	Range	log-Range
Mean	0.68	17.53	2.73
Standard Error	0.25	0.22	0.01
Median	1.30	15.13	2.72
Mode	-0.92	9.26	2.23
Standard Deviation	11.48	10.08	0.52
Sample Variance	131.69	101.66	0.27
Kurtosis	6.29	11.34	2.90
Skewness	-0.55	2.11	0.16
Minimum	-86.40	3.68	1.30
Maximum	58.27	101.79	4.62
	GOOG stock		
	D(Center)	Range	log-Range
Mean	0.33	7.46	1.85
Standard Error	0.13	0.11	0.01
Median	0.15	6.19	1.82
Mode	-0.94	5.38	1.95
Standard Deviation	6.04	4.92	0.55
Sample Variance	36.47	24.17	0.30
Kurtosis	35.81	15.47	3.18
Skewness	2.06	2.68	0.33
Minimum	-36.67	1.09	0.09
Maximum	86.90	54.00	3.99

Table 1.7: Descriptive Statistics for first-differenced center, range, and log-range of daily intervals (Jan.2, 2009-Jan.25, 2017)

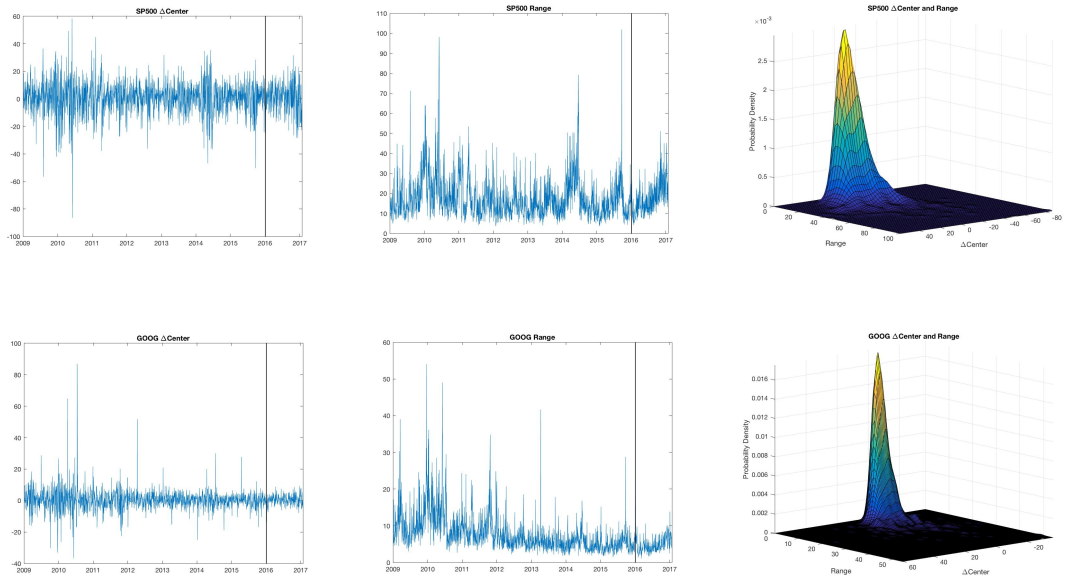


Figure 1.1: SP500 and GOOG. Time series plots of first-differenced center and range and bivariate density

We proceed with the modeling of the bivariate system first-differenced center/log-range from which we will construct the point forecasts. We split the total sample into an estimation sample from January 2, 2009 to December 31, 2015 with 1762 observations, and an evaluation sample from January 4, 2016 to January 25, 2017 with 268 observations. In Figure 1.2, we present the autocorrelograms of the first-differenced center and the range. For both, the SP500 and GOOG, the profiles of the ACF and PACF are very similar. The first-differenced center has only a mild autocorrelation of order one of about 0.2-0.3, which may be induced by a bid-ask effect. This is in agreement with what we observed when we model the more traditional end-of-the-day return. The ACF and PACF of the log-range present the profile of an autoregressive process with strong memory. An AR(6) for the SP500 index and an AR(8) for GOOG seem to be appropriate to capture these dynamics.

This mimics the autocorrelation that we observe in squared (end-of-the-day) returns when modelling a conditional variance, which is not very surprising because range or log-range are good proxies for volatility. The SIC selects a VAR(6) for the SP500 Index and a VAR(5) for GOOG. Conservatively, we proceed to estimate a VAR(6) for the SP500 and a VAR(8) for GOOG. The estimation results are presented in Table 1.8.

SP500 Index. VAR(6) for first-differenced center and log-range system

	D(Center)			log-Range		
	Coeff. estimate	SE (robust)	t-statistic	Coeff. estimate	SE (robust)	t-statistic
Constant	-0.9344	2.0444	-0.4571	0.0759	0.0784	9.4694
D-C(-1)	0.3404	0.0305	11.1607	-0.0112	0.0010	-11.2000
D-C(-2)	-0.1530	0.0321	-4.7664	-0.0027	0.0010	-2.7000
D-C(-3)	0.0314	0.0408	0.7696	-0.0030	0.0010	-3.0000
D-C(-4)	-0.0551	0.0297	-1.8552	-0.0022	0.0009	-2.4444
D-C(-5)	-0.0209	0.0301	-0.6944	0.0010	0.0010	1.0000
D-C(-6)	-0.0011	0.0305	-0.0361	-0.0009	0.0009	-1.0000
log-R(-1)	-0.5030	0.6925	-0.7264	0.0852	0.0265	3.2151
log-R(-2)	0.1281	0.6207	0.2064	0.1845	0.0258	7.1512
log-R(-3)	-0.1556	0.6141	-0.2534	0.1539	0.0228	6.7500
log-R(-4)	0.9157	0.6925	1.3223	0.0760	0.0253	3.0040
log-R(-5)	-0.2561	0.5895	-0.4344	0.1086	0.0252	4.3095
log-R(-6)	0.4266	0.5911	0.7217	0.1227	0.0229	5.3581
Adj. R-squared	0.1086			0.3975		

GOOG. VAR(8) for first-differenced center and log-range system

	D(Center)			log-Range		
	Coeff. estimate	SE (robust)	t-statistic	Coeff. estimate	SE (robust)	t-statistic
Constant	0.0465	0.6253	0.0744	0.2345	0.0427	5.4918
D-C(-1)	0.1807	0.0330	5.4825	-0.0019	0.0019	-0.9938
D-C(-2)	-0.0391	0.0266	-1.4718	-0.0010	0.0018	-0.5683
D-C(-3)	-0.0022	0.0439	-0.0496	-0.0001	0.0016	-0.0714
D-C(-4)	-0.0060	0.0347	-0.1740	-0.0001	0.0015	-0.0951
D-C(-5)	-0.0481	0.0226	-2.1304	0.0008	0.0016	0.4752
D-C(-6)	0.0240	0.0274	0.8750	0.0011	0.0017	0.6839
D-C(-7)	-0.0109	0.0227	-0.4789	0.0011	0.0017	0.6257
D-C(-8)	-0.0241	0.0259	-0.9289	0.0011	0.0017	0.6719
log-R(-1)	0.0001	0.4121	0.0002	0.3306	0.0262	12.5943
log-R(-2)	-0.1060	0.3291	-0.3220	0.0781	0.0250	3.1249
log-R(-3)	0.4226	0.4521	0.9345	0.1182	0.0264	4.4768
log-R(-4)	0.0565	0.3521	0.1606	0.0705	0.0257	2.7453
log-R(-5)	0.0148	0.3324	0.0444	0.1011	0.0235	4.2955
log-R(-6)	-0.1495	0.4707	-0.3176	0.0265	0.0265	1.0023
log-R(-7)	-0.2280	0.3365	-0.6775	0.0486	0.0243	2.0039
log-R(-8)	0.1448	0.3117	0.4645	0.0945	0.0221	4.2689
Adj. R-squared	0.0275			0.4616		

Table 1.8: SP500 and GOOG. Estimation of VAR for first-differenced center and log-range system (Jan.2, 2009-Dec. 31, 2016)

As expected, the first-differenced center equation does not have much predictive power. Only its own past (one or two lags) first-differenced centers are statistically significant at the conventional significance levels. There is not dynamic effect of past log-ranges

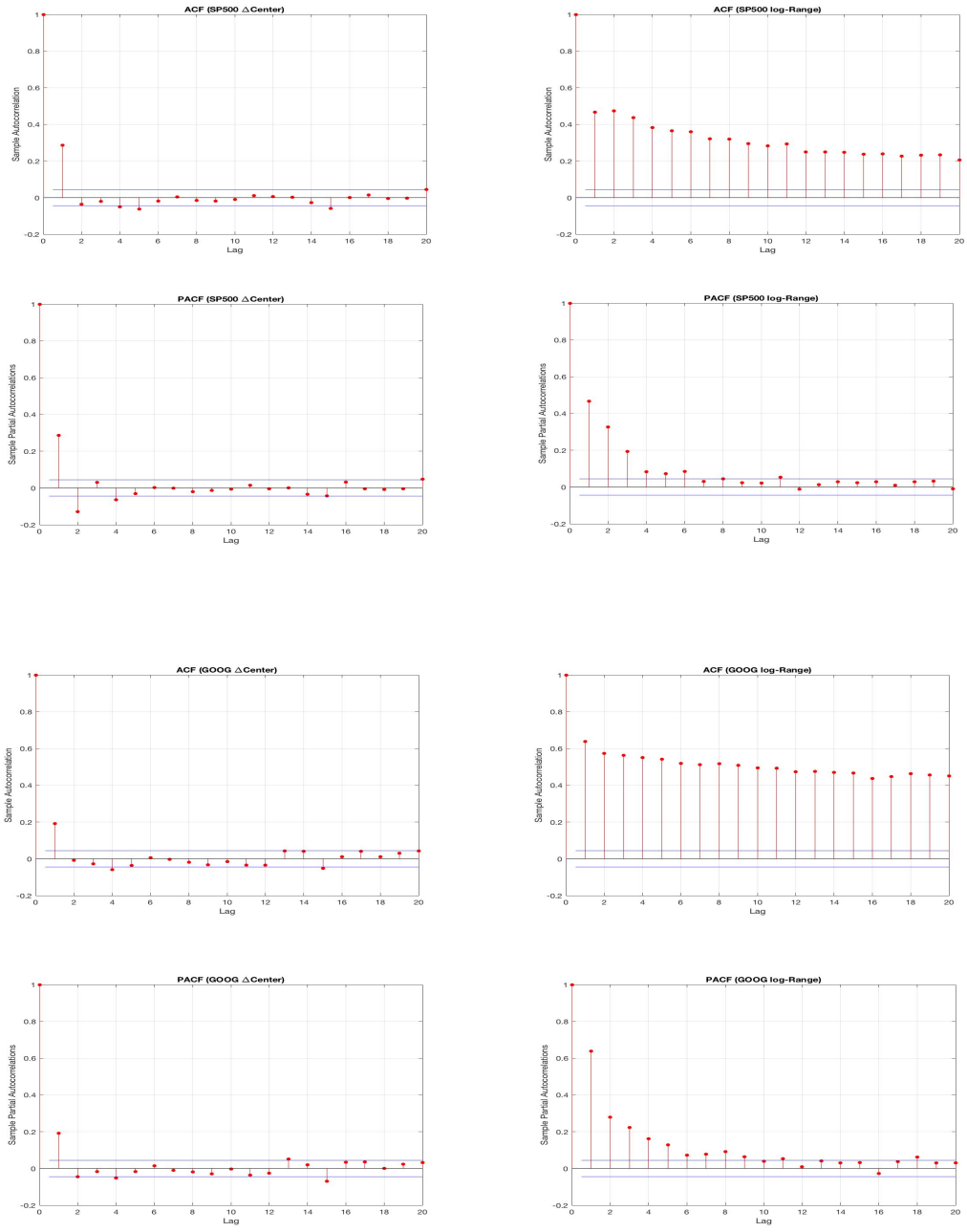


Figure 1.2: SP500 and GOOG. Autocorrelograms of first-differenced center and log-range

on center and the overall in-sample R-squared are 0.11 for the SP500 Index and 0.03 for GOOG. On the other hand, the goodness-of-fit in the log-range equation is much higher with R-squared of 0.40 for the SP500 and 0.46 for GOOG. In the case of SP500 Index, the first-differenced center Granger-causes the log-range in such a way that past first-differenced centers are negatively correlated with current log-ranges. i.e. positive and large changes in centers today will predict smaller ranges tomorrow. In both cases, SP500 and GOOG, the most relevant aspect is the strong and statistically significant autoregressive nature of the log-range as we have already observed in the ACF/PACF. The residuals corresponding to these systems are all clear of any autocorrelation. The residuals from the first-differenced center equations remain leptokurtic and the residuals from the log-range equations remain basically symmetric around zero with a sample kurtosis of about 3. With these characteristics, the joint density of first-differenced center and log-range will not be bivariate normal.

Formally, we test for bivariate normality by implementing the Generalized Auto-ContouR (G-ACR) (in-sample) tests based on the Probability Integral Transforms (PIT) of the joint density under the null hypothesis of bivariate normality (González-Rivera and Sun, 2015). In Table 1.9, we report the result of the t-statistics ($t_{k,\alpha}$) that canvas from the 1% to 99% PIT autocontours for $k = 1, 2, ...5$ lags. The null hypothesis is strongly rejected at the 5% significance level for mostly all but the 10%, 90% and 95% autocontours in the case of the SP500 index and the 5% , 90% and 95% autocontours in the case of GOOG. The aggregated test C_k also reinforce the strong rejection of bivariate normality. In Figure 1.3, we plot just some of the autocontours of the PITs (ΔCenter$_t$, log-range$_{t-1}$) and (ΔCenter$_t$, log-range$_{t-2}$) for SP500 and GOOG. Under the correct null, the distribution of the PITS

should be uniformly distributed within these autocontour squares. For GOOG we observe a much less uniform distribution of the PITs than for the SP500 index.

Over the evaluation period, January 4, 2016 to January 25, 2017 (268 observations), we assess the performance of the point forecasts of center and range for $h = 1, 2, 3$ days ahead. In Table 1.10, we report the values of RMSE, MAE and MDE losses associated with several forecasts of the center and range as well as the ACE measure. For both, SP500 and GOOG, we observe similar behavior. There are no differences across methods regarding the forecast of the center because the center equation is the same across methods and it does not need any transformation. In terms of RMSE, it is only for the range that the naive forecast is the worst performer, as expected, and the bias-corrected, smearing, and bootstrap forecasts are very much equivalent on delivering the same RMSE loss. Recall that the log-range is almost normal, so the Guerrero bias-correction is exact. However, in terms of MAE, the naive forecast is the best for the SP500 system and it is equivalent to the bias-corrected, smearing, and bootstrap forecasts for the GOOG system. In terms of MDE, there are not major differences across methods. In terms of ACE, we observe some minor differences. The bias-corrected, smearing, and bootstrap forecasts offer, on average, slightly more coverage of the realized intervals than the naive forecast. The coverage rates are large for $h = 1$, 52% (SP500) and 54% (GOOG) and they decrease with the forecasting horizon, i.e. for $h = 2$, 39% (SP500) and 38% (GOOG), and for $h = 3$, 34% (SP500) and 31% (GOOG). In Figures 1.4 and 1.5, we plot the time series of the one-day-ahead forecast intervals (smearing approach) compared to the realized intervals. With a few exceptions, we observe that the overlap between forecast and realized intervals is substantial.

SP500 Index

	t-statistics ($t_{k,\alpha}$)				
α	lag k				
	1	2	3	4	5
0.01	-2.55	-1.44	-2.99	-2.10	-2.76
0.05	-2.41	-0.62	-3.15	-1.55	-1.74
0.1	-1.60	0.31	-1.26	-0.20	-0.20
0.2	1.42	2.65	1.77	2.71	3.01
0.3	2.82	3.87	2.81	3.34	4.35
0.4	3.34	3.40	3.52	3.57	3.99
0.5	3.85	3.34	3.75	3.66	4.10
0.6	3.97	3.89	4.06	4.04	4.42
0.7	4.19	4.07	4.43	4.39	4.34
0.8	2.69	2.77	3.10	2.84	3.04
0.9	0.76	0.70	0.97	1.08	1.08
0.95	0.21	0.36	0.51	0.51	0.74
0.99	-3.64	-3.48	-3.31	-3.14	-2.97
C-statistic (C_k)	60.34	49.86	65.21	52.87	62.69

GOOG

	t-statistics ($t_{k,\alpha}$)				
α	lag k				
	1	2	3	4	5
0.01	-2.76	-2.76	-2.10	-3.42	-1.87
0.05	-2.02	-0.33	-1.17	-1.26	-1.44
0.1	2.74	3.73	2.23	2.77	2.44
0.2	7.08	6.48	5.84	6.79	6.33
0.3	8.53	8.06	7.76	8.57	7.95
0.4	8.70	7.84	7.71	7.98	8.18
0.5	7.30	6.89	6.56	6.69	6.99
0.6	6.61	6.17	6.05	5.89	6.31
0.7	5.44	5.43	5.42	5.49	5.25
0.8	4.04	3.94	4.02	4.06	3.83
0.9	1.14	1.02	1.19	1.18	1.00
0.95	-0.66	-0.66	-0.35	-0.36	-0.52
0.99	-3.65	-3.48	-3.32	-3.15	-2.98
C-statistic (C_k)	152.06	121.20	112.77	135.82	119.98

C_k aggregates all 13 autocontours for a given lag k; its 5% critical value is 22.36

Table 1.9: Generalized-AutoContouR (G-ACR) tests (González-Rivera and Sun, 2015) for SP500 and GOOG

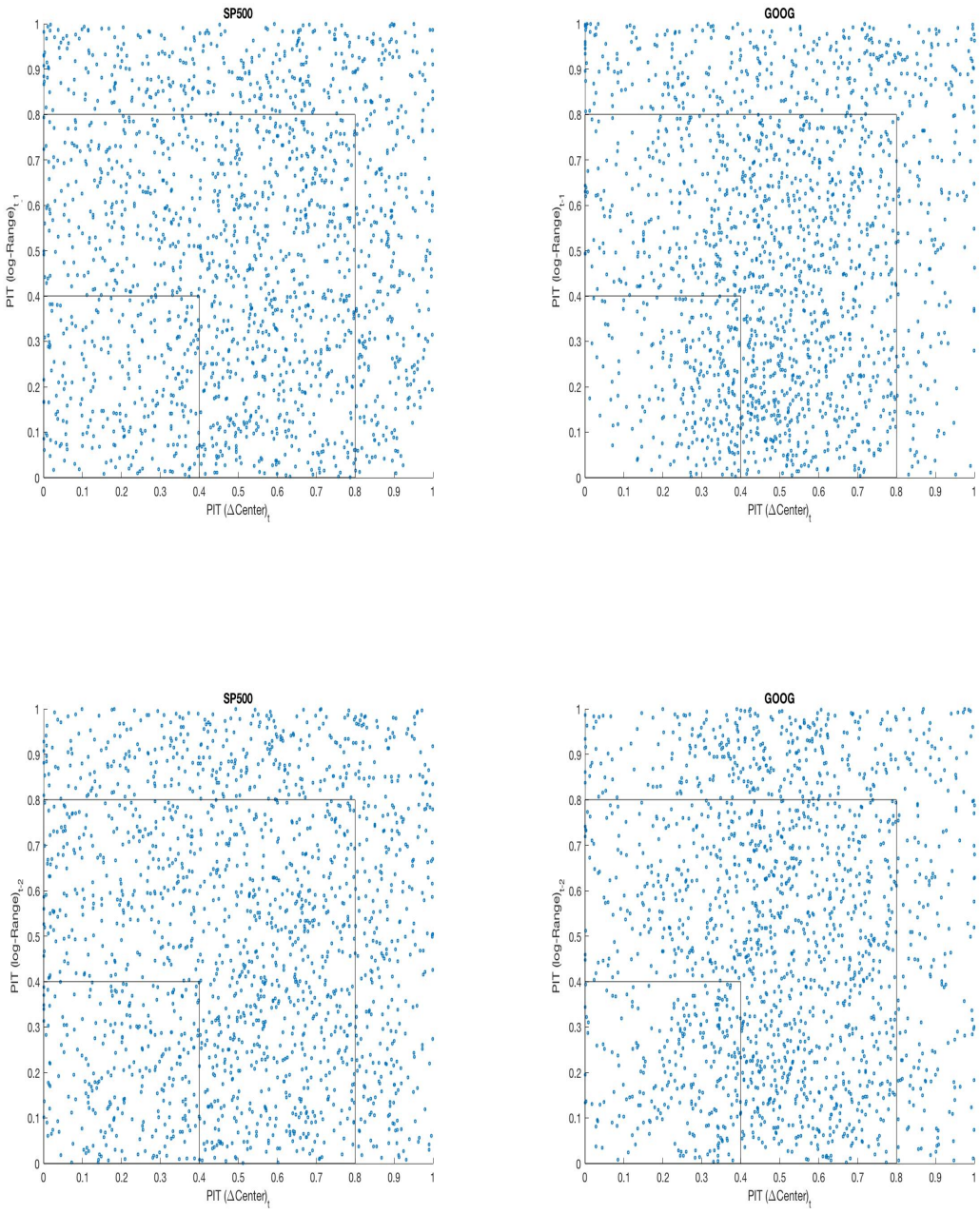


Figure 1.3: SP500 and GOOG. G-ACR specification tests for bivariate normality of first-differenced center and log-range. PITs autocontours

SP500

1-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	9.0925	12.7157	9.5614	6.2643	10.5232	0.5010
Guerrero	8.9864	12.7157	9.5614	6.4007	10.5606	0.5153
Smearing	8.9872	12.7157	9.5614	6.4052	10.5618	0.5155
Bootstrap	9.0070	12.7181	9.5767	6.4060	10.5656	0.5156
2-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	9.2584	20.0343	14.4127	6.3505	15.2939	0.3760
Guerrero	9.0845	20.0343	14.4127	6.4578	15.2953	0.3904
Smearing	9.0847	20.0343	14.4127	6.4632	15.2961	0.3907
Bootstrap	9.0910	20.0119	14.4044	6.4564	15.2984	0.3890
3-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	9.6236	24.7139	17.5703	6.6207	18.3425	0.3267
Guerrero	9.4354	24.7139	17.5703	6.7845	18.3592	0.3379
Smearing	9.4360	24.7139	17.5703	6.7918	18.3604	0.3381
Bootstrap	9.4291	24.6802	17.5748	6.7334	18.3546	0.3348

GOOG

1-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	5.5337	7.8902	5.7694	3.6162	6.3013	0.5256
Guerrero	5.3989	7.8902	5.7694	3.5985	6.3076	0.5375
Smearing	5.3963	7.8902	5.7694	3.6006	6.3084	0.5379
Bootstrap	5.3977	7.9077	5.7797	3.5948	6.3167	0.5366
2-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	5.7015	12.4717	8.9958	3.6068	9.4745	0.3651
Guerrero	5.5134	12.4717	8.9958	3.6171	9.4632	0.3777
Smearing	5.5093	12.4717	8.9958	3.6203	9.4633	0.3781
Bootstrap	5.5305	12.4986	9.0219	3.6114	9.4871	0.3759
3-step	RMSER	RMSEC	MAEC	MAER	MDE	ACE
Naïve	5.8619	15.8074	11.3515	3.7513	11.8010	0.2996
Guerrero	5.6739	15.8074	11.3515	3.7594	11.8110	0.3094
Smearing	5.6700	15.8074	11.3515	3.7641	11.8118	0.3097
Bootstrap	5.6933	15.8138	11.3669	3.7386	11.8209	0.3073

Table 1.10: SP500 and GOOG. Loss functions and ACE measure for center/range point forecasts. January 4, 2016 to January 25, 2017 (268 observations)

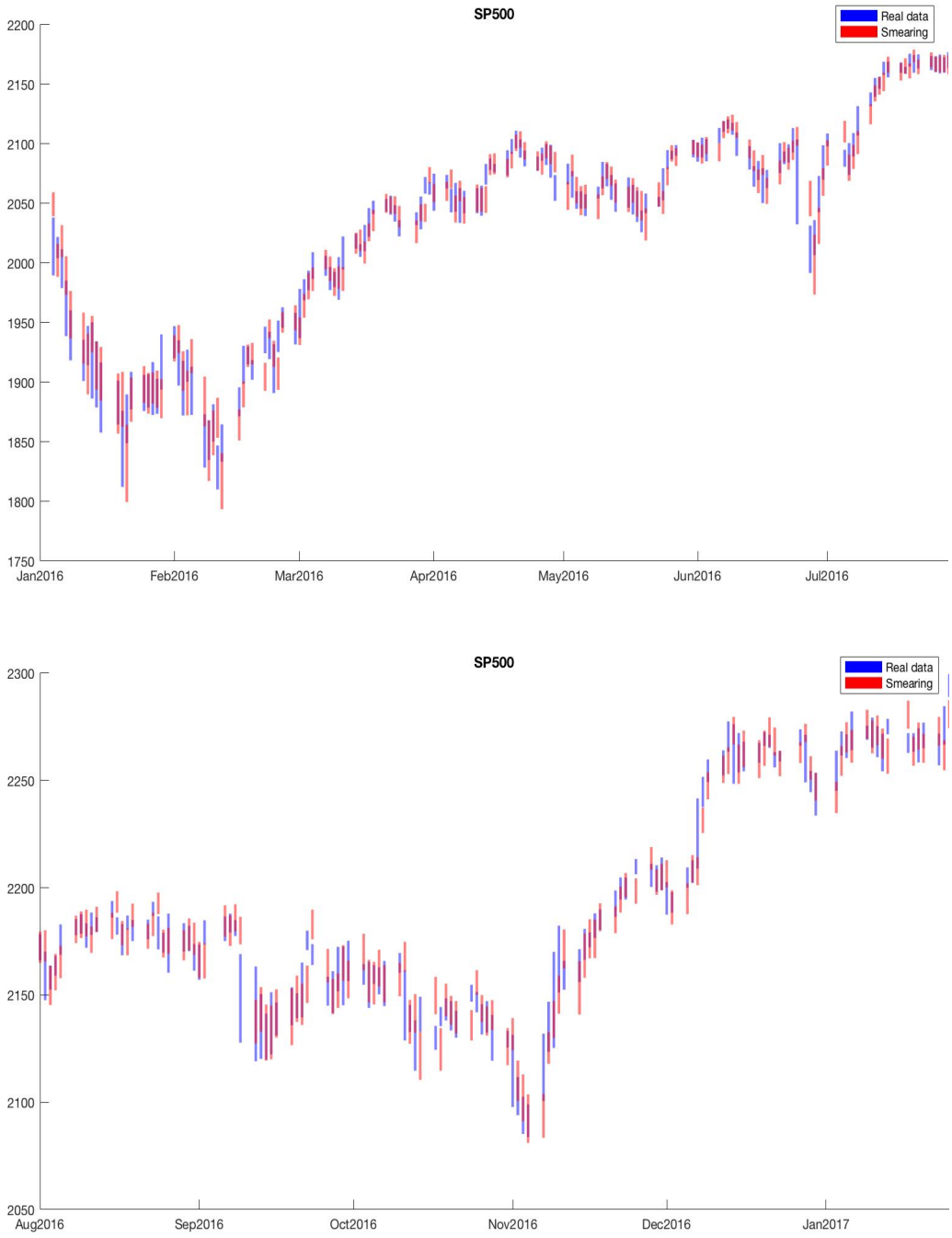


Figure 1.4: SP500. Daily one-step-ahead interval forecasts from January 4, 2016 to January 25, 2017

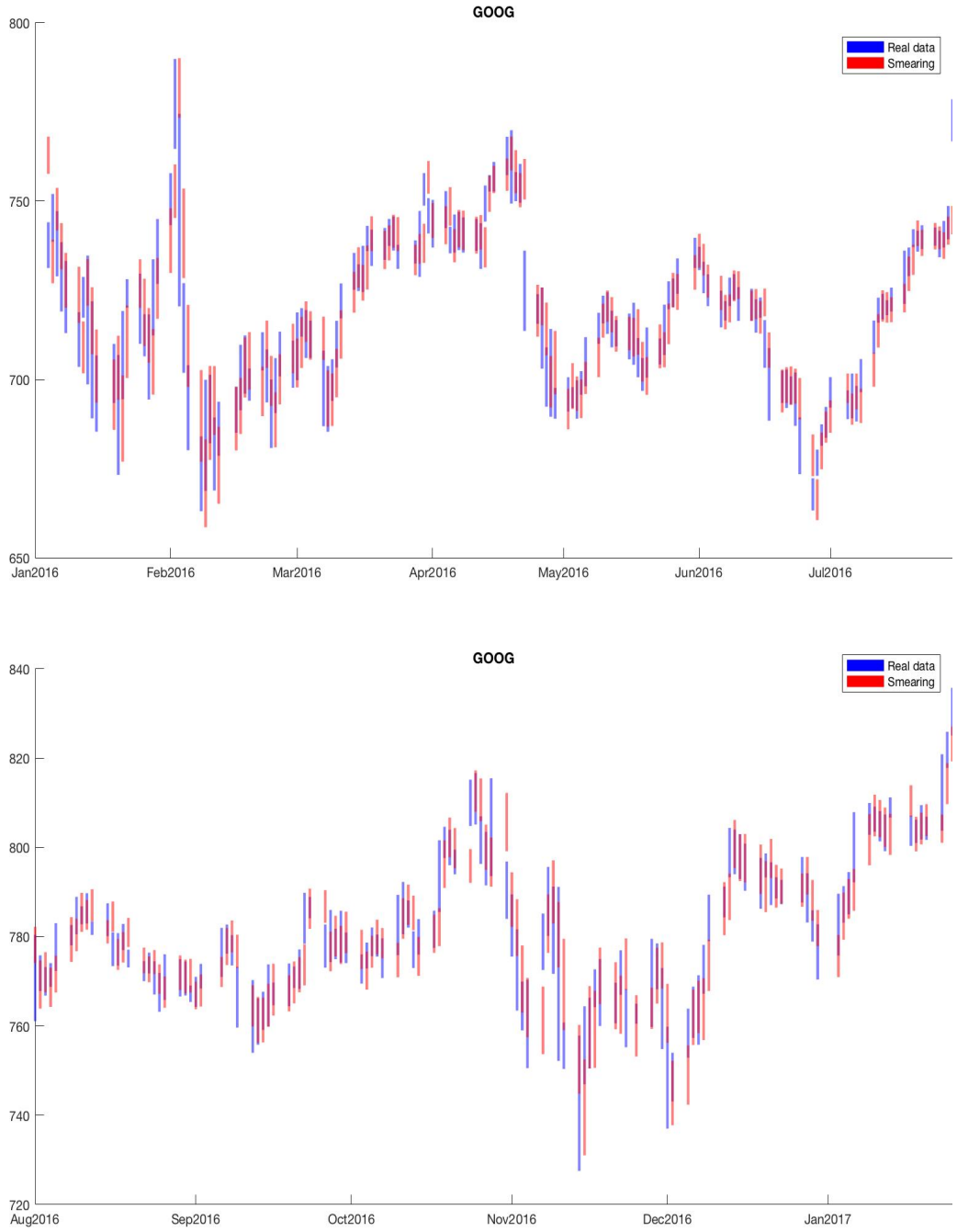


Figure 1.5: GOOG. Daily one-step-ahead interval forecasts from January 4, 2016 to January 25, 2017

Summarizing our findings:

1. The center and log-range are not bivariate normal. The center comes from a fat-tail density though the log-range is distributed close to normal. The log-range has strong memory though the first-differenced center has almost none. Features shared by the end-of-the-day returns and their volatility.

2. The center is Granger-causing the range with negative correlation; this is similar to the leverage effect in conditional volatility models.

3. The bias-corrected forecast works well because of the normality of the log-range but, if this is not the case, the smearing and bootstrap forecasts will be preferred. We find that there is substantial overlap between forecast and realized intervals up to 54%.

1.6 Conclusions

We have generated point forecast for an interval-valued time series. We started by estimating a Gaussian VAR(p) model for the center/log-range system because we avoid imposing the positive constraint in the range and the estimation is straightforward and delivers QMLE results. We have transformed the point forecasts to the center/range system by exploring and comparing several bias-corrected methods . We have shown the commonalities and differences between modeling the daily low/high price interval of the SP500 Index and GOOG stock and the more standard approach in financial econometrics of modeling the end-of-the-day return and its volatility process. The predictability of the interval is higher than the predictability of the end-of-the day return. We have found that, on average, one-step-ahead interval forecasts covered more than 50% of the realized intervals.

Bibliography

[1] Alizadeh, S., M. W. Brandt, and F.X. Diebold (2002), "Range-based estimation of stochastic volatility models", *Journal of Finance*, Vol. 57, No. 3, pp. 1047-1091.

[2] Ariño, M.A., and P. Franses (2000), "Forecasting levels of vector autoregressions log-transformed time series", *International Journal of Forecasting*, Vol. 16, pp. 111-116.

[3] Bardsen, G., and H. Lutkepohl (2011), "Forecasting levels of log variables in vector autoregressions", *International Journal of Forecasting*, Vol. 27, pp. 1108-1115.

[4] Blanco-Fernádez, A. and P. Winker (2016), "Data generation and statistical management of interval data", *AStA Advances in Statistical Analysis*, Vol. 10, No. 4, pp. 475-494.

[5] Duan, N. (1983), "Smearing estimate: A Nonparametric retransformation method", *Journal of the American Statistical Association*, Vol. 78, No. 383, pp. 605-610.

[6] Fresoli, D., E. Ruiz, and L. Pascual. (2015), "Bootstrap multi-step forecasts for non-Gaussian VAR models", *International Journal of Forecasting*, Vol. 31, pp. 834-848.

[7] González-Rivera, G., and W. Lin (2013), "Constrained Regression for Interval-valued Data," *Journal of Business and Economic Statistics*, Vol. 31, No. 4, pp. 473–490.

[8] Granger, C.W.J., and P. Newbold (1976), "Forecasting transformed series," *Journal of the Royal Statistics Society, Series B*, Vol. 38, pp. 189-203.

[9] Guerrero, V.M. (1993), "Time series analysis supported by power transformations" *Journal of Forecasting*, Vol. 12, pp. 37-48.

[10] Lima Neto, E., and F. de Carvalho (2010), "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics and Data Analysis*, Vol. 54, pp. 333-347.

[11] Mayr, J., and D. Ulbricht (2015), "Log versus level in VAR forecasting: 42 million empirical answers. Expect the unexpected", *Economics Letters*, Vol. 126, pp. 40-42.

[12] Parkinson, M. (1980), "The extreme value method for estimating the variance of the rate of return" *Journal of Business*, Vol. 53, No. 1, pp. 61-65.

[13] Pascual, L., J. Romo, and E. Ruiz, (2005), "Bootstrap prediction intervals for power-transformed time series", *International Journal of Forecasting*, Vol. 21, pp. 219-235.

[14] Tu, Y., and Y. Wang (2016), "Center and log range models for interval-valued data with an application to forecast stock returns", working paper.

[15] Yao, W., and Z. Zhao (2013), "Kernel density-based linear regression estimate", *Communications in Statistics. Theory and Methods*, Vol. 42, No. 24, pp. 4499-4512.

Chapter 2

Prediction Regions for Interval-valued Time Series

2.1 Introduction

Most of the econometric analysis of interval-valued data has focused on model estimation and inference, and though it is possible to construct point forecasts based on a given model or algorithm (e.g., chapter 1), the question of constructing probabilistic forecasts for interval data has not been addressed yet. This is the main question that we aim to analyze in this chapter. There are several routes to construct a probabilistic forecast for the lower/upper bounds system or for the center/range system, which involve some trade-offs between estimation and prediction decisions.

In this chapter, we contribute to the literature by approximating a probabilistic forecast for interval-valued time series. We offer alternative approaches to construct bi-

variate forecast regions of the center and the range (or lower and upper bounds) of the interval. We will start with a dynamic model for the center/log-range system. We specify a VAR system to be estimated by quasi-maximum likelihood (QML), maximizing a bivariate Gaussian density, that guarantees the consistency of the estimators.[1] We estimate only the center/log-range system, construct prediction regions for this system, and based on these estimates, construct prediction regions for the center/range system and for the upper/lower bounds system. We implement analytical and numerical approaches to move a prediction region for the center/log-range system to prediction regions for the other systems. If the center/log-range system is bivariate normally distributed, we obtain analytical forecast ellipsoids with a desired probability coverage. Furthermore, as proposed by Lutkephol (1991), we could also construct forecast regions by using Bonferroni rectangles, which are simpler and rather popular among practitioners. However, the center and/or the log-range are often not normally distributed and the joint system will not be bivariate normal. In these cases, we obtain forecasts of the center/log-range system using the bootstrap procedure proposed by Fresoli et al. (2015) for VAR models, which does not require any specific assumption on the forecast error distribution. After obtaining bootstrap replicates of future values of the center/log-range system, we construct forecast regions as ellipsoids, Bonferroni rectangles, or using the Tukey peeling. Implementing either analytical or bootstrap methods, the prediction regions constructed for the center/log-range system can be directly transformed into prediction regions for the center/range system. For instance, consider a normal ellipse with $(1-\alpha)\%$ probability coverage. The boundary of this ellipse is the $(1-\alpha)\%$ bivariate

[1] Tu and Wang (2016) used the estimator of Yao and Zhao (2013) that relies on kernel estimates of the likelihood. This estimator is computationally more demanding than QML and depends on the choice of tuning parameters. Their empirical results suggest that both estimators are very similar and, consequently, we focus on the QML estimator.

quantile. Its boundary points (center, log-range) can be transformed into another boundary of points (center, exp(log-range)) of a prediction region for the center/range system. The new region will not preserve the shape of an ellipse but it will have the same coverage because the exponential function is a monotonic transformation. An important advantage of our approach is that, by focusing on prediction regions rather than on point forecasts, we avoid the biases that are associated with the exp-transformation of the point forecasts of log-transformed variables, for which a bias correction is necessary to obtain the conditional mean of the variable of interest[2]; see, for example, Granger and Newbold (1976) and Guerrero (1993).

We compare the performance of the prediction regions considered in this chapter according to several metrics. The most basic required property is coverage so that regions are reliable when the empirical coverage is close to the nominal coverage. Beyond coverage, the literature on evaluating multivariate prediction regions is rather thin. To our knowledge, there is one additional metric that brings the volume of the region to interact with its coverage (Golestaneh *et al.*, 2017). In this chapter, we also contribute to this literature by introducing several new measures that account for (i) the location of out-of-the-region points with respect to a central point of the region, (ii) the tightness of the intervals that result from projecting the two-dimensional region into one-dimensional intervals, and (iii) the distance of the also projected out-of-the-region points to the projected one-dimensional

[2]For point forecasts of Gaussian VAR models, Ariño and Franses (2000) and Bardsen and Lutkepohl (2011) give explicit expressions for the optimal point forecasts of the levels when both variables are log-transformed. Furthermore, Bardsen and Lutkepohl (2011) show that, despite its theoretical advantages, optimal point forecasts are inferior to naive forecasts if specification and estimation uncertainty are taken into account. Hence, they conclude that, in practice when the interest is a point forecast, using the exponential of the log-forecasts is preferable to using the optimal forecasts; see also Mayr and Ulbricht (2015) for an empirical application to forecasting GDP. Finally, it is important to point out that the optimal transformations are not designed to obtain density forecasts.

interval. These new measures bring a notion of risk associated with the prediction region. In addition, we also provide a description of the distribution of the out-of-the-region points around the region to measure whether the region is probability-centered.

For the three systems (center/log-range, center/range, and upper/lower bounds), we perform several Monte Carlo simulations to assess the out-of-sample performance of the prediction regions constructed with analytical and bootstrap methods. We evaluate bivariate Gaussian and non-Gaussian center/log-range systems and their implied distributions for the center/range and upper/lower bounds systems. We note that even for Gaussian systems, bootstrap methods to construct ellipsoids and Bonferroni rectangles deliver the best performance, mainly when the estimation sample is small and estimation uncertainty is most relevant. For non-Gaussian systems, the performance depends on whether the joint distribution of the center/log-range system is symmetric or not. If symmetry is present, bootstrap ellipsoids and their transformations are recommended. For asymmetric non-Gaussian systems, bootstrap Bonferroni rectangles are preferred.

Using the analytical and bootstrap procedures described above, we construct forecast regions for a time series of daily low/high return intervals of the SP500 index. These intervals are more informative than just a daily one-point measurement (end-of-day return) as they encompass all returns during the day. There are commonalities between the analysis of return intervals and the standard analysis of end-of-the-day returns and their volatility. The center of the return interval has large kurtosis and does not have any autocorrelation. The log-range, which is close to be normally distributed, is a proxy for volatility as proposed by Parkinson (1980) and Alizadeh *et al.* (2002). It shows a strong autocorrelation

as that of an autoregressive process, which is similar to the patterns found in ARCH and stochastic volatility processes. We also find that there is Granger-causality from the center of the interval to the log-range such that positive and large changes in the center will predict narrower ranges, which is similar to the so-called leverage effect. However, an important difference pertains to the construction of the forecasts. In standard ARCH and stochastic volatility processes, the forecast of the return is mostly zero and together with a forecast of the conditional volatility and some conditional distribution of the return, it is possible to generate a density forecast of future returns. In the interval approach, we forecast jointly the future low/high return interval and construct prediction regions of the center and range of the interval at any desired horizon that do not require parametric distributional assumptions. Overall, the main advantage of the interval approach is that allows for the modeling of the joint conditional density of the return level and the return volatility, which in our sample are contemporaneous and negatively correlated, and consequently allows for the construction of bivariate density forecasts. We develop a new trading strategy by extending the strategy by He et al. (2010) for point forecasts of high/low prices to account for the density forecasts of high/low returns. We found this strategy profitable in the out-of-sample evaluation environment.

The organization of the chapter is as follows. In section 2, we establish notation by describing the VAR model for the center/log-range system, its estimation and construction of point forecasts. In section 3, we present analytical prediction regions for a Gaussian center/log-range system and how they translate into those for the center/range and upper/lower bounds systems. In section 4, we introduce bootstrap procedures to deal with

prediction regions for non-Gaussian center/log-range systems and their implications for those regions in the center/range and upper/lower bounds systems. In section 5, we propose several new metrics to evaluate the performance of the different prediction regions. In section 6, we report Monte Carlo simulations to compare the performance of the proposed procedures to construct forecast regions. In section 7, we model the SP500 low/high return interval and construct several prediction regions for the interval, based on which we develop a trading strategy. We conclude in section 8.

2.2 The Center/Log-Range System

Even if the final goal is to obtain probabilistic forecasts of the center/range or lower/upper bounds systems, we start by estimating a dynamic model for the center/log-range system that is not subject to any restriction as we are log-transforming the range. We consider a linear bivariate VAR(p) for the center/log-range system from which we will construct a probabilistic forecast for $(C_t, \log R_t)$. Let us call $y_{c,t} \equiv C_t$ and $y_{r,t} \equiv \log R_t$. The bivariate VAR(p) is given by

$$y_{c,t} = \alpha_1 + \sum_{i=1}^{p} \beta_{11}^{(i)} y_{c,t-i} + \sum_{i=1}^{p} \beta_{12}^{(i)} y_{r,t-i} + \varepsilon_{c,t} \tag{2.1}$$

$$y_{r,t} = \alpha_2 + \sum_{i=1}^{p} \beta_{21}^{(i)} y_{c,t-i} + \sum_{i=1}^{p} \beta_{22}^{(i)} y_{r,t-i} + \varepsilon_{r,t} \tag{2.2}$$

where the components of the error vector $(\varepsilon_{c,t}, \varepsilon_{r,t})'$ are white noise processes, possibly contemporaneous correlated, with covariance matrix Ω. The estimation of the parameters of the VAR(p) model proceeds by LS, which is consistent under mild assumptions.

The LS estimator is a full information ML estimator when the errors have a bivariate normal distribution. Otherwise, if the errors are non-normal, a QML estimator based on maximizing the Gaussian likelihood will be equivalent to a LS estimator. Let $\theta \equiv (\alpha_1, \alpha_2, \beta_{11}^{(1)}, ..., \beta_{11}^{(p)}, \beta_{12}^{(1)}, ..., \beta_{12}^{(p)}, \beta_{21}^{(1)}, ..., \beta_{21}^{(p)}, \beta_{22}^{(1)}, ..., \beta_{22}^{(p)})$ be the parameter vector to estimate. Following White (1982), the asymptotic distribution of the Gaussian QML estimator is $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, A^{-1}BA^{-1})$ where matrix A is the (minus) expectation of the Hessian and matrix B is the expectation of the outer product of the score of a Gaussian log-likelihood function. The QML environment will be the most common estimation approach given that bivariate normality of $(\varepsilon_{c,t}, \varepsilon_{r,t})'$ is difficult to entertain. To guarantee bivariate normality of the system, the conditional densities as well as the marginal densities must also be normal density functions. For financial data, there is evidence that the log-range $y_{r,t}$ (as a proxy for volatility) is near-normal (Alizadeh *et al.*, 2002). However, the center $y_{c,t}$ is less likely to be normally distributed because the prevalence of fat tails, at least in financial data at a relative high frequency, e.g. daily financial returns. In the empirical section, we will test the assumption of bivariate normality as a starting step to construct density forecasts of the full system.

Given an information set available at time T, if the loss function is quadratic, the optimal h-step-ahead point forecasts of the system $(y_{c,t}, y_{r,t})$ are the conditional means denoted by $y_{c,T+h|T}$ and $y_{r,T+h|T}$. Since the VAR(p) model is always invertible, the conditional mean is a linear function of the observations. Therefore, point forecasts of the center

and log-range are given by

$$y_{c,T+h|T} = \alpha_1 + \sum_{i=1}^{p} \beta_{11}^{(i)} y_{c,T+h-i|T} + \sum_{i=1}^{p} \beta_{12}^{(i)} y_{r,T+h-i|T} \tag{2.3}$$

$$y_{r,T+h|T} = \alpha_2 + \sum_{i=1}^{p} \beta_{21}^{(i)} y_{c,T+h-i|T} + \sum_{i=1}^{p} \beta_{22}^{(i)} y_{r,T+h-i|T} \tag{2.4}$$

where $y_{c,T+h-i|T} = y_{c,T+h-i}$ and $y_{r,T+h-i|T} = y_{r,T+h-i}$ for $i \geq h$. The corresponding forecast error vector is $(e_{c,T+h|T}, e_{r,T+h|T}) = (y_{c,T+h} - y_{c,T+h|T}, y_{r,T+h} - y_{r,T+h|T})$ with variance-covariance matrix $W_h = \Omega + \sum_{i=1}^{h-1} \Psi_i \Omega \Psi_i'$ where matrices Ψ_i come from the MA(∞) representation of the VAR(p) model. In practice, we plug in consistent estimates, i.e. $\hat{\theta}, \hat{\Omega}$, and $\hat{\Psi}_i$, in the VAR(p) to obtain the estimated h-step-ahead point forecasts and their estimated variance-covariance matrices that are denoted by $\hat{y}_{c,T+h|T}$, $\hat{y}_{r,T+h|T}$, and $\hat{W}_h$ respectively.

If the center/log-range system is bivariate normal, then pointwise bivariate density forecasts can be obtained as follows,

$$\begin{bmatrix} y_{c,T+h} \\ y_{r,T+h} \end{bmatrix} \rightarrow N\Big(\begin{bmatrix} \hat{y}_{c,T+h|T} \\ \hat{y}_{r,T+h|T} \end{bmatrix}, \begin{bmatrix} \hat{W}_{h,11} & \hat{W}_{h,12} \\ \hat{W}_{h,21} & \hat{W}_{h,22} \end{bmatrix} \Big) \tag{2.5}$$

Note that the variance-covariance matrices of the forecast densities in (2.5) do not incorporate parameter uncertainty, which will be negligible when the sample size T is large relative to the number of estimated parameters. When the center/log-range system is non-Gaussian, we can obtain bootstrap pointwise forecast densities by implementing the bootstrap procedure proposed by Fresoli, Ruiz, and Pascual (2015), which will be described in the forthcoming section 4. The bootstrap forecast densities incorporate parameter un-

certainty without relying on any specific forecast error distribution. Even in the Gaussian case, if the estimation sample is not very large, the effect of parameter uncertainty on the forecast may not vanished, and so the use of bootstrap forecast densities may be desired.

2.3 Gaussian Center/Log-Range System

2.3.1 Prediction regions for the center/log-range system

Using the forecast densities in (2.5), we can construct pointwise h-step-ahead forecast regions. The $100 \times (1-\alpha)\%$ h-step-ahead forecast ellipsoid for $Y_{T+h} \equiv (y_{c,T+h}, y_{r,T+h})'$ is given by

$$NE_{T+h} = \left[Y_{T+h} | (Y_{T+h} - \hat{Y}_{T+h|T})' \hat{W}_h^{-1} [Y_{T+h} - \hat{Y}_{T+h|T}]\right] \leq q_{1-\alpha}, \tag{2.6}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-square distribution with 2 degrees of freedom. The ellipse is a countour of the bivariate normal center/log-range system with $100 \times (1-\alpha)\%$ coverage.

A straightforward and easy to construct h-step-ahead forecast region is a Bonferroni rectangle with (at least) $100 \times (1-\alpha)\%$ coverage. This rectangle will have the following sides

$$\left[b_{c,\alpha/4}, b_{c,1-\alpha/4}\right] \equiv \left[\hat{y}_{c,T+h|T} - z_{\alpha/4}\sqrt{\hat{W}_{h,11}}, \quad \hat{y}_{c,T+h|T} + z_{\alpha/4}\sqrt{\hat{W}_{h,11}}\right] \tag{2.7}$$

$$\left[b_{r,\alpha/4}, b_{r,1-\alpha/4}\right] \equiv \left[\hat{y}_{r,T+h|T} - z_{\alpha/4}\sqrt{\hat{W}_{h,22}}, \quad \hat{y}_{r,T+h|T} + z_{\alpha/4}\sqrt{\hat{W}_{h,22}}\right], \tag{2.8}$$

where $z_{\alpha/4}$ is the $\alpha/4$-quantile of the standard normal distribution. Given the bivariate normality of the system (2.5), the marginal probability density functions of $y_{c,T+h}$ and $y_{r,T+h}$ are also normal.

To include the contemporaneous linear correlation between the center and log-range, we modify the Bonferroni rectangles as in Fresoli *et al.* (2015). The corners of the modified rectangle are

$$[b_{c,\alpha/4}, \quad b_{r,\alpha/4} + p_{21,h} b_{c,\alpha/4}], \qquad [b_{c,\alpha/4}, \quad b_{r,1-\alpha/4} + p_{21,h} b_{c,\alpha/4}], \tag{2.9}$$

$$[b_{c,1-\alpha/4}, \quad b_{r,\alpha/4} + p_{21,h} b_{c,1-\alpha/4}], \qquad [b_{c,1-\alpha/4}, \quad b_{r,1-\alpha/4} + p_{21,h} b_{c,1-\alpha/4}]$$

where $p_{21,h} = \hat{W}_{h,21}/\hat{W}_{h,11}$. The area of the modified Bonferroni rectangle is the same as that of the Bonferroni rectangle. However, the theoretical coverage rate may be slightly different depending on the quantiles associated with the modified terms, e.g., $b_{r,\alpha/4}+p_{21,h}b_{c,\alpha/4}$, which in turn depend on the magnitude and sign of $p_{21,h}$. Simulations results will provide some information on the coverage rate of the modified Bonferroni rectangle. To illustrate the shapes of the three forecast regions described above, in Figure 2.1 we plot the 1-step-ahead 95% ellipse, Bonferroni rectangle and modified Bonferroni rectangle for the center/log-range system generated by a VAR(4) model with parameter values as reported in Table 1 and Gaussian errors with contemporaneous correlations of -0.24. The forecast regions have been obtained after estimating the parameters based on T=1000 observations so that the parameter estimation uncertainty is negligible. In Figure 2.1, we also plot 1000 realizations of Y_{T+1}. We observe that both the ellipse and the modified Bonferroni rectangle are able to capture the negative correlation between the center and the log-range while the Bonferroni

rectangle cannot inform about this correlation. Note that the Bonferroni rectangle has large empty areas without any realization of Y_{T+1}.

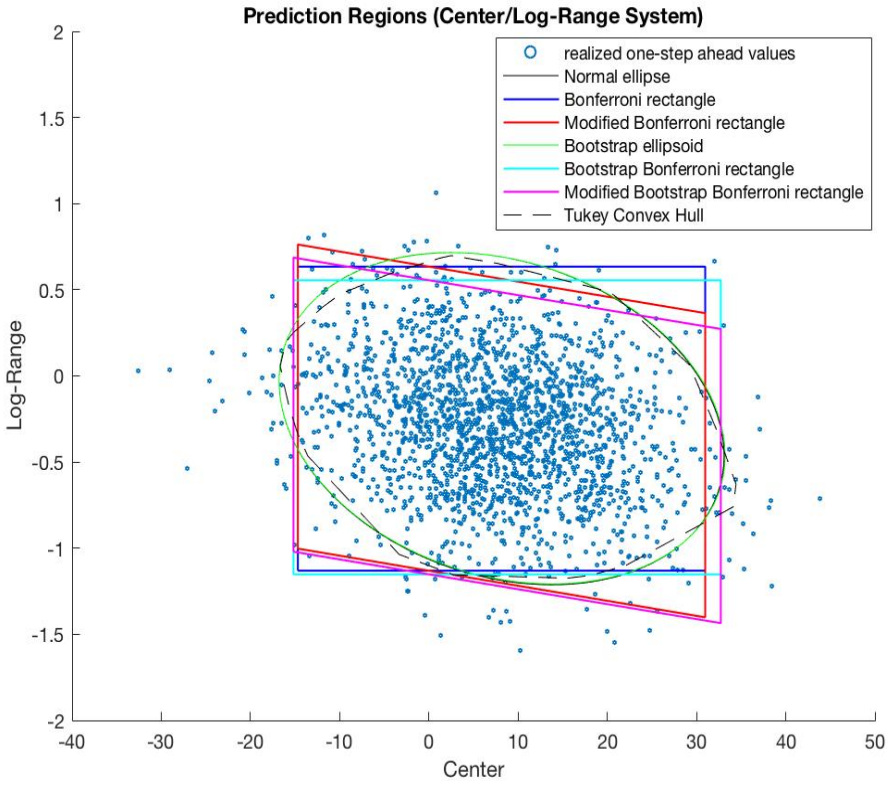


Figure 2.1: 95% prediction regions for the center/log-range system obtained from a simulated VAR(4) model with Gaussian errors and $T = 1000$.

2.3.2 Prediction regions for center/range and lower/upper systems

Moving from the center/log-range system to the center/range system or to the lower/upper bounds system, we can implement either analytical or numerical methods to construct prediction regions for the center/range system or for the lower/upper bounds

system. Under bivariate normality of center/log-range, the bivariate density of the center/range system is

$$f(y_{c,T+h}, R_{T+h}) = \frac{1}{2\pi\sqrt{|\hat{W}_h|}} \frac{1}{R_{T+h}} \exp[-\frac{1}{2}(Y_{T+h} - \hat{Y}_{T+h|T})'\hat{W}_h^{-1}(Y_{T+h} - \hat{Y}_{T+h|T})]. \quad (2.10)$$

Since the center of the interval $y_c \equiv (y_u + y_l)/2$ and the range $R \equiv (y_u - y_l)$ are linear combinations of the upper and lower bounds, it is easy to see that that the conditional bivariate density of the upper/lower bounds is also given by (2.10).

We construct analytical contours for the center/range and lower/upper bounds system by horizontally cutting the bivariate density (2.10) at a value determined by the nominal coverage $100 \times (1 - \alpha)\%$ that we wish to obtain. Such a value is obtained by numerical simulation. Based on the same simulated system described above, in Figure 2.2 we illustrate the shape of the forecast regions for the center/range system obtained using the analytical density in (2.10) by plotting the 95% forecast region and 1000 realizations of (C_{T+1}, R_{T+1}). We observe that, as expected, the region is not an ellipse.

As an illustration of the shapes of the regions for the lower/upper bounds system, in Figure 2.3 we plot the 95% forecast regions based on (2.10) and a close-up detail of the central area of the region. In Figure 2.4, we plot close-ups of the extreme areas of the regions.

For the center/range system, we also construct numerical contours based on the $100\times(1-\alpha)\%$ normal ellipse (2.6) of the center/log-range system by transforming the points (center, log-range) sitting on the boundary of (2.6) to points (center, exp(log-range)). The

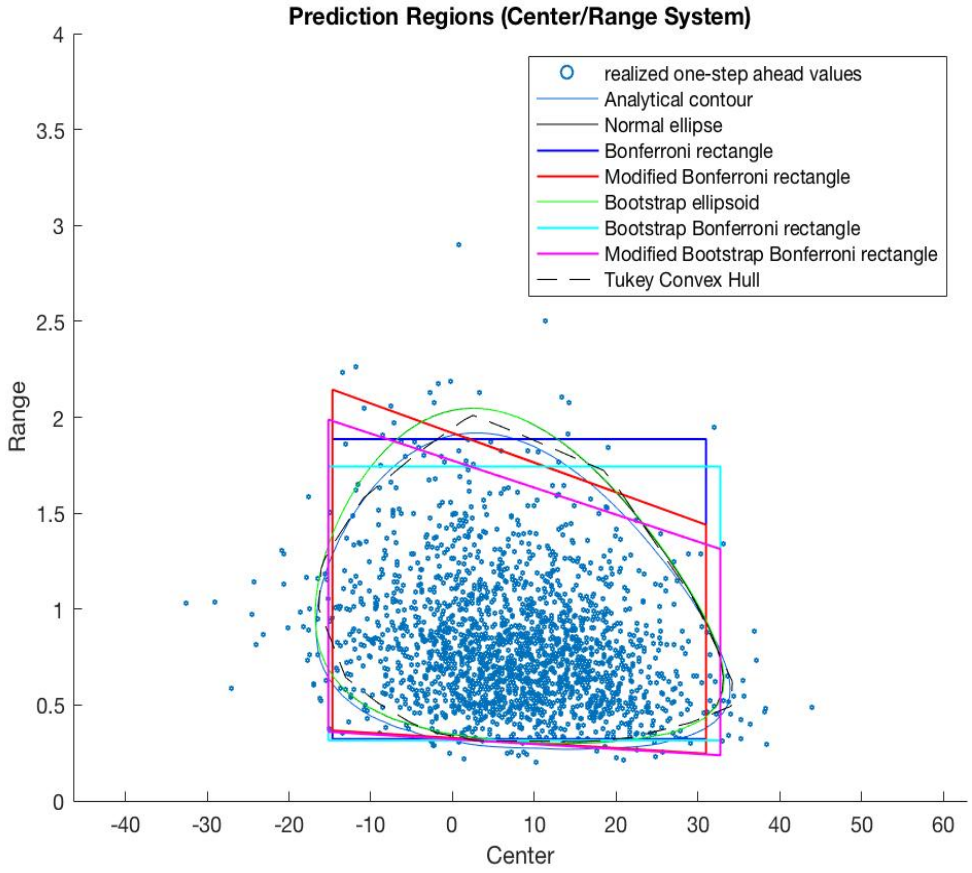


Figure 2.2: 95% prediction regions for the center/range system obtained by transforming the regions obtained for the center/log-range system as well as the analytical contour based on (2.10). Normal ellipse refers to the transformed normal ellipse T-NE and Bootstrap ellipse refers to the transformed bootstrap ellipse T-BE, which are identical.

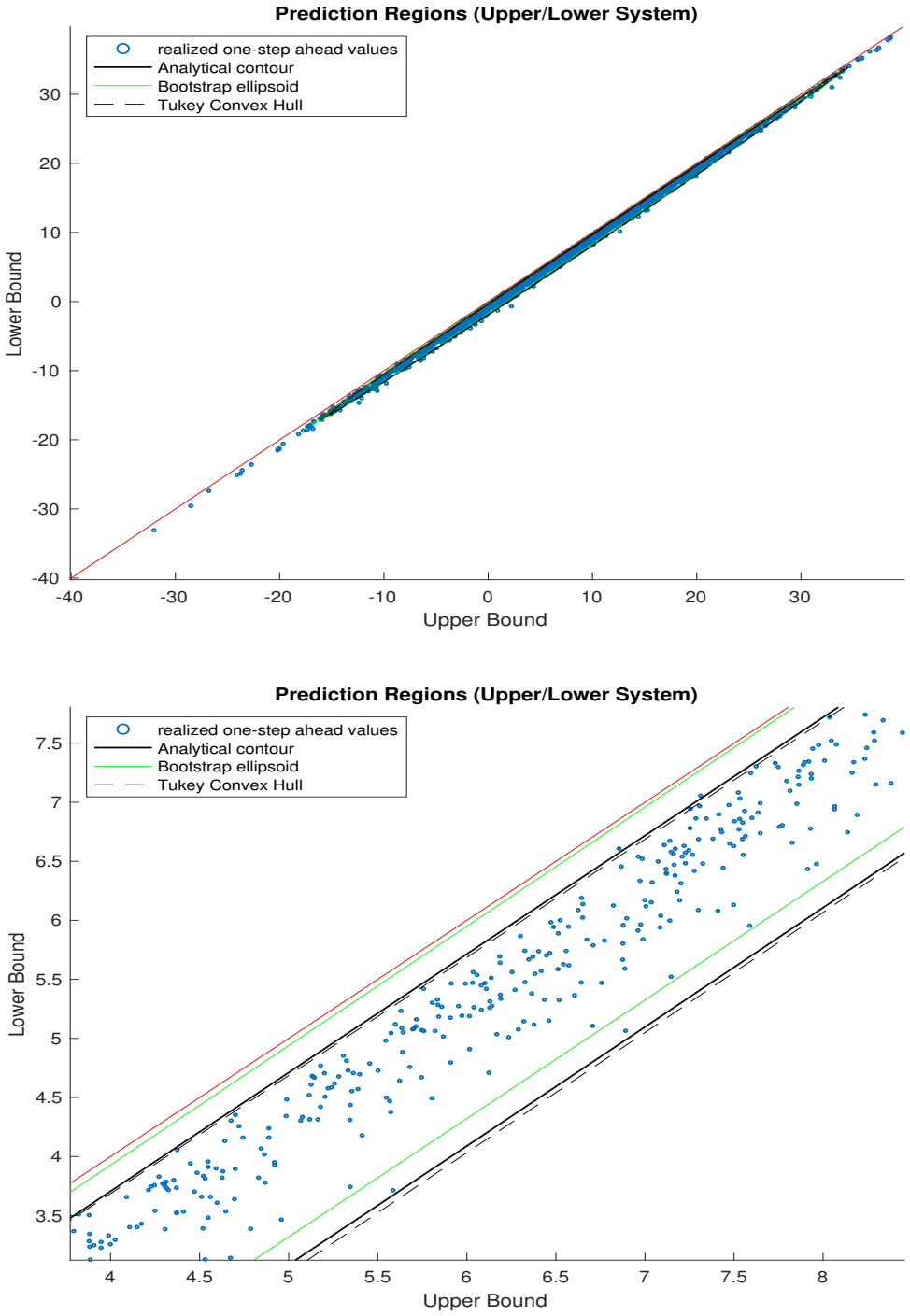


Figure 2.3: 95% prediction regions for the upper/lower bounds system. The lower panel is a close-up of the central area of the regions.

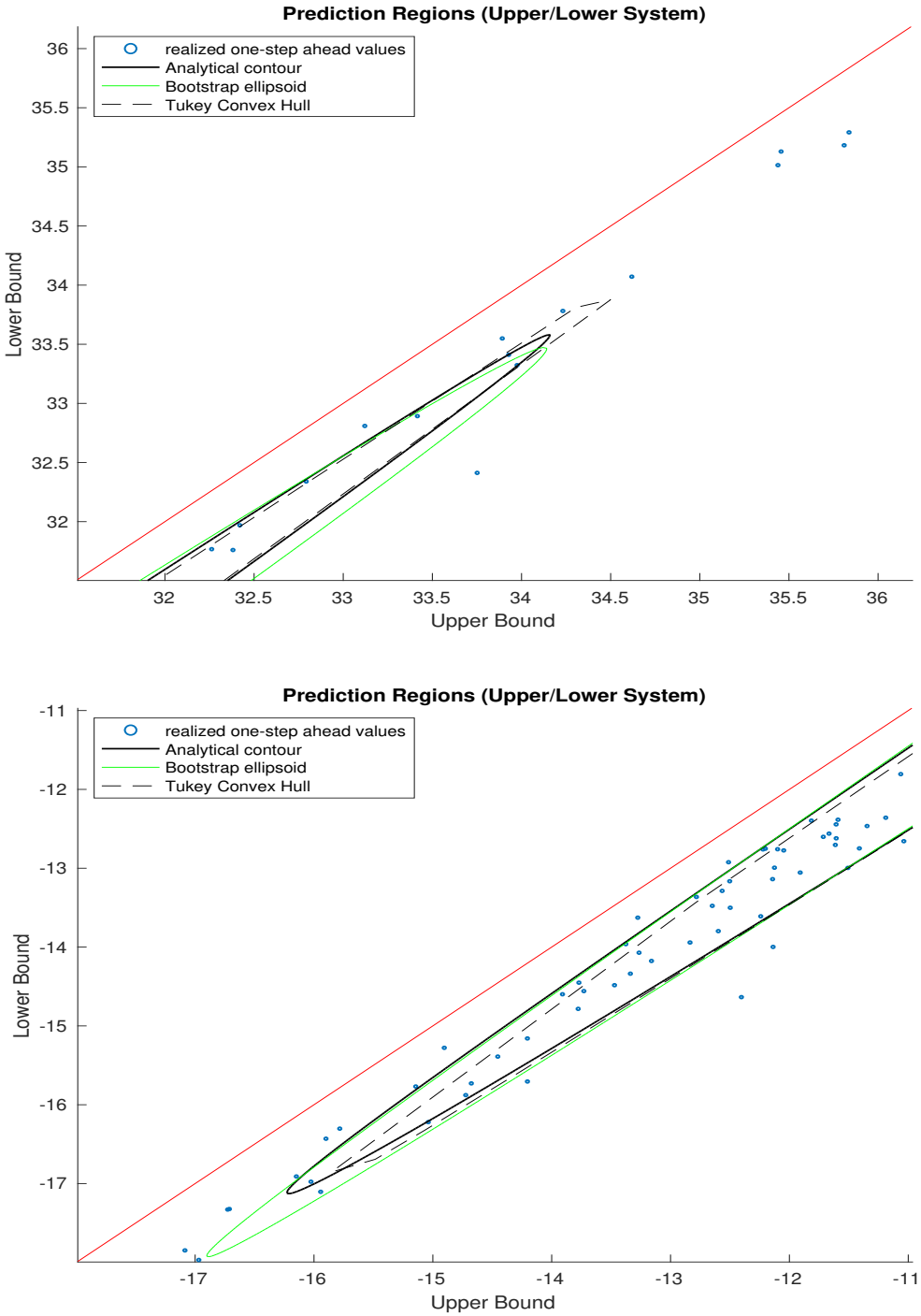


Figure 2.4: 95% prediction regions for the upper/lower bounds system. Detail of the extreme areas of the regions.

new shapes will not be ellipsoids but they will maintain the coverage, and have the advantage of delivering strictly positive values for the range[3].

The $100 \times (1-\alpha)\%$ transformed normal ellipse (T-NE) is given by

$$\text{T-NE}_{T+h} = \big\{ \big[(y_{c,T+h}, \exp[\log R_{T+h}])'\big] \quad \text{such that} \quad (Y_{T+h} - \hat{Y}_{T+h|T})'\hat{W}_h^{-1}(Y_{T+h} - \hat{Y}_{T+h|T}) = q_{1-\alpha} \big\} \tag{2.11}$$

In Figure 2.2, we illustrate the shape of the transformed ellipse using the same simulated example previously described. The transformed shape is similar to the analytical although are not identical.

Similarly, we transform the Bonferroni and modified Bonferroni rectangles by taking the exponential transformation of the log-range intervals (2.8) and the range terms in (2.9) respectively.

Transformed Bonferroni rectangle:

$$\big[\hat{y}_{c,T+h|T} - z_{\alpha/4}\sqrt{\hat{W}_{h,11}}, \quad \hat{y}_{c,T+h|T} + z_{\alpha/4}\sqrt{\hat{W}_{h,11}}\big] \tag{2.12}$$
$$\big[\exp(\hat{y}_{r,T+h|T} - z_{\alpha/4}\sqrt{\hat{W}_{h,22}}), \quad \exp(\hat{y}_{r,T+h|T} + z_{\alpha/4}\sqrt{\hat{W}_{h,22}})\big]$$

Transformed modified Bonferroni rectangle:

$$[b_{c,\alpha/4}, \quad \exp(b_{r,\alpha/4} + p_{21,h}b_{c,\alpha/4})], \qquad [b_{c,\alpha/4}, \quad \exp(b_{r,1-\alpha/4} + p_{21,h}b_{c,\alpha/4})], \tag{2.13}$$
$$[b_{c,1-\alpha/4}, \quad \exp(b_{r,\alpha/4} + p_{21,h}b_{c,1-\alpha/4})], \qquad [b_{c,1-\alpha/4}, \quad \exp(b_{r,1-\alpha/4} + p_{21,h}b_{c,1-\alpha/4})]$$

[3]This approach cannot be implemented to find prediction regions for the lower/upper bounds system because there is not a monotonic transformation from the boundary points of the center/log-range region to the boundary points of the lower/upper bounds region.

In Figure 2.2, we illustrate the shapes of the transformed Bonferroni rectangles. Observe that while the transformed modified Bonferroni rectangle also shows the correlation between center and range, the transformed Bonferroni rectangle does not and some portions of the area are empty.

2.4 Non-Gaussian Center/Log-Range System

2.4.1 Prediction regions for the center/log-range system

Following Fresoli *et al.* (2015), we implement the following bootstrap procedure to obtain bootstrap forecasts of the center/log-range system:

Step 1. Estimate the parameters of the VAR(p) model in (2.1)-(2.2) by LS and obtain the residual vector $\hat{\varepsilon}_t = (\hat{\varepsilon}_{c,t}, \hat{\varepsilon}_{r,t})'$. Center the residuals, i.e. $\hat{\varepsilon}_t - \bar{\varepsilon}_t$ where $\bar{\varepsilon}_t = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\varepsilon}_t$. Rescale the residuals using the factor $[\frac{T-p}{T-p-d}]^{1/2}$, where d is the number of parameters to estimate. Denote the empirical distribution of the centered and rescaled residuals as $\hat{F\varepsilon}$.

Step 2. Using the parameter estimates obtained in Step 1, generate in sample bootstrap series $\{y_{c,1}^{*(b)}, ..., y_{c,T}^{*(b)}\}$ and $\{y_{r,1}^{*(b)}, ..., y_{r,T}^{*(b)}\}$, for $t = 1, ..., T$, as follows,

$$y_{c,t}^{*(b)} = \hat{\alpha}_1 + \sum_{i=1}^{p}\hat{\beta}_{11}^{(i)} y_{c,t-i}^{*(b)} + \sum_{i=1}^{p}\hat{\beta}_{12}^{(i)} y_{r,t-i}^{*(b)} + \varepsilon_{c,t}^{*(b)}$$
$$y_{r,t}^{*(b)} = \hat{\alpha}_2 + \sum_{i=1}^{p}\hat{\beta}_{21}^{(i)} y_{c,t-i}^{*(b)} + \sum_{i=1}^{p}\hat{\beta}_{22}^{(i)} y_{r,t-i}^{*(b)} + \varepsilon_{r,t}^{*(b)},$$

where $(\varepsilon_{c,t}^{*}, \varepsilon_{r,t}^{*})'$ are random pairwise extractions with replacement from $\hat{F}_{\hat{\varepsilon}}$ and, for $t =$

$1, ..., p$, $y_{c,t}^{*(b)} = y_{c,t}$ and $y_{r,t}^{*(b)} = y_{r,t}$.[4] Using $y_{c,t}^{*(b)}, y_{r,t}^{*(b)}$, estimate the VAR(p) parameters $\hat{\alpha}^{*(b)} = \begin{bmatrix} \hat{\alpha}_1^{*(b)} \\ \hat{\alpha}_2^{*(b)} \end{bmatrix}$ and $\hat{\beta}^{*(b)} = \begin{bmatrix} \hat{\beta}_{11}^{*(i)(b)} & \hat{\beta}_{12}^{*(i)(b)} \\ \hat{\beta}_{21}^{*(i)(b)} & \hat{\beta}_{22}^{*(i)(b)} \end{bmatrix}_{i=1,...,p}$.

Step 3. Construct bootstrap h-step-head future values of the vector $(y_{c,T+h}, y_{r,T+h})'$ as follows,

$$\hat{y}_{c,T+h|T}^{*(b)} = \hat{\alpha}_1^{*(b)} + \sum_{i=1}^{p} \hat{\beta}_{11}^{*(i)(b)} \hat{y}_{c,T+h-i|T}^{*(b)} + \sum_{i=1}^{p} \hat{\beta}_{12}^{*(i)(b)} \hat{y}_{r,T+h-i|T}^{*(b)} + \varepsilon_{c,T+h}^{*(b)}$$
$$\hat{y}_{r,T+h|T}^{*(b)} = \hat{\alpha}_2^{*(b)} + \sum_{i=1}^{p} \hat{\beta}_{21}^{*(i)(b)} \hat{y}_{c,T+h-i|T}^{*(b)} + \sum_{i=1}^{p} \hat{\beta}_{22}^{*(i)(b)} \hat{y}_{r,T+h-i|T}^{*(b)} + \varepsilon_{r,T+h}^{*(b)},$$

where $\hat{y}_{c,T+h-i|T}^{*(b)} = y_{c,T+h-i}$, and $\hat{y}_{r,T+h-i|T}^{*(b)} = y_{r,T+h-i}$ for $i \geq h$, and $(\varepsilon_{c,T+h}^{*(b)}, \varepsilon_{r,T+h}^{*(b)})'$ are pairwise random draws with replacement from $\hat{F}_{\hat{\varepsilon}}$. Notice that, in order to obtain forecasts conditional on the available data set, the last p values of the original data are fixed in this step.

Step 4. Repeat steps 2 and 3 B times.

We obtain B bootstrap replicates of the vector $Y_{T+h|T}^{*(b)} = (\hat{y}_{c,T+h|T}^{*(b)}, \hat{y}_{r,T+h|T}^{*(b)})'$; see Fresoli, Ruiz, and Pascual (2015) for the asymptotic validity of the bootstrap procedure.

The bootstrap replicates obtained through the procedure proposed by Fresoli *et al.*(2015) can be used to obtain the following pointwise bootstrap ellipsoid with $100\times(1-\alpha)\%$ coverage

$$BE_{T+h} = \left[Y_{T+h} | [Y_{T+h} - \bar{Y}_{T+h|T}^{*}]' S_{Y^*}(h)^{-1} [Y_{T+h} - \bar{Y}_{T+h|T}^{*}]\right] \leq q_{1-\alpha}^{*}, \tag{2.14}$$

[4] Alternatively, we can use the permutation bootstrap initially proposed by LePage and Podgorski (1996), which is expected to have a better performance in the presence of heavy-tailed errors; see Cavaliere et al. (in press) for an application to non-causal time series.

where $\bar{Y}^*_{T+h|T}$ is the sample mean of the B bootstrap replicates $Y^{*(b)}_{T+h|T}$, $S_{Y^*}(h)$ is the corresponding sample covariance matrix and $q^*_{1-\alpha}$ is the $(1-\alpha)$ quantile of the empirical distribution of the quadratic form $[Y^{*(b)}_{T+h} - \bar{Y}^*_{T+h|T}]' S_{Y^*}(h)^{-1} [Y^{*(b)}_{T+h} - \bar{Y}^*_{T+h|T}]$.

Pointwise bootstrap prediction regions for the center/log-range system can also be constructed as Bonferroni rectangles with at least $100 \times (1-\alpha)\%$ coverage with the following corners

$$[q^*_{c,\alpha/4}, q^*_{r,\alpha/4}], \quad [q^*_{c,\alpha/4}, q^*_{r,1-\alpha/4}], \quad [q^*_{c,1-\alpha/4}, q^*_{r,\alpha/4}], \quad [q^*_{c,1-\alpha/4}, q^*_{r,1-\alpha/4}] \tag{2.15}$$

where $q^*_{c,\alpha/4}$ and $q^*_{r,\alpha/4}$ are the $\alpha/4$ quantiles from the respective marginal bootstrap distributions of the center and the log-range.

If we wish to correct for the contemporaneous correlation between the center and the log-range, we construct a pointwise bootstrap modified Bonferroni rectangle with the following corners

$$\begin{gathered} [q^*_{c,\alpha/4}, q^*_{r,\alpha/4} + p^B_{21,h} q^*_{c,\alpha/4}], \quad [q^*_{c,\alpha/4}, q^*_{r,1-\alpha/4} + p^B_{21,h} q^*_{c,\alpha/4}], \\ [q^*_{c,1-\alpha/4}, q^*_{r,\alpha/4} + p^B_{21,h} q^*_{c,1-\alpha/4}], \quad [q^*_{c,1-\alpha/4}, q^*_{r,1-\alpha/4} + p^B_{21,h} q^*_{c,1-\alpha/4}] \end{gathered} \tag{2.16}$$

where $p^B_{21,h} = S_{Y^*}(h)_{21} / S_{Y^*}(h)_{11}$.

Note that neither the bootstrap ellipsoid nor the Bonferroni rectangles need to be probability-centered when the joint distribution of the center/log-range system is not normal; see, for example Beran (1993) for the desirable properties of multivariate forecast regions. In this case, these regions will be only approximations to the true shape of the

bootstrap forecasts. Alternatively, probability-centered forecast regions can be constructed using the convex hull peeling method of Tukey (1975); see Green (1985) for a description.[5] The Tukey peeling method consists of constructing a series of convex prediction polygons. Given a data cloud, the first layer of the Tukey convex hull is the convex polygon formed by the boundary of the data. It continues by peeling the first layer off and finding the second layer for the remaining data. This process is repeated until no convex polygon can be constructed anymore. In our case, we have a two-dimensional bootstrap data cloud $Y^{*(b)}_{T+h|T} = (\hat{y}^{*(b)}_{c,T+h|T}, \hat{y}^{*(b)}_{r,T+h|T})'$. We construct layers of convex polygons and we choose the polygon that provides the closest coverage to the desired nominal coverage rate. This is the Tukey nonparametric region. The bootstrap forecast regions for the center/log-range system can obviously be also constructed even if the errors are normal. As an illustration, Figure 2.1 we plot the 95% bootstrap ellipse and the Bonferroni and modified Bonferroni rectangles when the data is generated by the same data generating process described in the previous section. These regions are based on B=4000 bootstrap replicates. Given the large sample size $T = 1000$ to estimate the parameters, the uncertainty due to parameter estimation is negligible. Consequently, the normal and bootstrap ellipses have identical shapes. The Tukey hull follows very closely the ellipses. The bootstrap Bonferroni rectangles are also very similar to their normal counterparts.

[5] One can also construct prediction regions using the High Density Regions proposed by Hyndman (1996) based on kernel estimates of the joint bootstrap empirical density or using the Monge-Kantorovich distance as proposed by Chernozhukov *et al.* (2017).

2.4.2 Prediction regions for center/range and lower/upper systems

As in the previous section, we can construct prediction regions for the center/range system based on the bootstrap $100\times(1-\alpha)\%$ ellipsoid (2.14) of the center/log-range system. By transforming the points (center, log-range) sitting on the boundary of (2.14) to points (center, exp(log-range)), we obtain the $100\times(1-\alpha)\%$ transformed bootstrap ellipse (T-BE)

$$\begin{aligned} \text{T-BE}_{T+h} = \big\{ \big[(y_{c,T+h}, \exp[\log R_{T+h}])'\big] \\ \text{such that} \quad (Y_{T+h} - \bar{Y}^*_{T+h|T})' S_{Y^*}(h)^{-1} (Y_{T+h} - \bar{Y}^*_{T+h|T}) = q^*_{1-\alpha} \big\} \end{aligned} \tag{2.17}$$

Similarly, we obtain the transformed bootstrap Bonferroni rectangle for the center/range system with corners

$$\begin{aligned} &[q^*_{c,\alpha/4}, \exp(q^*_{r,\alpha/4})], \quad [q^*_{c,\alpha/4}, \exp(q^*_{r,1-\alpha/4})], \\ &[q^*_{c,1-\alpha/4}, \exp(q^*_{r,\alpha/4})], \quad [q^*_{c,1-\alpha/4}, \exp(q^*_{r,1-\alpha/4})] \end{aligned} \tag{2.18}$$

and the transformed bootstrap modified Bonferroni rectangle with corners

$$\begin{aligned} &[q^*_{c,\alpha/4}, \exp(q^*_{r,\alpha/4} + p^B_{21,h} q^*_{c,\alpha/4})], \quad [q^*_{c,\alpha/4}, \exp(q^*_{r,1-\alpha/4} + p^B_{21,h} q^*_{c,\alpha/4})], \\ &[q^*_{c,1-\alpha/4}, \exp(q^*_{r,\alpha/4} + p^B_{21,h} q^*_{c,1-\alpha/4})], \quad [q^*_{c,1-\alpha/4}, \exp(q^*_{r,1-\alpha/4} + p^B_{21,h} q^*_{c,1-\alpha/4})] \end{aligned} \tag{2.19}$$

where $p^B_{21,h}$ is defined as in (4.3).

We also construct the Tukey nonparametric region for the data cloud of bootstrap realizations of center and range $(\hat{y}^{*(b)}_{c,T+h|T}, \exp(\hat{y}^{*(b)}_{r,T+h|T}))'$. In Figure 2.2, we plot these regions for the same simulated system considered above.

Finally, for the lower/upper bounds system, we calculate first the bootstrap upper and lower bounds based on the bootstrap realizations of the center and range as follows

$$y^{*(b)}_{u,T+h} = \hat{y}^{*(b)}_{c,T+h|T} + \frac{1}{2}\exp(\hat{y}^{*(b)}_{r,T+h|T}) \tag{2.20}$$

$$y^{*(b)}_{l,T+h} = \hat{y}^{*(b)}_{c,T+h|T} - \frac{1}{2}\exp(\hat{y}^{*(b)}_{r,T+h|T}) \tag{2.21}$$

and construct a bootstrap ellipsoid for the upper and lower bounds as

$$BE^{UL}_{T+h} = \left[Y^{UL}_{T+h} | \ [Y^{UL}_{T+h} - \bar{Y}^{UL*}_{T+h|T}]' S^{UL}_{Y^*}(h)^{-1} [Y^{UL}_{T+h} - \bar{Y}^{UL*}_{T+h|T}] \leq q^{UL*}_{1-\alpha}\right] \tag{2.22}$$

where $Y^{UL}_{T+h} = (y_{u,T+h}, y_{l,T+h})'$ and $\bar{Y}^{UL*}_{T+h|T}$ and $S^{UL}_{Y^*}(h)$ represents the mean vector and variance covariance matrix, respectively, of the bootstrap upper/lower bound realizations.

Finally, a Tukey nonparametric region can be constructed for the data cloud of bootstrap realizations of upper and lower bounds $(y^{*(b)}_{u,T+h}, y^{*(b)}_{l,T+h}))'$. Note that for this system, we do not construct Bonferroni rectangles because they may contain unfeasible subregions of points where the lower bound is greater than the upper bound.

In Figures 2.1-2.4, we illustrate the shapes of the different prediction regions. We run a single simulation to construct the one-step-ahead 95%-probability forecast regions based on the estimation of a VAR(4) model ($T = 1000$) for the center/log-range system whose errors follow a bivariate normal distribution with contemporaneous correlation of

-0.24. For the bootstrap procedures, we use $B = 4000$. In Figure 2.1, we plot seven regions for the center/log-range system. As expected, the Normal ellipse and the bootstrap ellipse have identical shapes. The Tukey convex hull follows very closely the ellipses. The modified Bonferroni rectangles are able to capture the negative correlation between center and log-range. In Figure 2.2, we plot the seven regions for the center/range system. As expected, the transformed Normal ellipse and the transformed bootstrap ellipse have identical shapes. The Tukey convex hull and the analytical contour based on (2.10) follow very closely the transformed ellipses. In Figures 2.3 and 2.4, we plot the 95% bootstrap forecast ellipsoid and the Tukey region for the upper/lower bounds system. The analytical contour based on (2.10) and the Tukey convex hull are very close to each other. However, for this particular realization, the bootstrap ellipsoid is somehow different mainly in the center and upper right corner of the distribution of the lower/upper bounds system.

2.5 Evaluation of the Prediction Regions

We present several criteria to evaluate the prediction regions. As in the case of loss functions, it is only the objective of the forecaster that will define which criterium is the most appropriate.

At the most basic level, the forecaster will aim for reliability, that is, those prediction regions that provide the closest coverage to the nominal coverage rate. In an out-of-sample environment, for a regions with $100 \times (1 - \alpha)\%$ nominal coverage, the **average**

coverage rate is defined as

$$C_{(1-\alpha)} = \frac{1}{N}\sum_{t=1}^{N} I_t^{(1-\alpha)} \tag{2.23}$$

where N is the number of out-of-sample forecasts and $I_t^{(1-\alpha)}$ is an indicator variable that is equal to 1 if the observed outcome falls within the prediction region and 0 otherwise.

Following Golestaneh *et al.* (2017), we combine reliability with sharpness, a preference for regions with smaller area or volume, and they propose the following **average coverage-volume score** for regions with $100 \times (1-\alpha)\%$ nominal coverage

$$CV_{(1-\alpha)} = \Big|\frac{1}{N}\sum_{t=1}^{N}\big[I_t^{(1-\alpha)} - (1-\alpha)\big] \times \big[V_t^{(1-\alpha)}\big]^{\frac{1}{p}}\Big| \tag{2.24}$$

where $V_t^{(1-\alpha)}$ is the volume of the prediction region with nominal coverage rate $(1-\alpha)$ at time t, and p is the dimension of the outcome variable, which in our case is $p = 2$. The forecaster would prefer a lower score as he is aiming for regions with high reliability and small area.

Another aspect to the evaluation of forecast regions is to consider the observations outside of the $100 \times (1-\alpha)\%$ region and to assess how far they are from a central point within the prediction region. We propose the following **average outlier distance**

$$O_{(1-\alpha)} = \frac{1}{G}\sum_{t=1}^{N}\big[1 - I_t^{(1-\alpha)}\big] \times D(y_t, M_t) \tag{2.25}$$

where G is the number of observations outside the region, D is a distance measure (e.g. Euclidean distance) of each outside-the-region outcome y_t from M_t, which is a central point

in the region. We choose M_t to be the median of the realizations generated at each time t according to the methods explained in Section 4. However, defining the median for a multi-dimensional dataset (2-dimensional in our case) is not as straighforward as it is for a one-dimensional dataset. To obtain M_t, we implement the definition of median in a multi-dimensional setting introduced by Zuo (2003), known as 'projection depth median', and programmed in the Matlab package (Liu and Zuo, 2015). A brief description follows.

With a one-dimensional dataset, $Z = \{Z_i\}$, $i = 1, ...n$, a robust measurement of the outlyingness of a point z (a scalar) relative to Z is the outlying function

$$o_1(z, Z) \quad = \quad \frac{|z - Med(Z)|}{MAD(Z)}$$

where Med is the median of data set Z and $MAD(Z) = Med\{|Z_i - Med(Z)|, i = 1, ..., n\}$. When z and Z are p-dimensional ($p > 1$), the above outlying function is applied by projecting z and Z into a one-dimensional space, i.e., $o_1(u^T z, u^T Z)$, where $u \in S$ and $S = \{v \in R^p : \|v\| = 1\}$ is a set of unit vectors in the p-dimensional space. The projection depth of point z with respect to Z is defined as $PD(z, Z) = (1 + O(z, Z))^{-1}$, where $O(z, Z) \equiv \sup_{u \in S} o_1(u^T z, u^T Z)$. Under some mild conditions (Zuo, 2013), there exists a unique single point ($z^* \in R^p$, not necessarily from Z) that maximizes $PD(z, Z)$ for a data set Z. This z^* is defined as the 'projection depth median' of Z.

The $100 \times \alpha\%$ outside-the-region observations can be considered 'risk' that the forecaster has to bear and, in this sense, he would like to minimize $O_{(1-\alpha)}$. For two regions with similar coverage, the forecaster will choose that with a lower average outlier dispersion.

We also evaluate the prediction region by the sharpness or tightness of the intervals that result from projecting the two-dimensional region into one-dimensional intervals. We draw a large number of directions, which are given by the lines drawn from the zero origin of the unit circle to any point in its boundary. For each direction, we find the two bounding tangent lines to the prediction region that are perpendicular to that direction. We calculate the length of the projected interval bounded by the tangent lines. See Figure 2.5 (top panel) for a graphical representation. Denote $d_i \in \Upsilon$ as the i^{th} direction in Υ, where Υ is the set of all directions, and let D be the number of directions. At time t, the average projection length over all directions is

$$P_t = \frac{1}{D}\sum_{i=1}^{D}(u_{d_i} - l_{d_i}),$$

where u_{d_i} is the upper bound and l_{d_i} the lower bound of the projected interval in the i^{th} direction. Then, over the prediction sample, the **average length of the projected intervals** associated with the $(1-\alpha)\%$ prediction region is

$$P_{(1-\alpha)} \quad = \quad \frac{1}{N}\sum_{t=1}^{N} P_t \tag{2.26}$$

The forecaster would prefer prediction regions that deliver tight projected intervals.

We now consider the realized data points over the prediction period in conjunction with the projected intervals. For each direction, we also project each point into that direction and measure whether the point falls into or outside of the projected interval (see Figure 2.5, top panel) . An indicator function I will assign the value 0 if the point falls inside and

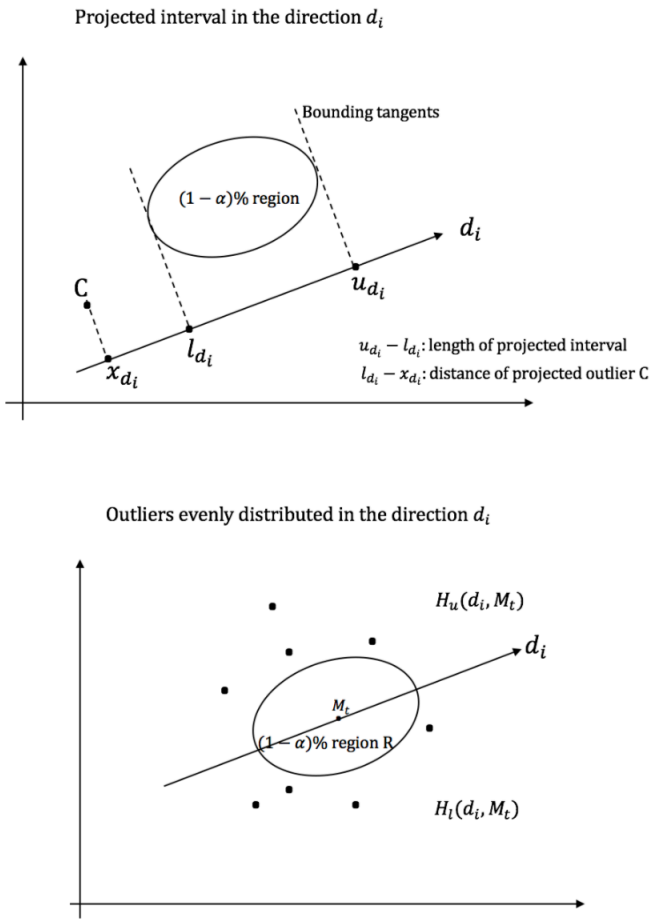


Figure 2.5: Projected interval and projected outliers (top panel). Outlier distribution around a region (bottom panel)

1 if the point falls outside of the projected interval. At time t, over all D directions, we calculate the average distance of the projected outliers to the projected interval as

$$OP_t = \frac{1}{D}\sum_{i=1}^{D}[(l_{d_i} - x_{d_i})I(x_{d_i} < l_{d_i}) + (x_{d_i} - u_{d_i})I(x_{d_i} > u_{d_i})]$$

where x_{d_i} is the coordinate of the data point projected on the i^{th} direction. Then, over the prediction sample, the **average distance of the projected outliers** associated with the $100 \times (1-\alpha)\%$ prediction region is

$$OP_{(1-\alpha)} = \frac{1}{N}\sum_{t=1}^{N} OP_t \tag{2.27}$$

The forecaster prefers prediction regions with projected outliers close to the projected intervals.

We expect that when the length of the projected interval is large, the distance of the projected outliers to the interval will be smaller. To take into account this a trade-off, we propose a combined criterium $POP_t = P_t \times OP_t$ so that, over the prediction sample, the average trade-off associated with the $100 \times (1-\alpha)\%$ prediction region is

$$POP_{(1-\alpha)} = \frac{1}{N}\sum_{t=1}^{N} POP_t \tag{2.28}$$

A smaller $POP_{(1-\alpha)}$ would be preferred by the forecaster.

Finally, we assess whether the prediction region is probability-centered. We check whether the points outside of the prediction region are evenly distributed around the region. At time t, we consider a cloud of data points and calculate the median M_t as in Zuo (2003).

We also consider a number of directions D that pass through M_t. We define $H_u(d_i, M_t)$ as the half-plane above the line generated by the direction d_i and $H_l(d_i, M_t)$ as the half-plane below the same line. See Figure 2.5 (bottom panel) for a graphical representation. For a given direction d_i and the $100 \times (1-\alpha)\%$ prediction region R, we consider the number of points outside of the region, i.e., $x \in R^c$, which are either in $H_u(d_i, M_t)$ or in $H_l(d_i, M_t)$, that is

$$C_u(d_i, M_t)) = \{\#x | (x \in R^c) \cap (x \in H_u(d_i, M_t))\}$$

$$C_l(d_i, M_t) = \{\#x | (x \in R^c) \cap (x \in H_l(d_i, M_t))\}$$

where $C_u(d_i)$ and $C_l(d_i)$ are functions providing the number of of outlier points falling in the upper half-plane or lower half-plane respectively. If the outliers are evenly distributed around the $100 \times (1-\alpha)\%$ prediction region, we expect the following statistic $S_{(1-\alpha)}(M_t)$ to be close to zero

$$S_{(1-\alpha)}(M_t) = \frac{1}{D} \sum_{i=1}^{D} |C_u(d_i, M_t)) - C_l(d_i, M_t)| \tag{2.29}$$

Though $S_{(1-\alpha)}(M_t)$ will not be feasible with real data (we will have only one realized observation at time t that could be in or out of the prediction region), in a simulated environment, we will be able to assess the probability-centered property of each prediction region.

2.6 Monte Carlo Simulations

We perform extensive Monte Carlo simulations to assess the performance of the prediction regions constructed with the analytical and semiparametric methods explained in sections 3 and 4. The regions are evaluated according to the seven criteria described in section 5.

We generate a small sample of $T = 200$ observations and a large sample of $T = 1000$ observations (estimation samples) from a VAR(4) for the center/log-range system (1.1)-(1.2) with parameter values reported in Table 2.1. We consider four cases regarding distributional assumptions from which the errors are drawn: (1) center and log-range errors are both normally distributed; (2) center errors are Student-t with 5 degrees of freedom and log-range errors are normally distributed; (3) center errors are Student-t with 5 degrees of freedom and range errors are exponentially (λ) distributed; and (4) center errors are normal and range errors exponentially (λ) distributed. Note that the distributional assumptions are on the marginal densities of the errors of each equation. It is only in case (1) that the bivariate density of the center and log-range is normal; in the rest of the cases, we do not know the exact bivariate densities.[6]

[6]For the system to have the desired marginal density functions and the stated correlation structure, we have generated bivariate errors from a Gaussian copula and re-transform the PITs of the corresponding univariate normal variates according to the desired density, e.g. Student-t, to obtain the new error variates, which need to be adjusted to have the desired mean and variance.

	Center equation	log-Range equation
Constant	-0.9344	0.0759
C(-1)	0.3404	-0.0112
C(-2)	-0.1530	-0.0027
C(-3)	0.0314	-0.0030
C(-4)	-0.0551	-0.0022
log-R(-1)	-0.5030	0.0852
log-R(-2)	0.1281	0.1845
log-R(-3)	-0.1556	0.1539
log-R(-4)	0.9157	0.0760

Variance-covariance matrix of the errors:

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 111.24 & -1.02 \\ -1.02 & 0.16 \end{bmatrix}$$

Contemporaneous correlation between the center and log-range errors = -0.24.

Table 2.1: Monte Carlo simulations. VAR(4) parameter values for the center/log-range system

We consider 1- and 3-step-ahead prediction regions with 95% nominal coverage.[7] We calculate the empirical coverage by simulating 1000 future values of the required vector at time T, i.e. center/log-range, center/range, and upper/lower bounds, at the forecast horizon, and calculating the proportion of these values that falls within the constructed prediction regions. The number of Monte Carlo replications is 500, the number of bootstrap samples is $B = 2000$, and the number of directions to calculate the average length of the projected intervals and outliers is $D = 100$.

2.6.1 Center and Log-Range are Normally distributed

The errors of the center equation of the VAR(4) are drawn from a normal density as well as the errors of the log-range equation. In Tables 2.2-2.3, we report the evaluation of the prediction regions for the three systems (center/log-range, center/range and upper/lower

[7]The results for $h = 3$ are provided in the Supplementary Material, Tables S1-S4.

bounds) for forecast horizon $h = 1$ with estimation samples $T = 1000$ and $T = 200$. Note that we only estimate the VAR(4) once for the center/log-range, construct prediction regions for this system, and based on these estimates, we proceed to construct prediction regions for the other two systems.

Given the bivariate normality of the center/log-range system, the prediction regions based on the normal ellipse (2.6) and on the analytical methods (2.10) would be exact if parameter estimation were not a concern. For a large estimation sample $T = 1000$ (Table 2.2), all regions, except for the Tukey convex hull, are very reliable with empirical coverage C_{95} of mostly 95%. Bootstrap ellipse and bootstrap Bonferroni rectangles, which account for parameter uncertainty, deliver the closest value to 95% in the three systems. Bonferroni rectangles have the largest areas compared to the normal ellipse and to the regions based on analytical methods, but because they provide good coverage, they enjoy one of the lowest average coverage-volume scores CV_{95}. The larger area of the Bonferroni rectangles is somehow compensated by a lower average outlier distance O_{95}, though this metric is very similar for all prediction regions considered across the three systems. The tightest regions i.e., those projecting tight one-dimensional intervals measured by POP_{95}, correspond to the normal ellipse, bootstrap ellipsoids, and those regions based on exact analytical expressions. These are also the regions with outliers more evenly distributed around their boundaries.

For small estimation sample $T = 200$ (Table 2.3), the bootstrap regions provide a clear advantage with respect to the other regions. In small samples, parameter uncertainty plays a more important role than in large samples. Boostrap methods are designed to take into account estimation uncertainty. Across systems, bootstrap ellipsoids and bootstrap

Bonferroni rectangles are still very reliable with empirical coverage close to 95%. They also enjoy the smallest score CV_{95}. The tightest regions, i.e small POP_{95}, are provided by the bootstrap ellipsoid and its transformed regions followed by the Tukey region.

Considering the overall performance assessed by the metrics C, CV, and POP, for large estimation samples, normal ellipses, bootstrap ellipsoids, and those regions based on analytical methods are the best performers, and as expected, better than the Tukey convex hull. Bonferroni rectangles, though providing good coverage, tend to be conservative in area, which in turn provides some advantages regarding the lower dispersion of the outliers. For small samples, the bootstrap ellipsoid is the best performer. These conclusions hold regardless of whether $h = 1$ or $h = 3$.

Large Estimation Sample $T = 1000$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (2.6)	0.9469	8.7059	0.0645	18.8891	25.8596	0.0275	0.7090	0.0089
Bonferroni rectangle (2.7)-(2.8)	0.9484	9.1321	0.0646	18.3628	24.9220	0.0375	0.9309	0.0088
Modified Bonferroni rectangle (2.9)	0.9516	9.1321	0.0655	18.5670	24.9306	0.0372	0.9255	0.0095
BE: Bootstrap ellipsoid (2.14)	0.9493	8.7842	0.0801	19.0161	26.0994	0.0262	0.6773	0.0087
Bootstrap Bonferroni rectangle (2.15)	0.9493	9.2184	0.0819	18.4758	25.1222	0.0373	0.9293	0.0096
Modified Bootstrap Bonferroni rectangle (2.16)	0.9521	9.2184	0.0848	18.6347	25.1309	0.0371	0.9240	0.0104
Tukey convex hull	0.9414	8.6389	0.1028	18.5770	26.1263	0.0283	0.7265	0.0107

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9470	9.8933	0.0782	18.7056	26.2543	0.0263	0.6885	0.0161
T-NE: T-Normal ellipse (2.3.2)	0.9469	10.1606	0.0756	18.9612	25.9781	0.0280	0.7259	0.0089
T-Bonferroni rectangle (2.12)	0.9484	10.7382	0.0758	18.4471	25.4849	0.0349	0.8849	0.0088
T-Modified Bonferroni rectangle (2.13)	0.9513	10.9309	0.0788	18.5675	25.5757	0.0343	0.8742	0.0101
T-BE: T-Bootstrap ellipsoid (2.4.2)	0.9493	10.2637	0.0940	19.0892	28.8440	0.0267	0.7346	0.0087
T-Bootstrap Bonferroni rectangle (2.4.2)	0.9493	10.8482	0.0966	18.5638	25.6943	0.0347	0.8832	0.0096
T-Modified Bootstrap Bonferroni rectangle (2.19)	0.9519	11.0436	0.1036	18.6581	25.7867	0.0342	0.8722	0.0109
Tukey convex hull	0.9411	10.2071	0.1237	18.6050	26.2517	0.0287	0.7406	0.0119

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9470	9.8763	0.0781	26.3045	46.5574	0.0463	2.1514	0.0095
Bootstrap ellipsoid (2.22)	0.9488	10.4679	0.1010	24.3205	47.3650	0.0414	1.9393	0.0126
Tukey convex hull	0.9411	10.1884	0.1232	26.1956	46.5761	0.0506	2.3167	0.0114

Table 2.2: Evaluation of the h-step ahead 95% prediction regions from a **GAUSSIAN center/log-range system** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

Small Estimation Sample $T = 200$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (2.6)	0.9323	8.5326	0.1687	18.4343	25.4321	0.0369	0.9220	0.0208
Bonferroni rectangle (2.7)-(2.8)	0.9352	8.9625	0.1559	18.0943	24.5124	0.0490	1.1831	0.0198
Modified Bonferroni rectangle (2.9)	0.9378	8.9625	0.1421	18.1143	24.5209	0.0488	1.1768	0.0201
BE: Bootstrap ellipsoid (2.14)	0.9465	8.9272	0.1302	19.0901	26.6052	0.0276	0.7146	0.0172
Bootstrap Bonferroni rectangle (2.15)	0.9455	9.3834	0.1384	18.4711	25.6546	0.0408	1.0206	0.0181
Modified Bootstrap Bonferroni rectangle (2.16)	0.9480	9.3834	0.1320	18.4979	25.6636	0.0406	1.0152	0.0178
Tukey convex hull	0.9334	8.7457	0.1775	18.3070	26.8745	0.0310	0.8045	0.0222

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9327	9.7378	0.1927	18.4312	25.8139	0.0356	0.9029	0.0247
T-NE: T-Normal ellipse (2.3.2)	0.9323	9.9933	0.1971	18.5044	25.5514	0.0376	0.9434	0.0209
T-Bonferroni rectangle (2.12)	0.9352	10.5726	0.1836	18.1739	25.0743	0.0457	1.1273	0.0199
T-Modified Bonferroni rectangle (2.13)	0.9373	10.7899	0.1750	18.1043	25.1757	0.0450	1.1140	0.0208
T-BE: T-Bootstrap ellipsoid (2.4.2)	0.9465	10.4975	0.1528	19.1640	26.7346	0.0281	0.7324	0.0173
T-Bootstrap Bonferroni rectangle (2.4.2)	0.9455	11.1567	0.1645	18.5525	26.2772	0.0378	0.9673	0.0181
T-Modified Bootstrap Bonferroni rectangle (2.19)	0.9472	11.3987	0.1633	18.4061	26.3916	0.0371	0.9548	0.0187
Tukey convex hull	0.9338	10.4532	0.2114	18.3920	27.0735	0.0311	0.8151	0.0224

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9327	9.7528	0.1931	25.9290	45.7581	0.0629	2.8261	0.0198
Bootstrap ellipsoid (2.22)	0.9470	10.7473	0.1919	25.1246	48.4270	0.0429	2.0190	0.0180
Tukey convex hull	0.9338	10.4590	0.2118	25.8989	48.0060	0.0550	2.5477	0.0220

Table 2.3: Evaluation of the h-step ahead 95% prediction regions from a **GAUSSIAN center/log-range system** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

2.6.2 Student-t(5) Center and Normal Log-Range

In Tables 2.4-2.5, we report the performance of the different predictions regions when the errors of the center equation of the VAR(4) are leptokurtic and the errors of the log-range equation are normal. Thus, the bivariate system center/log-range is not normally distributed but symmetric. Consequently, the normal ellipse, Bonferroni rectangles, and their corresponding transformed regions tend to undercover with empirical coverage rates of about 94% in large samples, and about 93% in small samples because they do not consider the fat tails of the errors in the center equation. The bootstrap regions, which are robust to distributional assumptions and capture estimation uncertainty, are better performers with coverage rates close to 95% in large and small samples. They also provide the smallest score CV_{95} and, according to S_{95}, tend to have a more evenly distribution of outliers around the regions. For small samples, the performance of bootstrap regions is even more striking with the bootstrap ellipsoid being the best region in terms of C_{95}, CV_{95}, and S_{95}.These results hold for both horizons $h = 1$ or $h = 3$.

The Tukey regions, which do not require any distributional assumption, are in-between the bootstrap regions and the regions based on normality. Note that the Tukey regions have a superior advantage according to POP_{95}. Sacrificing a bit of coverage, the Tukey region provides the tightest one-dimensional projections across systems, estimation samples, and forecast horizons.

Large Estimation Sample $T = 1000$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (2.6)	0.9412	8.6797	0.0890	22.6435	28.0054	0.1126	3.1971	0.0088
Bonferroni rectangle (2.7)-(2.8)	0.9398	9.1163	0.1026	21.9114	24.8483	0.1269	3.1407	0.0090
Modified Bonferroni rectangle (2.9)	0.9429	9.1163	0.0850	22.3330	24.8572	0.1266	3.1360	0.0100
BE: Bootstrap ellipsoid (2.14)	0.9488	8.9704	0.0760	23.7767	26.6225	0.1029	2.7149	0.0079
Bootstrap Bonferroni rectangle (2.15)	0.9492	9.6376	0.0850	21.5902	27.3796	0.0984	2.6631	0.0092
Modified Bootstrap Bonferroni rectangle (2.16)	0.9523	9.6376	0.0854	22.0367	27.3895	0.0982	2.6586	0.0100
Tukey convex hull	0.9413	8.8669	0.1045	21.0923	29.6214	0.0785	2.2802	0.0106

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9411	9.9212	0.1058	22.5047	26.1741	0.1090	2.8408	0.0143
T-NE: T-Normal ellipse (2.3.2)	0.9412	10.1810	0.1048	22.7112	28.3778	0.1129	3.1091	0.0087
T-Bonferroni rectangle (2.12)	0.9398	10.7807	0.1219	21.9890	25.4332	0.1215	3.0773	0.0090
T-Modified Bonferroni rectangle (2.13)	0.9426	10.9858	0.1053	22.3062	25.5306	0.1206	3.0676	0.0109
T-BE: T-Bootstrap ellipsoid (2.4.2)	0.9488	10.5704	0.0897	23.8431	34.9965	0.1027	3.1692	0.0079
T-Bootstrap Bonferroni rectangle (2.4.2)	0.9492	11.4053	0.1009	21.6829	27.9724	0.0944	2.6087	0.0091
T-Modified Bootstrap Bonferroni rectangle (2.19)	0.9518	11.6538	0.1043	21.9587	28.0838	0.0937	2.5993	0.0109
Tukey convex hull	0.9416	10.5931	0.1239	21.1648	29.8935	0.0776	2.2727	0.0123

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (2.10)	0.9411	9.8875	0.1053	31.6801	46.4097	0.1959	9.0588	0.0090
Bootstrap ellipsoid (2.22)	0.9486	10.9656	0.1034	30.8649	49.3236	0.1642	8.0201	0.0112
Tukey convex hull	0.9416	10.5863	0.1236	29.8004	53.1740	0.1386	7.2162	0.0106

Table 2.4: Evaluation of the h-step ahead 95% prediction regions from a system with **center STUDENT-t(5) distributed and NORMAL log-range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

Small Estimation Sample $T = 200$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (3.1)	0.9280	8.5034	0.1946	21.5578	25.3124	0.1243	3.0850	0.0198
Bonferroni rectangle (3.2)-(3.3)	0.9281	8.9418	0.2017	20.9695	24.4282	0.1400	3.3598	0.0200
Modified Bonferroni rectangle (3.4)	0.9309	8.9418	0.1841	21.1982	24.4369	0.1398	3.3548	0.0200
BE: Bootstrap ellipsoid (4.1)	0.9465	9.1088	0.1302	23.8342	27.0845	0.1018	2.6898	0.0151
Bootstrap Bonferroni rectangle (4.2)	0.9444	9.7872	0.1448	21.5150	27.8271	0.1044	2.8057	0.0176
Modified Bootstrap Bonferroni rectangle (4.3)	0.9471	9.7872	0.1376	21.7377	27.8371	0.1042	2.8013	0.0174
Tukey convex hull	0.9339	8.9892	0.1761	20.7920	30.3991	0.0836	2.4223	0.0213

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9280	9.6810	0.2227	21.5371	25.7212	0.1207	3.0448	0.0233
T-NE: T-Normal ellipse (3.6)	0.9280	9.9215	0.2269	21.6251	30.6254	0.1244	3.4822	0.0199
T-Bonferroni rectangle (3.7)	0.9281	10.5185	0.2370	21.0455	24.9815	0.1344	3.2963	0.0199
T-Modified Bonferroni rectangle (3.8)	0.9303	10.7436	0.2278	21.1401	25.0872	0.1333	3.2845	0.0209
T-BE: T-Bootstrap ellipsoid (4.4)	0.9465	10.6969	0.1525	23.9020	36.2004	0.1017	3.1213	0.0150
T-Bootstrap Bonferroni rectangle (4.5)	0.9444	11.6073	0.1712	21.5993	28.4407	0.0999	2.7454	0.0174
T-Modified Bootstrap Bonferroni rectangle (4.6)	0.9461	11.9048	0.1724	21.5778	28.5761	0.0991	2.7347	0.0181
Tukey convex hull	0.9343	10.7344	0.2128	20.8409	30.6042	0.0839	2.4434	0.0218

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9280	9.6699	0.2229	30.3216	45.5948	0.2173	9.7108	0.0178
Bootstrap ellipsoid (4.9)	0.9464	11.1371	0.1877	31.5378	50.2535	0.1626	7.9359	0.0153
Tukey convex hull	0.9343	10.7409	0.2131	29.3517	54.4157	0.1498	7.7539	0.0196

Table 2.5: Evaluation of the h-step ahead 95% prediction regions from a system with **center STUDENT-t(5) distributed and NORMAL log-range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

2.6.3 Student-t(5) Center and Exponential Range

In Tables 2.6-2.7, we report the performance of the different predicitons regions when the errors of the center equation of the VAR(4) are drawn from a Student-t with 5 degrees of freedom and the errors of the log-range equation are those resulting from assuming that the *range* itself is exponentially distributed. The exponential errors introduce some asymmetry that is not fully corrected when they are transformed into errors of the log-range equation. The resulting bivariate system center/log-range is not normally distributed as it exhibits leptokurtosis and asymmetry.

For small and large samples, the bootstrap regions (ellipsoids and Bonferroni rectangles) provide the best coverage C_{95} with empirical rates very close to 95%, followed by the Tukey region that covers around 94% of the events. The same regions have the smallest scores CV_{95} and the smallest POP_{95}. As in the previous case, the Tukey region has a clear advantage over the other regions when we are interested in the smallest POP_{95}. It is interesting to note that the boostrap Bonferroni rectangles are able to distribute outliers more evenly around their perimeters than any other prediction regions. These results hold for the two horizons considered $h = 1$ and $h = 3$.

Large Estimation Sample $T = 1000$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (3.1)	0.9368	8.7108	0.1201	21.1329	27.9739	0.1133	3.2150	0.0280
Bonferroni rectangle (3.2)-(3.3)	0.9341	9.1404	0.1482	20.7347	24.8541	0.1273	3.1531	0.0298
Modified Bonferroni rectangle (3.4)	0.9357	9.1404	0.1375	20.8467	24.8628	0.1271	3.1475	0.0280
BE: Bootstrap ellipsoid (4.1)	0.9492	9.3037	0.0779	22.1716	32.8760	0.0937	2.9613	0.0233
Bootstrap Bonferroni rectangle (4.2)	0.9502	9.7724	0.0905	22.0051	27.3986	0.0988	2.6772	0.0099
Modified Bootstrap Bonferroni rectangle (4.3)	0.9524	9.7724	0.0920	22.4823	27.4082	0.0986	2.6727	0.0108
Tukey convex hull	0.9428	9.0187	0.1055	21.4827	29.6579	0.0790	2.3001	0.0122

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9433	9.9723	0.0914	22.5306	26.1769	0.1083	2.8242	0.0186
T-NE: T-Normal ellipse (3.6)	0.9368	10.2383	0.1419	21.0668	30.7229	0.1121	3.4644	0.0289
T-Bonferroni rectangle (3.7)	0.9341	10.8315	0.1767	20.6600	25.4505	0.1212	3.0713	0.0309
T-Modified Bonferroni rectangle (3.8)	0.9337	11.0282	0.1877	20.3441	25.5449	0.1205	3.0672	0.0312
T-BE: T-Bootstrap ellipsoid (4.4)	0.9492	11.0252	0.0923	22.0965	32.9957	0.0930	2.8497	0.0240
T-Bootstrap Bonferroni rectangle (4.5)	0.9502	10.1569	0.0941	21.9978	27.5380	0.0976	2.6566	0.0100
T-Modified Bootstrap Bonferroni rectangle (4.6)	0.9545	10.3666	0.0996	23.1236	27.6214	0.0970	2.6486	0.0109
Tukey convex hull	0.9425	9.4339	0.1060	21.2823	29.9939	0.0763	2.2415	0.0117

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9433	9.9574	0.0911	31.8056	46.4025	0.1954	9.0708	0.0105
Bootstrap ellipsoid (4.9)	0.9498	9.4454	0.0911	35.2534	46.7651	0.1925	8.9307	0.0074
Tukey convex hull	0.9425	9.4304	0.1056	30.0265	53.4562	0.1364	7.1413	0.0099

Table 2.6: Evaluation of the h-step ahead 95% prediction regions from a system with **center STUDENT-t(5) distributed and EXPONENTIAL range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

Small Estimation Sample $T = 200$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (3.1)	0.9261	8.5167	0.2056	20.6548	25.3141	0.1247	3.0939	0.0310
Bonferroni rectangle (3.2)-(3.3)	0.9251	8.9470	0.2239	20.4164	24.4319	0.1404	3.3677	0.0322
Modified Bonferroni rectangle (3.4)	0.9256	8.9470	0.2220	20.3442	24.4403	0.1401	3.3622	0.0321
BE: Bootstrap ellipsoid (4.1)	0.9465	9.3814	0.1325	22.2371	27.8365	0.0945	2.5540	0.0242
Bootstrap Bonferroni rectangle (4.2)	0.9462	9.9340	0.1571	22.2250	27.8960	0.1042	2.8043	0.0183
Modified Bootstrap Bonferroni rectangle (4.3)	0.9481	9.9340	0.1491	22.4270	27.9057	0.1040	2.7995	0.0178
Tukey convex hull	0.9354	9.0890	0.1825	21.4695	30.1396	0.0856	2.4598	0.0222

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9316	9.6907	0.1955	21.7117	25.7230	0.1197	3.0188	0.0232
T-NE: T-Normal ellipse (3.6)	0.9261	9.9385	0.2419	20.5949	28.0041	0.1237	3.2429	0.0318
T-Bonferroni rectangle (3.7)	0.9251	10.5207	0.2657	20.3489	24.9834	0.1339	3.2846	0.0332
T-Modified Bonferroni rectangle (3.8)	0.9236	10.7321	0.2891	19.9189	25.0827	0.1332	3.2787	0.0352
T-BE: T-Bootstrap ellipsoid (4.4)	0.9465	11.0513	0.1571	22.1626	34.6414	0.0937	2.7120	0.0247
T-Bootstrap Bonferroni rectangle (4.5)	0.9462	10.4326	0.1637	22.2054	28.0685	0.1026	2.7783	0.0184
T-Modified Bootstrap Bonferroni rectangle (4.6)	0.9493	10.6796	0.1551	22.6885	28.1671	0.1019	2.7678	0.0174
Tukey convex hull	0.9356	9.6360	0.1871	21.1415	30.5515	0.0826	2.4063	0.0227

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9316	9.6745	0.1953	30.6485	45.6115	0.2165	9.6754	0.0181
Bootstrap ellipsoid (4.9)	0.9488	9.7179	0.1664	36.3154	47.9860	0.1862	8.7110	0.0134
Tukey convex hull	0.9356	9.6336	0.1873	29.8330	54.4258	0.1477	7.6642	0.0194

Table 2.7: Evaluation of the h-step ahead 95% prediction regions from a system with **center STUDENT-t(5) distributed and EXPONENTIAL range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) in the first column are the corresponding equations in the text.

2.6.4 Normal Center and Exponential Range

In Tables 2.8-2.9, we report the performance of the different predictions regions when the errors of the center equation of the VAR(4) are drawn from a normal distribution and the errors of the log-range equation are those resulting from assuming that the *range* itself is exponentially distributed. The resulting bivariate system center/log-range is not normally distributed as asymmetry is introduced through the log-range equation.

For large samples, all regions have an empirical coverage C_{95} between 94 and 95% with the bootstrap ellipoid and the bootstrap Bonferroni rectangle being very close to 95%. It is interesting to note that the normal ellipse in the center/log-range system and its analytically derived regions for the center/range and upper/lower systems provide a very competitive coverage of almost 95% and the smallest scores CV_{95}. The bootstrap ellipsoid and its transformed regions come as the next best performer with some advantage regarding the POP_{95} criterium. In small samples, the boostrap methods provide the best coverage with an empirical rate of almost 95%. The bootstrap ellipsoid delivers the best performance when considering CV_{95} and POP_{95}. The normal ellipse and its analytically derived formulas tend to undercover with rates around 93%. We obtain similar results for $h = 1$ and $h = 3$.

Large Estimation Sample $T = 1000$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (3.1)	0.9435	8.7270	0.0781	17.5784	25.8654	0.0278	0.7161	0.0299
Bonferroni rectangle (3.2)-(3.3)	0.9426	9.1588	0.0872	17.4485	24.9354	0.0376	0.9352	0.0304
Modified Bonferroni rectangle (3.4)	0.9445	9.1588	0.0830	17.4346	24.9442	0.0373	0.9282	0.0283
BE: Bootstrap ellipsoid (4.1)	0.9489	8.9423	0.0747	17.5336	26.5065	0.0237	0.6226	0.0281
Bootstrap Bonferroni rectangle (4.2)	0.9506	9.3471	0.0837	18.7679	25.1703	0.0372	0.9274	0.0099
Modified Bootstrap Bonferroni rectangle (4.3)	0.9521	9.3471	0.0877	18.9354	25.1792	0.0369	0.9216	0.0114
Tukey convex hull	0.9427	8.7490	0.1003	18.7555	26.1170	0.0284	0.7289	0.0120

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9501	9.9939	0.0690	18.6000	26.2658	0.0252	0.6607	0.0186
T-NE: T-Normal ellipse (3.6)	0.9435	10.2792	0.0925	17.4961	28.6147	0.0271	0.7610	0.0304
T-Bonferroni rectangle (3.7)	0.9426	10.8600	0.1040	17.3532	25.5328	0.0341	0.8679	0.0311
T-Modified Bonferroni rectangle (3.8)	0.9426	11.0572	0.1134	17.0028	25.6268	0.0338	0.8619	0.0311
T-BE: T-Bootstrap ellipsoid (4.4)	0.9489	10.5592	0.0886	17.4463	29.2689	0.0231	0.6586	0.0285
T-Bootstrap Bonferroni rectangle (4.5)	0.9506	9.7196	0.0871	18.7588	25.3102	0.0362	0.9074	0.0100
T-Modified Bootstrap Bonferroni rectangle (4.6)	0.9541	9.8967	0.0932	19.3375	25.3842	0.0357	0.8981	0.0111
Tukey convex hull	0.9422	9.1509	0.1053	18.6275	26.3466	0.0270	0.6995	0.0118

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9501	9.9568	0.0687	26.2449	46.5699	0.0459	2.1299	0.0107
Bootstrap ellipsoid (4.9)	0.9498	9.2576	0.0893	28.2307	45.8963	0.0513	2.3358	0.0090
Tukey convex hull	0.9422	9.1278	0.1048	26.2808	46.8771	0.0480	2.2145	0.0107

Table 2.8: Evaluation of the h-step ahead 95% prediction regions from a system with **center NORMALLY distributed and EXPONENTIAL range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

Small Estimation Sample $T = 200$	EVALUATION CRITERIA							
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
NE: Normal ellipse (3.1)	0.9316	8.5318	0.1691	17.7070	25.4302	0.0373	0.9311	0.0327
Bonferroni rectangle (3.2)-(3.3)	0.9322	8.9649	0.1732	17.6845	24.5118	0.0494	1.1909	0.0332
Modified Bonferroni rectangle (3.4)	0.9329	8.9649	0.1708	17.5768	24.5204	0.0491	1.1835	0.0328
BE: Bootstrap ellipsoid (4.1)	0.9470	9.0687	0.1279	17.7878	27.0149	0.0253	0.6623	0.0280
Bootstrap Bonferroni rectangle (4.2)	0.9473	9.5201	0.1538	19.0599	25.7098	0.0409	1.0245	0.0190
Modified Bootstrap Bonferroni rectangle (4.3)	0.9490	9.5201	0.1457	19.0867	25.7188	0.0407	1.0184	0.0185
Tukey convex hull	0.9362	8.8415	0.1775	18.7765	26.8598	0.0311	0.8061	0.0231

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9368	9.7262	0.1603	18.4753	25.8099	0.0345	0.8731	0.0238
T-NE: T-Normal ellipse (3.6)	0.9316	9.9864	0.1991	17.6350	25.5484	0.0367	0.9196	0.0334
T-Bonferroni rectangle (3.7)	0.9322	10.5660	0.2060	17.6041	25.0697	0.0453	1.1154	0.0341
T-Modified Bonferroni rectangle (3.8)	0.9308	10.7842	0.2258	17.2095	25.1711	0.0448	1.1076	0.0360
T-BE: T-Bootstrap ellipsoid (4.4)	0.9470	10.6754	0.1508	17.7025	27.1477	0.0248	0.6518	0.0284
T-Bootstrap Bonferroni rectangle (4.5)	0.9473	10.0196	0.1607	19.0394	25.8880	0.0397	0.9992	0.0191
T-Modified Bootstrap Bonferroni rectangle (4.6)	0.9501	10.2369	0.1506	19.2279	25.9780	0.0391	0.9879	0.0182
Tukey convex hull	0.9353	9.3801	0.1893	18.5592	27.0879	0.0300	0.7844	0.0235

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}	S_{95}
Analytical method (3.5)	0.9368	9.7297	0.1604	26.0705	45.7636	0.0625	2.8067	0.0200
Bootstrap ellipsoid (4.9)	0.9488	9.5331	0.1753	29.0966	47.0634	0.0520	2.3831	0.0168
Tukey convex hull	0.9353	9.3782	0.1894	26.1874	48.1843	0.0535	2.4796	0.0214

Table 2.9: Evaluation of the h-step ahead 95% prediction regions from a system with **center NORMALLY distributed and EXPONENTIAL range** ($h = 1$); 500 Monte Carlo simulations from a VAR(4). The numbers in parenthesis e.g., (x.x) in the first column are the corresponding equations in the text.

In summary, considering the three systems (center/log-range, center/range, and upper/lower) and assessing the overall performance of the prediction regions by the summary metrics C_{95}, CV_{95}, and POP_{95}, we conclude the following:

1. If the center/log-range system is bivariate normal (case 6.1) or approximately normal (case 6.4) and the estimation sample is large, the prediction regions based on the normal ellipse (2.6) and on the analytical methods (2.10) are the best performers. However, with a small estimation sample, we recommend implementing a bootstrap ellipsoid (2.14) and its transformed regions (2.4.2), and (2.22).

2. If the center/log-range system is not bivariate normal but the joint distribution is symmetric (case 6.2) and the estimation sample is large, any of the bootstrap regions (ellipsoids and Bonferroni rectangles) (2.14, (2.15), and (2.16), their transformed (2.4.2), (2.4.2), (2.19), as well as (2.22) are the best performers. In small samples, a bootstrap ellipsoid (2.14) and its transformed regions (2.4.2), as well as (2.22) are preferred.

3. If the center/log-range system is not bivariate normal and the joint distribution is leptokurtic and asymmetric (case 6.3), for large and small samples, we recommend implementing the bootstrap Bonferroni rectangles (2.15), and (2.16) and their transformed (2.4.2), (2.19), as well as (2.22).

2.7 Prediction Regions for SP500 Low/High Return Interval

We collect the daily intervals of low/high prices of the SP500 index from January 2, 2009 to April 20, 2018 for a total of 2341 observations. Since prices are non-stationary, we construct the daily interval of low/high returns by calculating the daily minimum and

maximum returns with respect to the closing price of the previous day. In this way, we will model stationary intervals. In the Appendix B Table B1, we provide the descriptive statistics of the center, range and log-range of the low/high return intervals. The center average is zero with a standard deviation of 0.64. The center exhibits fat tails with a coefficient of kurtosis of 7 and it is slightly skewed to the left. The range has a mean of 1.15 and a larger standard deviation, 0.83, than the center, it is positively skewed, and it is negatively correlated with the center with a coefficient of correlation of -0.12. The log-transformation of the range corrects the asymmetry and large kurtosis of the range so that log-range is only slightly skewed to the right and has a coefficient of kurtosis of about 3. The coefficient of correlation of center and log-range is about -0.10. The Q-statistics for the center indicate no autocorrelation while those for the range and log-range indicate high autocorrelation. In Figure 2.6, we plot the time series of the center and the range as well as their unconditional bivariate density function. The heavy tails in the center and the almost normality of the log-range are similar distributional characteristics to those of the simulation case in section 6.2 (Student-t(5) center and normal log-range).

We proceed with the modeling of the bivariate system of center/log-range. We split the total sample into an estimation sample from January 2, 2009 to December 31, 2016 (2014 observations) and a prediction/evaluation sample from January 1, 2017 to April 20, 2018 (327 observations). The autocorrelograms of the center seem to indicate no autocorrelation in contrast to those of the log-range that exhibit a profile of an AR(6) with strong memory (see Figure B1 in the Appendix B). These features mimic the autocorrelation that we observe in the end-of-the day returns and in their squared returns when modeling the

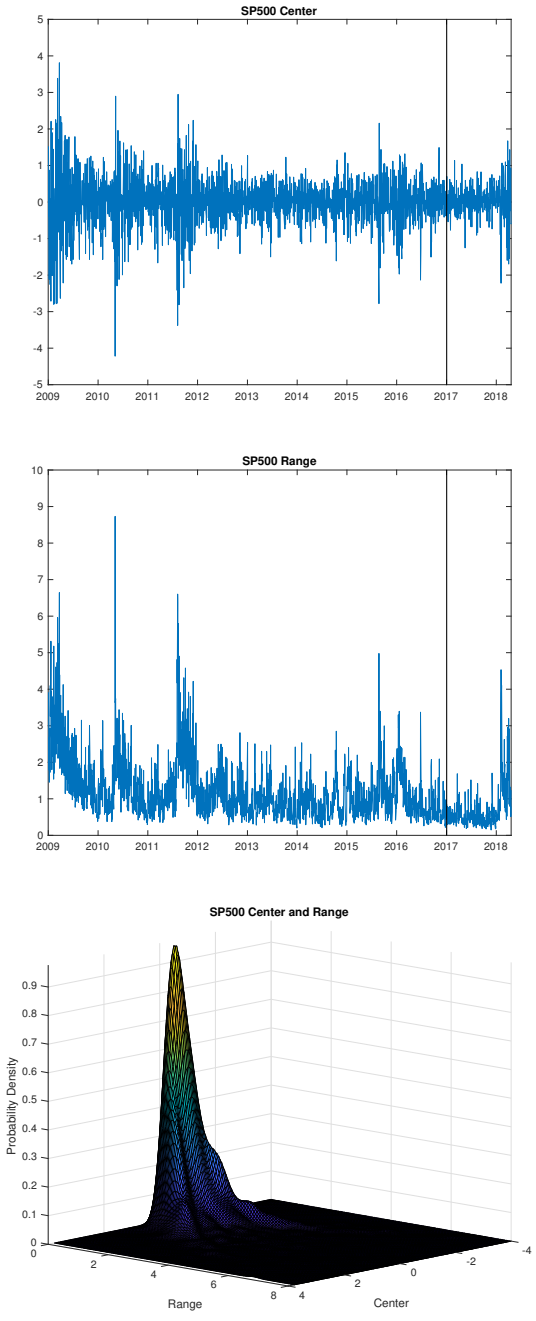


Figure 2.6: Time series plots of center (top panel), range (middle panel) and unconditional bivariate density (bottom panel) of SP500 low/high return interval from January 2, 2009 to April 20, 2018.

conditional variance, which is not very surprising because range or log-range are good proxies for volatility. The SIC also selects a VAR(6) and we proceed with the VAR estimation. The results are presented in Table B2 of the Appendix B. As expected, all the regressors (lagged center and lagged log-range) in the equation for the center are not statistically significant and we re-estimate a restricted VAR where the center equation has only a constant. On the contrary, the equation for the log-range present interesting dynamics. The center Granger-causes the log-range such that the lagged centers are negatively correlated with the current log-range, i.e. positive and large changes in the center return today will predict a narrower range tomorrow. This is similar to a leverage effect in a conditional variance equation. Another relevant aspect is the strong and statistically significant autoregressive nature of log-range in agreement with the ACF/PACF profiles. The goodness of fit for the log-range equation is high with an adjusted R-squared of 52%. The residuals corresponding to this system are all clear of any autocorrelation. The center residuals and log-range residuals are contemporaneous negatively correlated with a correlation coefficient of -0.17. The residuals from the center equation have the same characteristics as the center, that is, are leptukortic with a sample kurtosis of 7 and slightly skewed to the left. The residuals from the log-range equation remain almost symmetric around zero and they have a sample kurtosis of 3. With these characteristics, the conditional joint density of the center and log-range cannot be bivariate normal.

Formally, we test for conditional bivariate normality by implementing the Generalized AutoContouR (G-ACR) (in-sample) tests based on the Probability Integral Transformations (PIT) of the joint density under the null hypothesis of bivariate normality

(González-Rivera and Sun, 2015). In Table B3 of the Appendix B, we report the results of the t-statistics ($t_{k,\alpha}$) that canvas the density from the 1% to the 99% PIT autocontours for lags $k = 1, 2, ...5$. The null hypothesis is strongly rejected at the 5% significance level for mostly all but the 10%, 90% and 95% autocontours. The portmanteau test C_k also reinforces the strong rejection of bivariate normality. In Figure B2 of the Appendix B, we plot the autocontours of the contemporaneous PITs ($\text{center}_t, \text{log-range}_t|\text{center}_t$). Under the correct null hypothesis, the distribution of the PITs should be uniformly distributed within these autocontour squares. It is obvious that this is not the case.

We evaluate the out-of-sample performance of the one-step-ahead 95% prediction regions from January 1, 2017 to April 20, 2018 (327 observations). The results are reported in Tables 2.10. For the system center/log-range, the bootstrap Bonferroni rectangles (2.15) and (2.16) offer the best coverage C_{95} with empirical rates of mostly 95% and they are the most reliable with the lowest average coverage-volume scores CV_{95}. Together with the Tukey convex hull, they also provide the lowest average outlier distance O_{95}. Both rectangles (2.15) and (2.16) also provide the tightest projected one-dimensional regions measured by POP_{95}. For the system center/range, we find that the transformed modified bootstrap Bonferroni rectangle (2.19) is the best performer according to most metrics C_{95}, CV_{95} and POP_{95}. For the system upper/lower bounds, the Tukey convex hull offers the best coverage and the lowest scores for O_{95} and POP_{95}. As expected, the analytic methods (2.10) are not reliable as they tend to undercover. On the contrary, the bootstrap ellipsoid (2.14) and its transformed region (2.4.2), and (2.22) tend to overcover. All these results are very consistent with the Monte Carlo findings of the previous section.

SP500 Low/High Returns	EVALUATION CRITERIA						
CENTER/log-RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}
NE: Normal ellipse (3.1)	0.9541	2.2238	0.0094	1.6077	2.4307	0.0038	0.0093
Bonferroni rectangle (3.2)-(3.3)	0.9450	2.3134	0.0114	1.4657	2.8953	0.0020	0.0058
Modified Bonferroni rectangle (3.4)	0.9480	2.3134	0.0043	1.4534	2.9125	0.0015	0.0045
BE: Bootstrap ellipsoid (4.1)	0.9602	2.3616	0.0252	1.6859	2.5819	0.0027	0.0070
Bootstrap Bonferroni rectangle (4.2)	0.9480	2.4732	0.0040	1.3783	3.1022	0.0009	0.0027
Modified Bootstrap Bonferroni (4.3)	0.9511	2.4732	0.0031	1.3631	3.1222	0.0005	0.0015
Tukey convex hull	0.9450	2.1422	0.0103	1.3317	2.5003	0.0020	0.0049

CENTER/RANGE system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}
Analytical method (3.5)	0.9358	1.8135	0.0319	1.5232	2.2083	0.0085	0.0188
T-NE: T-Normal ellipse (3.6)	0.9541	1.8879	0.0024	1.7135	2.2555	0.0055	0.0125
T-Bonferroni rectangle (3.7)	0.9450	1.9778	0.0180	1.4464	2.5824	0.0039	0.0100
T-Modified Bonferroni rectangle (3.8)	0.9480	1.9904	0.0131	1.4425	2.6327	0.0030	0.0078
T-BE: T-Bootstrap ellipsoid (4.4)	0.9602	2.0147	0.0080	1.8537	2.4211	0.0042	0.0103
T-Bootstrap Bonferroni rectangle (4.5)	0.9480	2.1867	0.0127	1.3101	2.8656	0.0026	0.0074
T-Modified Bootstrap Bonferroni (4.6)	0.9511	2.2146	0.0021	1.2391	2.9464	0.0012	0.0036
Tukey convex hull	0.9450	2.0480	0.0252	1.2361	2.4935	0.0035	0.0088

UPPER/LOWER system	C_{95}	$V^{1/2}$	CV_{95}	O_{95}	P_{95}	OP_{95}	POP_{95}
Analytical method (3.5)	0.9358	1.8112	0.0321	1.6816	3.1735	0.0083	0.0265
Bootstrap ellipsoid (4.9)	0.9602	2.2105	0.0214	1.6737	3.5801	0.0047	0.0170
Tukey convex hull	0.9450	2.0481	0.0253	1.3109	3.6115	0.0027	0.0099

Table 2.10: SP500 Low/High Returns. Evaluation of the one-step ahead 95% prediction regions (Jan.1, 2017-April 20, 2018). In the first column, the numbers in parenthesis e.g., (x.x) are the corresponding equations in the text.

In Figures 2.7 and 2.8, we plot the one-step ahead 95% prediction regions for the center/log-range and center/range systems respectively. We choose six random dates over the prediction sample (March 15, May 11, August 30, December 8, 2017 and February 22, April 6, 2018). In all six dates, the one-step-ahead realized values of the (center, log-range) and (center, range) fall within the regions; only the realized values on December 8, 2017 and April 6, 2018 are slightly more extreme and they fall towards the boundaries of the prediction regions. For the center/log-range system, the normal ellipse and the bootstrap ellipse are very similar but in the center/range system, the bootstrap ellipse tends to be wider adapting to the kurtosis of the center and the asymmetry of the range. The differences among the Bonferroni rectangles are more obvious in the center/range system. In the center/log-range system, the Tukey convex hull has a cone shape over all the six dates though the shape becomes more irregular in the center/range system.

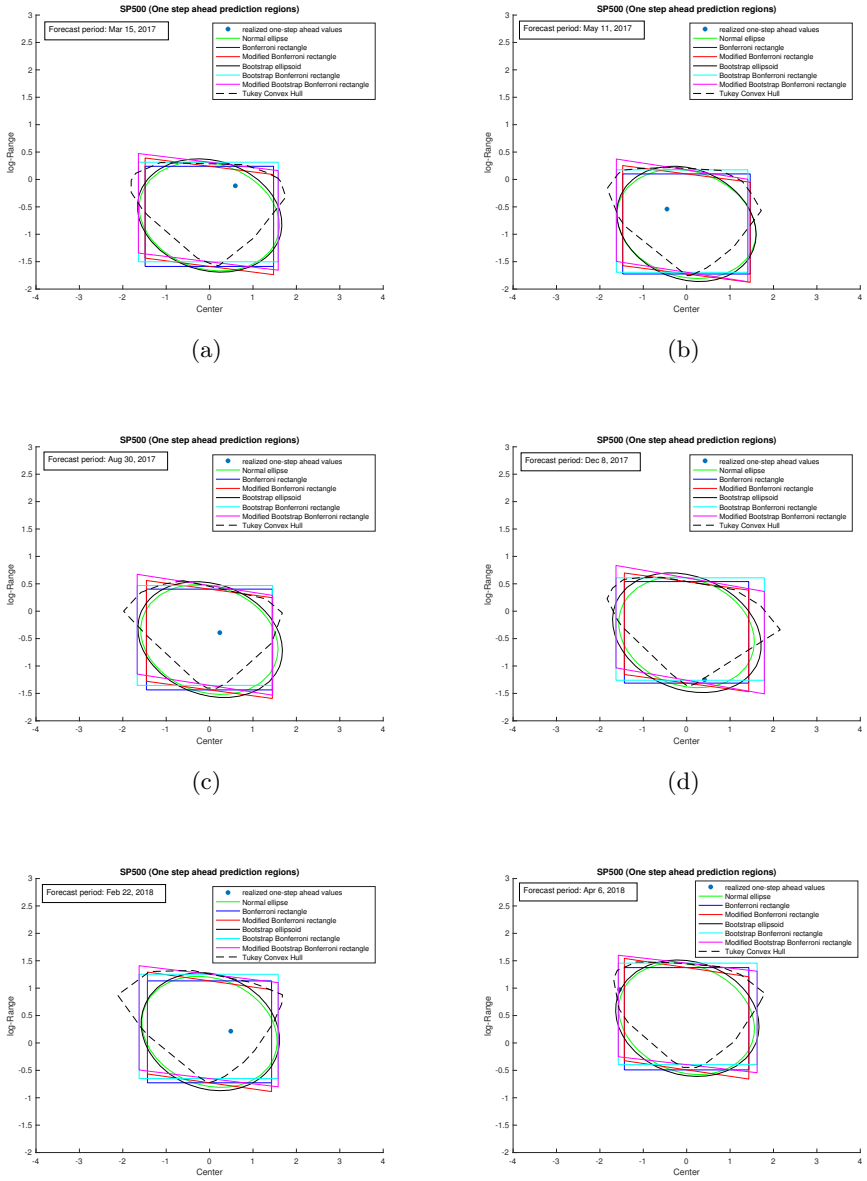


Figure 2.7: One-step-ahead 95% prediction regions for the center/log-range system of the SP500 return intervals corresponding to different dates of the out-of-sample period.

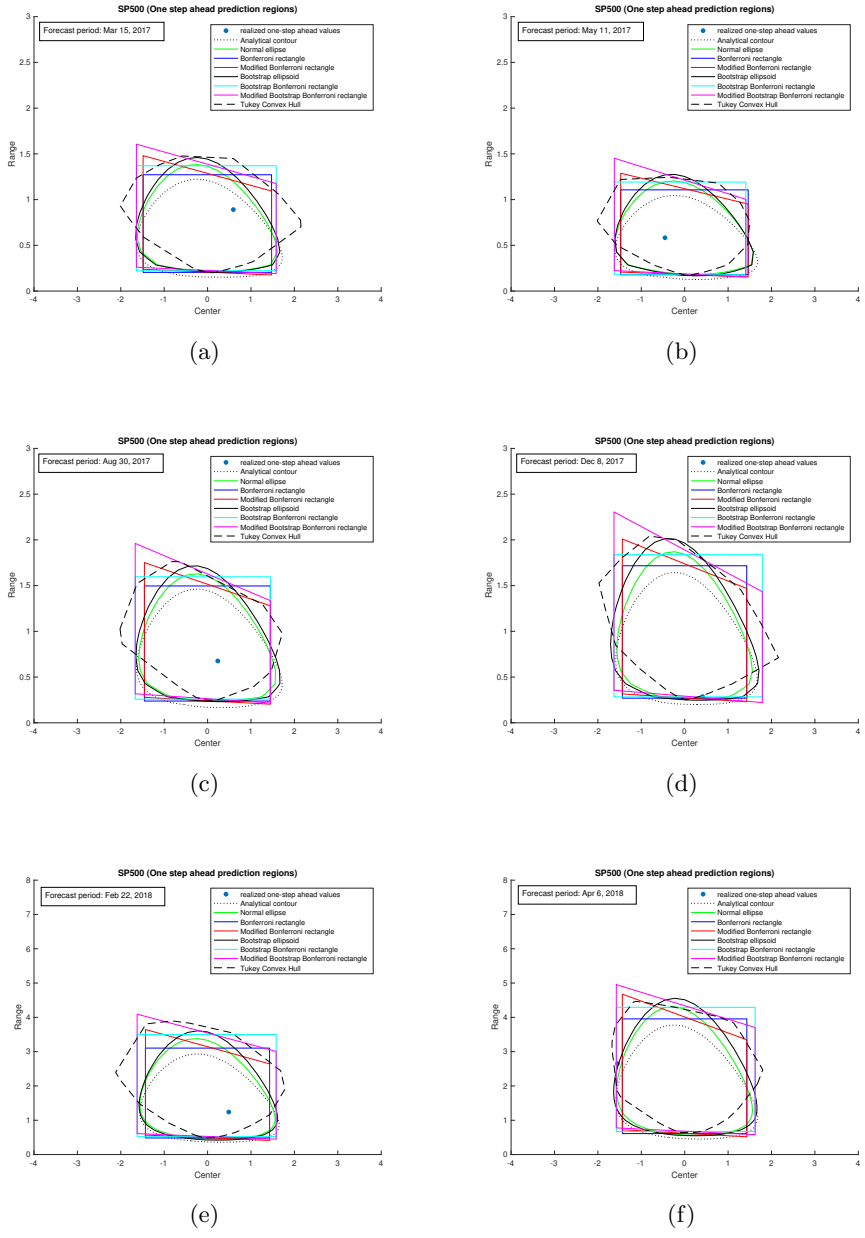


Figure 2.8: One-step-ahead 95% prediction regions for the center/range system of the SP500 return intervals corresponding to different dates of the out-of-sample period.

2.8 A Trading Strategy

We develop a trading strategy based on the prediction regions for the SP500 daily high and low returns. We extend the strategy proposed in He et al. (2010, thereafter HKW) for point forecasts of crude oil high/low prices to account for the probabilistic distribution forecasts of high and low returns. Denote O_t as the opening return at day t, calculated using the opening price at day t with respect to the closing price at day $t-1$. Consider the following ratio $s = \frac{|O_t - \hat{L}_{t+h}|}{|\hat{U}_{t+h} - O_t|}$, where $\hat{L}_{t+h}$ and $\hat{U}_{t+h}$ are the low and the high return forecasts respectively. If the magnitude of $\hat{U}_{t+h} - O_t$ is larger than that of $O_t - \hat{L}_{t+h}$, then the return is more likely to go up than down in the next h days. And if this is observed for several days, it is reasonable to believe that the market is forming an upward trend. Therefore, a "buy alert signal" should be generated (see HKW). Similar argument can be applied to the "sell alert signal". Unlike HKW where the comparison of $\left|\hat{U}_{t+h} - O_t\right|$ and $\left|O_t - \hat{L}_{t+h}\right|$ is based on the point forecasts $\hat{U}_{t+h}$ and $\hat{L}_{t+h}$, we compare the probability of $\left|\hat{U}_{t+h} - O_t\right| > \left|O_t - \hat{L}_{t+h}\right|$ with the probability of $\left|\hat{U}_{t+h} - O_t\right| < \left|O_t - \hat{L}_{t+h}\right|$, which is equivalent to comparing the probabilities of $s < 1$ (buy signal) and $s > 1$ (sell signal). In Figure 2.9, we illustrate the trading strategy. Notice that s is the absolute value of the slope of any line that connects point $A \equiv (O_t, O_t)$ and any other point below the 45 degree line. The ellipse represents the h-step ahead prediction region of the high and low returns. The slope of line AB is equal to one and it is perpendicular to the 45 degree line. Hence, the area under the 45 degree line can be divided by the line AB into two areas: $s > 1$ to the left of line AB, and $s < 1$ to the right of line AB. Therefore, counting the bootstrap realizations in the two subareas of the prediction region, we can estimate the probability of

$s < 1$ with that of $s > 1$ for a given $(1-\alpha)\%$ confidence region. Then, the trading strategy consists of the following steps:

- At day t, plot Figure 1 based on O_t and the h-step ahead prediction region of high and low returns. Within the prediction region, if the number of bootstrap realizations (obtained as in 3.4 and 3.5) on the right hand side of the line AB is larger than that on the left hand side of the line AB, a "buy alert signal" is generated.

- Buy the asset on day $t+m-1$ using the closing price on that day if the "buy alert signal" is observed for m consecutive days beginning with day t.

- After buying the asset, on any other day d, watch for the"sell alert signal", that is, the number of bootstrap realizations on the left hand side of the line AB should be larger than that on the right hand side of the line AB within the prediction region.

- Sell the asset on day $d+m-1$ using the closing price on that day if the "sell alert signal" is observed for m consecutive days beginning with day d. Otherwise, hold the asset.

We evaluate this trading strategy by considering the forecasts of the SP500 high/low returns for the out-of-sample period (Jan. 1, 2017 to Apr. 20, 2018, 327 observations). We consider the bootstrap ellipse (BE^{UL}_{T+h}) and the Tukey Convex Hull (TH^{UL}_{T+h}) prediction regions with a 95% nominal coverage rate. For the implementation, the choice of m should not be too small because it will introduce substantial noise in trading but it should not be too large either because we could miss profitable trades. We consider $m = 1, 2, 3, 4$ and $h = 1, 2, 3$. We apply a transaction cost of 0.1%, and we annualize the profit/loss for

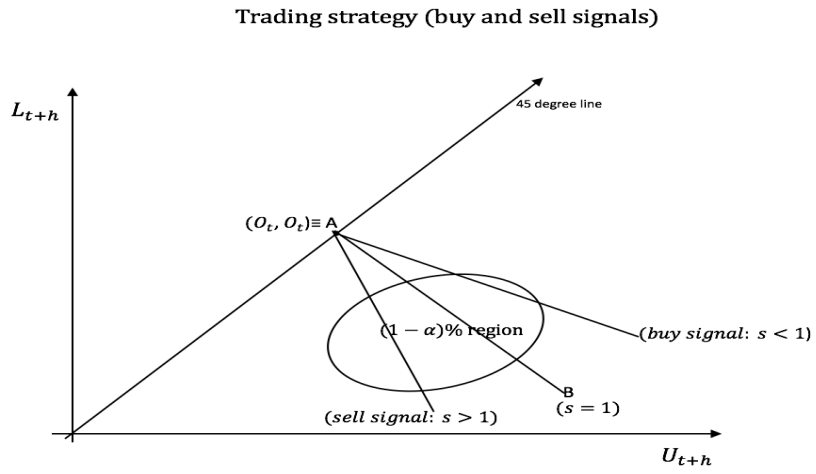


Figure 2.9: Buy and sell signals from trading strategy.

each trade because each trade will have a different holding period.The annualized return is calculated as $AR = (\frac{C_{t+j}-C_t}{C_t} - 0.001)(\frac{365}{j})$ where C_t and C_{t+j} $(j > 0)$ are the closing prices for the buying and selling days respectively. The investor can buy the asset again before the previous bought asset is sold. At the end of the evaluation period, if there are still assets that have not been sold, these assets will not be considered when calculating the profits.

In Table 2.11, we report the averaged annualized returns, the max and min annualized returns, and the percentage of trade with positive returns for all cases. HKW is the trading strategy in He et al. (2010) based on the point forecasts of the high and low returns obtained by the bootstrap algorithm applied to our estimated model. For all cases but two, the averaged annualized returns are positive. The choice of m is very relevant because the gap between the max and min annualized returns narrows as m increases for all h. A large m means that the investor is looking for a stronger signal and, though she may miss some trades with extreme positive returns, she will also avoids those extreme negative returns that can be catastrophic. There is also a monotonic positive relation between m and the percentage of trades with positive returns. For average annualized returns, the performance of BE is better than that of TH in most cases, and the performance of BE or TH is better than HKW in particular when $m = 4$.

	m=1			m=2			m=3			m=4		
	HKW	BE	TH	HKW	BE	TH	HKW	BE	TH	HKW	BE	TH
h=1												
Averaged annualized returns	26.16%	33.12%	33.14%	-5.71%	10.89%	5.94%	43.49%	42.52%	38.04%	38.76%	40.25%	40.25%
max annualized returns	853.58%	853.58%	853.58%	239.83%	239.83%	239.83%	155.24%	155.24%	155.24%	85.18%	85.18%	85.18%
min annualized returns	-3760.12%	-3760.12%	-3760.12%	-1621.35%	-1621.35%	-1621.35%	-116.43%	-116.43%	-116.43%	4.64%	4.64%	4.64%
% of trades with positive returns	61.04%	61.54%	61.54%	78.26%	81.48%	80.77%	75.00%	75.00%	75.00%	100%	100%	100%
h=2												
Averaged annualized returns	29.86%	30.39%	29.15%	2.02%	-1.53%	0.79%	45.44%	48.10%	44.43%	38.76%	38.76%	38.76%
max annualized returns	853.58%	853.58%	853.58%	239.83%	239.83%	239.83%	155.24%	155.24%	129.53%	85.18%	85.18%	85.18%
min annualized returns	-3760.12%	-3760.12%	-3760.12%	-1621.35%	-1621.35%	-1621.35%	-116.43%	-116.43%	-116.43%	4.64%	4.64%	4.64%
% of trades with positive returns	61.54%	62.82%	61.25%	76.00%	76.00%	76.00%	75.00%	71.43%	71.43%	100%	100%	100%
h=3												
Averaged annualized returns	29.92%	11.48%	3.37%	4.55%	9.53%	3.22%	42.37%	20.67%	17.26%	43.01%	63.75%	61.42%
max annualized returns	853.58%	853.58%	853.58%	239.83%	239.83%	239.83%	175.56%	175.56%	175.56%	85.18%	102.78%	102.78%
min annualized returns	-3760.12%	-3760.12%	-3760.12%	-1621.35%	-1621.35%	-1621.35%	-116.43%	-116.43%	-116.43%	20.14%	26.47%	26.47%
% of trades with positive returns	62.67%	58.97%	57.69%	80.77%	78.57%	77.78%	88.89%	71.43%	57.14%	100%	100%	100%

Table 2.11: Trading strategy comparison. SP500 average annualized returns over the out-of-sample period Jan. 1, 2017 to April 20, 2018.

2.9 Conclusion

The interest in interval data arises because interval measurements offer a more complete description of a data set. In time series, each time realization has joint information on the level and the dispersion of the process under study. However, statistical analysis of interval-valued data requires that the natural order of the interval is preserved. Though there are several works that consider the problem of estimation with constraints, we are not aware of any work that considers the construction of forecasts for interval-valued data satisfying the natural constraint in each period of time, i.e. lower bound is not larger than the upper bound, or equivalently, the range of the interval must be strictly positive. Our contribution lies on approximating a probabilistic forecast of an interval-valued time series by offering alternative approaches to construct bivariate prediction regions of the center and the range, or the lower and upper bounds, of the interval.

To overcome the positive constraint of the range, we have estimated a Gaussian bivariate system for the center/log-range system, which also delivers QML properties for our

estimators. However, the interest of the researcher is not the prediction of the center/log-range but the center/range or upper/lower bounds of the interval. By implementing either analytical or bootstrap methods we have directly transformed the prediction regions for the center/log-range system into those for the center/range and upper/lower bounds systems. It is important to remark that we do not focus on point forecast purposely. By focusing on prediction regions rather than on point forecasts, we avoid the biases that are associated with the exp-transformation of the point forecasts of log-transformed variable. In this case, bias-correction techniques are necessary if one's interest is the conditional mean of the future variable. A prediction region for the center/log-range does not need any bias correction when we transform it to a prediction region of the center/range system because the quantile is preserved under a monotonic transformation like the exp-transformation. However, these transformed prediction regions can have very irregular shapes even in the most straightforward scenario of bivariate normality of the center/log-range system. If a central point forecast is of interest, the researcher can always calculate the centroid of the region.

Beyond the standard coverage rate, we have proposed several new metrics to evaluate the performance of different prediction regions. We have introduced a notion of risk to the evaluation of the regions by considering the location of the out-of-the-region outcomes with respect to some central point in the region. The researcher would like to minimize risk once the empirical coverage of the region is close to the nominal coverage. We have considered Gaussian and non-Gaussian systems and our recommendation leans towards bootstrap methods, even for Gaussian systems. Bootstrap ellipsoids and their transformed are best

when the joint distribution of the center/log-range system is symmetric. If it is not, then bootstrap Bonferroni rectangles will be preferred.

We have analyzed the time series of the daily low/high return interval of the SP500 index. We modeled and predicted the *joint* conditional density of the return level and the return volatility. We showed that the construction of several prediction regions of the center and range of the return interval do not require strong parametric distributional assumptions. We also developed a trading strategy based on the constructed prediction regions for SP500 high and low daily returns, and showed its profitability.

Bibliography

[1] Alizadeh, S., M. W. Brandt, & F.X. Diebold (2002). Range-based Estimation of Stochastic Volatility Models. *Journal of Finance*, 57(3), pp. 1047-1091.

[2] Ariño, M.A., & P. Franses (2000). Forecasting Levels of Vector Autoregressions Log-transformed Time Series. *International Journal of Forecasting*, 16, pp. 111-116.

[3] Bardsen, G., & H. Lutkepohl (2011). Forecasting Levels of Log Variables in Vector Autoregressions. *International Journal of Forecasting*, 27, pp. 1108-1115.

[4] Beran, R. (1993). Probability-Centered Prediction Regions. *Annals of Statistics*, 21(4), pp. 1967-1981.

[5] Blanco-Fernádez, A. & P. Winker (2016). Data Generation and Statistical Management of Interval Data. *AStA Advances in Statistical Analysis*, 10(4), pp. 475-494.

[6] Cavaliere, G., H.B. Nielsen & A. Rahbek (in press). Bootstrapping Non-causal Autoregressions with Applications to Explosive Bubble Modeling. *Journal of Business & Economic Statistics*.

[7] Chernozhukov V., A. Galichon, M. Hallin & M. Henry (2017). Monge-Lontorovich Depth, Quantiles, Ranks and Signs.*Annals of Statistics*, 45(1), pp. 223-256.

[8] Fresoli, D., E. Ruiz, & L. Pascual (2015). Bootstrap Multi-step Forecasts for Non-Gaussian VAR Models. *International Journal of Forecasting*, 31, pp. 834-848.

[9] Golestaneh, F., P. Pinson, R. Azizipanah-Abarghooee, & H.B. Gooi (2018). Ellipsoidal Prediction Regions for Multivariate Uncertainty Characterization. *IEEE Transactions on Power Systems*, 33(4), pp. 4519-4530.

[10] González-Rivera, G., & W. Lin (2013). Constrained Regression for Interval-valued Data. *Journal of Business and Economic Statistics*, 31(4), pp. 473–490.

[11] González-Rivera, G., & Y. Sun (2015). Generalized Autocontours: Evaluation of Multivariate Density Models. *International Journal of Forecasting*, 31(3), pp. 799-814.

[12] ranger, C.W.J. & P. Newbold (1976). Forecasting Transformed Series. *Journal of the Royal Statistical Society, B*, 38, pp. 189-203.

[13] Green, W. (1985). Peeling Data. *Encyclopedia of Statistical Sciences*, 6, pp. 660-664.

[14] Guerrero, V.M. (1993). Time Series Analysis Supported by Power Transformation. *Journal of Forecasting*, 12, pp. 37-48.

[15] He, A.W.W., Kwok, J.T.K, & Wan, A.T.K (2010). An empirical model of daily highs and lows of West Texas Intermediate crude oil prices. *Energy Economics*, 32, pp. 1499-1506.

[16] Hyndman, R.J. (1996). Computing and Graphing Highest Density Regions. *The American Statistician*, 50, pp. 120-126.

[17] LePage, R. & K. Podgorski (1996). Resampling Permutations in Regressions without a Second Moment. *Journal of Multivariate Analysis*, 57, pp. 119-141.

[18] Lima Neto, E., & F. de Carvalho (2010). Constrained Linear Regression Models for Symbolic Interval-Valued Variables. *Computational Statistics and Data Analysis*, 54, pp. 333-347.

[19] Liu, X. & Y. Zuo (2015). CompPD: A MATLAB Package for Computing Projection Depth. *Journal of Statistical Software*, 65(2).

[20] Lutkephol, H. (1991). *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

[21] Mayr, J., & D. Ulbricht (2015). Log versus Level in VAR Forecasting: 42 Million Empirical Answers. Expect the Unexpected. *Economics Letters*, 126, pp. 40-42.

[22] Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business*, 53(1), pp. 61-65.

[23] Pascual L., J. Romo & E. Ruiz (2005). Bootstrap Prediction Intervals for Power Transformed Time Series. *International Journal of Forecasting*, 21, pp. 219-235.

[24] Tu, Y., & Y. Wang (2016). Center and Log Range Models for Interval-valued Data with An Application to Forecast Stock Returns. Working paper.

[25] Tukey, J. (1975). Mathematics and Picturing Data, in *Proceedings of the 1975 International Congress of Mathematics*, 2, pp. 523-531.

[26] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrics*, 50, pp.1-25.

[27] Yao, W., & Z. Zhao (2013). Kernel Density-based Linear Regression Estimate. *Communications in Statistics. Theory and Methods*, 42(24), pp. 4499-4512.

[28] Zuo, Y. (2003). Projection-based Depth Functions and Associated Medians. *Annals of Statistics*, 31(5), pp. 1460-1490.

[29] Zuo Y. (2013). Multidimensional Medians and Uniqueness. *Computational Statistics & Data Analysis*, 66, pp. 82-88

Chapter 3

A Truncated Mixture Transition Model for Interval-valued Time Series

3.1 Introduction

Interval data refers to data sets where the observation is an interval in contrast to a single point. Intervals arise in a variety of situations. There are instances when the data is directly collected in interval format. A standard example is survey design that avoids asking participants about private or sensitive information, e.g. income, and the answer is provided in interval format, e.g. [$50K, $100K]. In these cases, interval data is the only data format available to the researchers. In other instances, intervals arise as a result of aggregating data. The data may be collected at the individual level, e.g., gas

prices in a gas station, but the research question deals with a larger unit, e.g., gas prices at the county level. Rather than providing an average of gas station prices, aggregating the data in interval format for each county is more informative because it preserves the internal price variation of each county. Financial data, e.g., tick-by-tick stock transaction data recorded at the ultra-high frequency, generally collapses to a lower frequency single point, e.g. the daily closing price. Aggregating the data into intervals, e.g. daily max/min price interval, is more useful because it provides information on both the price level and the daily price volatility. A similar example is the interval of daily low/high temperature that provides relevant information for decision making. Finally, intervals can also arise because there is uncertainty on the measurement of the variable of interest. Regardless of the data generation mechanism of intervals, we define an interval-valued time series (ITS) as a collection of interval data observed over time.

The literature on modeling interval data and ITS can be divided into two categories depending on the data representation: the center/range system (e.g. center and range are respectively the midpoint and the distance between the upper and lower bounds.) or the upper/lower bound system. In the center/range system, the interval constraint is that the range cannot be smaller than zero. Lima Neto and de Carvalho (2010) propose modeling center and range separately while imposing non-negative constraints on the parameters of the range equation, which are unnecessarily too restrictive and complicate the estimation of the system. Tu and Wang (2016) overcome this restriction by log-transforming the range. However, it requires bias correction for the conditional mean and can fail when zero is present in range data. In the upper/lower bound system, an equivalent interval constraint

is that the upper bound cannot be smaller than the lower bound. González-Rivera and Lin (2013) propose a constrained regression model (GL) that preserves this natural order of the interval. They assume that the bivariate errors of the system of bounds follow a bivariate truncated normal distribution, where the truncation encloses the constraint that the upper bound is not smaller than the lower bound. However, this assumption is restrictive as the consistency of the estimators heavily depends on it.

The previous literature explores a variety of ways to preserve the interval constraint, and mainly focuses on modeling the conditional mean of ITS. To the best of our knowledge, none of the existing work has considered modeling the potential conditional heteroskedasticity in ITS, a feature that has been widely recognized in point-valued time series (PTS). One exception is that GL may produce conditional heteroskedasticity as a byproduct. In addition, many PTS exhibit non-Gaussian features that may also appear in ITS, such as flat streches, burst of activities, outliers and changepoints (see e.g., Le et.al. 1996, Wong and Li 2000), opening a door for models capable of generating more flexible predictive densities, an issue that has not been addressed in the current ITS literature. By contrast, there is a vast amount of literature on modeling conditional heteroskedasticity and non-Gaussian behaviors for PTS. Particularly, Le et. al. (1996) propose a Mixture Transition Distribution (MTD) model for the univariate PTS that seeks to account for the non-Gaussian features. Their idea is to specify the conditional distribution for the variable of interest as a mixture distribution, where each component contains only one lag from the information set. The fact that MTD is able to handle conditional heteroskedasticity is noted and discussed by Berchtold and Raftery (2002). MTD is generalized by Wong

and Li (2000) under the name of Mixture Autoregressive (MAR) model to entertain more flexibility by allowing each component to depend on the full information set. Hassan and Lii (2006) extend MTD for the marked point process under a bivariate setting.

In this chapter, we propose a model for ITS in the upper/lower bound system in the spirit of the MTD model. We specify the joint distribution of the upper bound (x_t) and lower bound (y_t) conditional on the information set as a mixture of truncated bivariate normal distribution, where for each component the bivariate normal distribution is truncated at $x_t \geq y_t$. The information set enters the conditional distribution as a linear function through the pseudo location parameter of the truncated bivariate normal distribution for each component.[1] The model comes with several benefits. First, it can preserve the natural order of ITS, that is, the upper bound not smaller than the lower bound. Second, it can capture conditional heteroskedasticity without modeling it explicitly, as the dynamics enter the covariance matrix via the truncation and the mixture framework. Third, the mixture distribution that the model based on provides great flexibility to approximate the underlying true conditional distribution, and hence can improve the quality of density forecast.

It is well known that the maximum likelihood estimator (MLE) does not have a closed-form solution for mixture models resulting from the complexity of the likelihood. In the literature, EM algorithm is a standard device to find the MLE for mixture models due to its simplicity and monotonicity in the likelihood (see e.g. Hamilton 1990, Le et. al. 1996, Hassan and Lii 2006). However, such a standard EM algorithm fails in our model as no closed-form solution can be obtained in the M step. This is caused by the

[1]The pseudo location parameter of a truncated bivariate normal distribution can be interpreted as the location parameter of the bivariate normal distribution (before the truncation). It is called pseudo because it no longer represents the mean (location) of the truncated distribution after the truncation is imposed.

normalizing factor in the truncated normal distribution, which possesses a complex form after taking the derivative. To overcome this problem, we propose a new EM algorithm.[2] The innovation is made by constructing a high level pseudo complete data generating process that brings in more latent variables than the standard EM algorithm. Specifically, at each time the observation is generated in four steps. First, a membership variable (latent) is generated from a multinomial distribution that suggests which component the observation truly comes from. Second, conditional on the observation coming from the component indicated in the previous step, a variable (latent) is obtained from a geometric distribution that indicates the number of invalid observations ($x_t < y_t$) before the occurrence of a valid observation ($x_t \geq y_t$). Third, generate the corresponding number of invalid observations (latent) independently from the area of the bivariate normal distribution where $x_t < y_t$. Fourth, draw one observation from the area of the bivariate normal distribution where $x_t \geq y_t$, and treat it as the valid observation. The Monte Carlo simulations indicate that the new EM algorithm performs well with the finite sample. We show that the MLE is consistent under some regular assumptions. We apply the model to IBM daily stock return ITS and show that it outperforms the competing models.

The organization of the chapter is as follows. In Section 2, we introduce the truncated mixture transition model and discuss its properties. In Section 3, we propose the new EM algorithm. In Section 4, we show the consistency of the MLE. In Section 5, we perform the Monte Carlo simulations. In Section 6, we apply our model to IBM daily stock return ITS. We conclude in Section 7.

[2]The new EM algorithm can be applied generally to data sets with any kinds of truncation either in the time series setting or for cross-sectional probability clustering.

3.2 The Truncated Mixture Transition Model

3.2.1 Definition

Let x_t be the upper bound, and y_t be the lower bound of the interval observed at time t. The interval time series data has the following format

$$\{ \quad (x_t, y_t), t = 1, \ldots T \quad \}$$

where by construction $x_t \geqslant y_t$, and we denote $Y_t = (x_t, y_t)'$ hereafter. We say that Y_t is generated by a truncated mixture transition ($TMT(P,Q)$) model if its conditional density function given the past information set can be written as

$$f(Y_t|\mathcal{F}^{t-1}) = \sum_{j=1}^{P} \alpha_j f_j(Y_t|Y_{t-Q}^{t-1}) \tag{3.1}$$
$$\sum_{j=1}^{P} \alpha_j = 1, \alpha_j > 0, j = 1, \ldots, P$$

where P is the number of components and is assumed to be fixed, and Q is the number of lags in each component.[3] $\mathcal{F}^{t-1}$ is the information set up to time $t-1$, and $Y_{t-Q}^{t-1} = (Y_{t-Q}, Y_{t-Q+1}, ..., Y_{t-1})$. $f_j(Y_t|Y_{t-Q}^{t-1})$ is a truncated bivariate normal probability density function truncated at $x_t \geqslant y_t$. That is, for each component, the upper bound is not smaller than the lower bound. The truncated density has the following form (see e.g. Nath

[3]The analysis in this chapter can be modified to accommodate the case where Q is allowed to be component specific.

1972)

$$f_j(Y_t|Y_{t-Q}^{t-1}) = \frac{1}{2\pi|\Sigma_j|F_{t,j}} exp[-\frac{1}{2}(Y_t - \mu_{t,j})'\Sigma_j^{-1}(Y_t - \mu_{t,j})] \tag{3.2}$$

where $\mu_{t,j} = C_j + B_{j,1}Y_{t-1} + ... + B_{j,Q}Y_{t-Q}$, C_j (2×1) is a constant vector, $B_{j,r}$ (2×2) $(r = 1, ..., Q)$ is a matrix, Σ_j (2×2) is a positive semi-definite matrix, and $|A|$ is the determinant of matrix A. (3.2) differs from a bivariate normal distribution in the extra normalization term: $F_{t,j} = 1 - \Phi(\frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}})$, which represents the cumulative distribution of the truncated area $(x_t \geq y_t)$. Φ is the standard normal cumulative distribution function, and $w = (1, -1)'$.

3.2.2 Theoretical properties

Given the definition above, we can write down the conditional mean of Y_t:

$$E(Y_t|\mathcal{F}^{t-1}) = \sum_{j=1}^{P} \alpha_j (M_{o,t,j}^1 + \mu_{t,j}) \tag{3.3}$$

where

$$M_{o,t,j}^1 = \frac{\Sigma_j w}{\sqrt{w'\Sigma_j w}} \frac{\phi(\frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}})}{1 - \Phi(\frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}})} \tag{3.4}$$

ϕ is the standard normal density function. Unlike the normal density, where $\mu_{t,j}$

is the mean for component j, the additional term, $M^1_{o,t,j}$, represents the mean shift after the truncation (see Nath 1972 for moments of truncated normal distribution). As a result, the conditional mean is no longer $\mu_{t,j}$ but a nonlinear function of $\mathcal{F}^{t-1}$. We also show that the natural order of interval time series is preserved at the conditional mean level: $w'E(Y_t|\mathcal{F}^{t-1})) \geq 0$. The proof can be found in Appendix B.1.

A promising feature of TMT model is that it can produce a time-varying conditional variance to capture conditional heteroskedasticity. To see this, the conditional variance is given by:

$$
\begin{aligned}
& V(Y_t|\mathcal{F}^{t-1}) \qquad (3.5)\\
=& E(Y_tY_t'|\mathcal{F}^{t-1}) - E(Y_t|\mathcal{F}^{t-1})E(Y_t|\mathcal{F}^{t-1})' \\
=& \sum_{j=1}^{P} \alpha_j (M^2_{o,t,j} + \mu_{t,j}(M^1_{o,t,j})' + M^1_{o,t,j}\mu_{t,j}' + \mu_{t,j}\mu_{t,j}') \\
& -(\sum_{j=1}^{P} \alpha_j (M^1_{o,t,j} + \mu_{t,j}))(\sum_{j=1}^{P} \alpha_j (M^1_{o,t,j} + \mu_{t,j}))'
\end{aligned}
$$

where

$$
M^2_{o,t,j} = \Sigma_j + \frac{\Sigma_j w w' \Sigma_j}{w'\Sigma_j w} \frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}} \frac{\phi(\frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}})}{1-\Phi(\frac{-w'\mu_{t,j}}{\sqrt{w'\Sigma_j w}})} \qquad (3.6)
$$

3.3 Estimation

In this section, we discuss the estimation of the TMT model using maximum likelihood (ML). The goal is to estimate the set of parameters $\Psi = \{\alpha_j, A_j, \Sigma_j | \forall j\}$ by maximizing the likelihood:

$$L(\Psi) \quad = \tfrac{1}{T-Q} \quad \sum_{t=Q+1}^{T} log[\sum_{j=1}^{P} \alpha_j f_j(Y_t | Y_{t-Q}^{t-1}, B_j, \Sigma_j)] \tag{3.7}$$

where $A_j = (C_j, B_{j,1}, ..., B_{j,Q})$. We first consider an unconditional version of (3.7), where $\mu_{t,j} = \mu_j$ doesn't depend on the information set. The corresponding log-likelihood function for $\Theta = \{\alpha_j, \mu_j, \Sigma_j | \forall j\}$ can be written as

$$L(\Theta) \quad = \quad \frac{1}{T} \sum_{t=1}^{T} log[\sum_{j=1}^{P} \alpha_j f_j(Y_t | \mu_j, \Sigma_j)] \tag{3.8}$$

Estimating Θ is easier than Ψ because the conditional distribution of Y_t doesn't depend on the information set and can be viewed as if the data is drawn i.i.d. from the mixture distribution. Therefore, we will first illustrate the ML estimation of (3.8) and then (3.7).

Clearly, no closed-form solution can be obtained from maximizing (3.8). In fact, the likelihood functions of mixture models are usually non-concave, and often have several local maxima (see e.g. Redner and Walker 1984). Dempster et. al. (1977) propose the expectation maximization (EM) algorithm, and it has been widely applied to find the ML

estimators for mixture models due to its simplicity and monotonicity property (see Dempster et. al. 1977), e.g., Hamilton (1990) uses EM algorithm to estimate the regime switching model. The statistical properties of EM algorithm have been studied extensively in the literature (see e.g. Wu 1983, Meng 1994, McLachlan and Krishnan 2007, and Balakrishnan, et. al. 2017).

A review of the EM algorithm for normal mixture models in unconditional setting (each $f_j(.)$ in (3.8) represents a normal distribution) can be found in Appendix B.2. Lee and Scott (2010) apply the EM algorithm to a truncated normal mixture model with each component truncated by a rectangle, e.g., $s \leq Y_t \leq k$, where s and k are vectors with the same dimension as Y_t. Although our model has a different type of truncation ($x_t \geq y_t$, or $w'Y_t \geq 0$) , their arguments can be adapted to derive an EM algorithm. However, this EM algorithm fails to have a closed-form solution in the M step, mainly due to the truncation term ($\frac{\phi(.)}{1-\Phi(.)}$) in the density (see Appendix B.3 for details). As a result, numerical maximization is needed in M step (see e.g. Lange 1995), sacrificing the simplicity of the EM algorithm. In the following, we propose a new EM algorithm that solves this problem.

3.3.1 A new EM algorithm for truncated normal mixture model (unconditional case)

The new EM algorithm begins by transforming the data generating process into a missing data framework as follow. To obtain the observation Y_t, a latent variable z_t is generated from a multinomial distribution, indicating which component the observation truly comes from. Next, conditional on z_t, another latent variable n_t can be generated from a geometric distribution. n_t represents the number of invalid draws ($x_t < y_t$) from

the respective component (a bivariate normal distribution) before a valid draw ($x_t \geq y_t$) arrives. The valid draw (the $(n_t + 1)^{th}$ draw) is then treated as the t^{th} observation (Y_t). In other words, only the valid draw can be observed while all the invalid draws (if any) are latent. Denote $Y_t^A = \{Y_{t,1}, Y_{t,2}, ..., Y_{t,n_t}, Y_{t,n_t+1}\}$ as all the draws for time t. We now formalize the above data generating process.

Let z_t follow a multinomial distribution:

$$g(z_t|\Theta) = \prod_{j=1}^{P} \alpha_j^{z_{tj}} \tag{3.9}$$

Given the role n_t plays in the above pseudo complete data generating process, it is natural to specify its distribution conditional on z_t as a geometric distribution, a discrete probability distribution that describes the number of failures before the first occurrence of success.

$$q(n_t|z_t, \Theta) = \prod_{j=1}^{P} \Big[(1 - F_j)^{n_t} F_j\Big]^{z_{tj}} \tag{3.10}$$

where $F_j = 1 - \Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})$ is the cumulative distribution for the truncated area ($x_t \geq y_t$) for component j, and represents the probability of getting a valid draw from the bivariate normal distribution. Then, the conditional density of Y_t^A is specified as below

$$h(Y_t^A|z_t,n_t,\Theta) = \prod_{j=1}^{P}\Big[\frac{f_j^N(Y_{t,n_t+1})}{F_j}\prod_{k=1}^{n_t}(\frac{f_j^N(Y_{t,k})}{1-F_j})\Big]^{z_{tj}} \tag{3.11}$$

where $f_j^N(.)$ is the bivariate normal density of component j. Next, the joint density function of the pseudo complete data ($\{Y_t^A, z_t, n_t\}$) can be constructed,

$$\begin{aligned} l(Y_t^A,z_t,n_t|\Theta) &= g(z_t|\Theta)q(n_t|z_t,\Theta)h(Y_t^A|z_t,n_t,\Theta) \\ &= \prod_{j=1}^{P}\Big[\alpha_j f_j^N(Y_{t,n_t+1})\prod_{k=1}^{n_t} f_j^N(Y_{t,k})\Big]^{z_{tj}} \end{aligned} \tag{3.12}$$

and we can write down the pseudo complete log-likelihood function.

$$L^C(\Theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} z_{tj}[log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \sum_{k=1}^{n_t} logf_j^N(Y_{t,k})] \tag{3.13}$$

E Step: the above likelihood (3.13) is replaced with its conditional expectation. See Appendix B.4 for details.

$$\begin{aligned} &Q(\Theta|\Theta^l) \\ =&E(L^C(\Theta)|Y,\Theta^l) \\ =&\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj}[log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \tilde{n}_{t,j}(\int logf_j^N(Y_{t,k})(\frac{f_j^{N,l}(Y_{t,k})}{1-F_j^l})dY_{t,k})] \end{aligned} \tag{3.14}$$

where $f_j^{N,l}(.)$ and F_j^l are respectively $f_j^N(.)$ and F_j conditional on Θ^l (the parameter set of the previous (l^{th}) iteration). $\tilde{n}_{t,j} = E(n_t|z_{tj} = 1, Y, \Theta^l) = \frac{1-F_j^l}{F_j^l}$, and

$$\begin{aligned} \tilde{z}_{tj} &= P(z_{tj} = 1|Y, \Theta^l) \\ &= \frac{P(z_{tj} = 1, Y_t|\Theta^l)}{P(Y_t|\Theta^l)} \\ &= \frac{\alpha_j^l f_j^l(Y_t)}{\sum_{r=1}^P \alpha_r^l f_r^l(Y_t)} \end{aligned} \tag{3.15}$$

M Step: We can obtain a closed-form solution by maximizing $Q(\Theta|\Theta^l)$. See Appendix B.5 for details.

$$\alpha_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj}}{N} \tag{3.16}$$

$$\mu_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj}(Y_t + \tilde{n}_{t,j}(M_{d,j}^{1,l} + \mu_j^l))}{\sum_{t=1}^T \tilde{z}_{tj}(1 + \tilde{n}_{t,j})} \tag{3.17}$$

$$\Sigma_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj}[(Y_t - \mu_j^{l+1})(Y_t - \mu_j^{l+1})' + \tilde{n}_{t,j} M_{d',j}^2]}{\sum_{t=1}^T \tilde{z}_{tj}(1 + \tilde{n}_{t,j})} \tag{3.18}$$

where $M_{d',j}^2 = M_{d,j}^{2,l} + (\mu_j^l - \mu_j^{l+1})(M_{d,j}^{1,l})' + (M_{d,j}^{1,l})(\mu_j^l - \mu_j^{l+1})' + (\mu_j^l - \mu_j^{l+1})(\mu_j^l - \mu_j^{l+1})'$. $M_{d,j}^{1,l}$ and $M_{d,j}^{2,l}$ are respectively $M_{d,j}^1$ and $M_{d,j}^2$ conditional on Θ^l.

$$M_{d,j}^1 = \frac{-\Sigma_j w}{\sqrt{w'\Sigma_j w}} \frac{\phi(\frac{w'\mu_j}{\sqrt{w'\Sigma_j w}})}{1 - \Phi(\frac{w'\mu_j}{\sqrt{w'\Sigma_j w}})} \tag{3.19}$$

$$M_{d,j}^2 = \Sigma_j + \frac{\Sigma_j w w' \Sigma_j}{w' \Sigma_j w} \frac{w' \mu_j}{\sqrt{w' \Sigma_j w}} \frac{\phi(\frac{w' \mu_j}{\sqrt{w' \Sigma_j w}})}{1 - \Phi(\frac{w' \mu_j}{\sqrt{w' \Sigma_j w}})} \tag{3.20}$$

It is interesting to notice that (3.17) and (3.18) are the first two moments of the pseudo complete sample weighted by z. For example, the numerator in (3.17) not only includes the observed valid draw (Y_t) but also imputes the sum of the latent invalid draws at time t with its conditional expectation that is feasible at the current iteration: $\tilde{n}_{t,j}(M_{d,j}^{1,l} + \mu_j^l)$. Similar pattern can also be observed in the denominator with $1+\tilde{n}_{t,j}$ being the total number of draws at time t. Moreover, our EM algorithm includes the standard EM algorithm for normal mixture models (Appendix) as a special case. To see this, suppose no truncation is imposed, we have $F_j = 1$, and $\tilde{n}_{t,j} = 0$. Therefore, the E step and M step become the same as in Appendix.

Finally, repeat E step and M step until convergence. Clearly, the new EM algorithm provides a closed-form solution and is able to maintain the monotonicity property. Furthermore, the constraints on parameters are satisfied by construction, e.g., Σ^{l+1} is positive semi-definite, $\sum_{j=1}^{P} \alpha_j^{l+1} = 1$, and $\alpha_j^{l+1} > 0$.

3.3.2 A new EM algorithm for truncated normal mixture model (conditional case)

We now discuss the conditional case. The EM algorithm is applied to the likelihood (3.7) to estimate $\Psi = \{\alpha_j, A_j, \Sigma_j | \forall j \in P\}$. Similar to section 3.1, the pseudo complete log-likelihood function can be constructed:

$$L^C(\Psi) = \frac{1}{T-Q}\sum_{t=Q+1}^{T}\sum_{j=1}^{P} z_{tj}[log\alpha_j + logf_{t,j}^N(Y_{t,n_t+1}) + \sum_{k=1}^{n_t} logf_{t,j}^N(Y_{t,k})] \quad (3.21)$$

E Step: the conditional expectation of complete log-likelihood function can be written as

$$\begin{aligned} & Q(\Psi|\Psi^l) \\ =& E[L^C(\Psi)|Y,\Psi^l] \\ =& \frac{1}{T-Q}\sum_{t=Q+1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj}[log\alpha_j + logf_{t,j}^N(Y_{t,n_t+1}) + \tilde{n}_{t,j}(\int logf_{t,j}^N(Y_{t,k})(\frac{f_{t,j}^{N,l}(Y_{t,k})}{1-F_{t,j}^l})dY_{t,k})] \end{aligned} \quad (3.22)$$

where $\tilde{z}_{tj} = \frac{\alpha_j^l f_{t,j}^l(Y_t)}{\sum_{r=1}^P \alpha_r^l f_{t,r}^l(Y_t)}$. $\tilde{n}_{t,j} = E(n_t|z_{tj}=1,Y,\Psi^l) = \frac{1-F_{t,j}^l}{F_{t,j}^l}$. $f_{t,j}^{N,l}(.)$ and $F_{t,j}^l$ are respectively $f_j^N(.)$ and F_j conditional on Ψ^l and with μ_j replaced by $\mu_{t,j} = C_j + B_{j,1}Y_{t-1} + ... + B_{j,Q}Y_{t-Q}$.

M Step: maximizing $Q(\Psi|\Psi^l)$ gives the iterated rules for Ψ. See Appendix B.6 for details.

$$\alpha_j^{l+1} = \frac{\sum_{t=Q+1}^{T} \tilde{z}_{tj}}{T-P} \tag{3.23}$$

$$A_j^{l+1} = (\bar{X}_j'\bar{Y}_j + \tilde{X}_j'\tilde{M}^1_{d',\bar{T},j})'(\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)^{-1} \tag{3.24}$$

$$\Sigma_j^{l+1} = \frac{\sum_{t=Q+1}^{T} \tilde{z}_{tj}[(Y_t - A_j^{l+1}X_{t-1})(Y_t - A_j^{l+1}X_{t-1})' + \tilde{n}_{t,j}M^2_{d',t,j}]}{\sum_{t=P+1}^{T} \tilde{z}_{tj}(1+\tilde{n}_{t,j})} \tag{3.25}$$

where $\tilde{M}^1_{d',\bar{T},j} = (\tilde{M}^1_{d',Q+1,j}, ..., \tilde{M}^1_{d',T,j})'$, and $\tilde{M}^1_{d',t,j} = \sqrt{\tilde{z}_{tj}\tilde{n}_{t,j}}(M^1_{d,t,j} + \mu^l_{t,j})$. $M^1_{d,t,j}$ is $M^1_{d,j}$ with μ_j^l replaced by $\mu^l_{t,j} = C_j^l + B^l_{j,1}Y_{t-1} + ... + B^l_{j,Q}Y_{t-Q}$, and $M^2_{d',t,j}$ is $M^2_{d',j}$ with μ_j^l and μ_j^{l+1} replaced by $\mu^l_{t,j}$ and $\mu^{l+1}_{t,j}$ respectively. Furthermore, $\bar{X}_j = \sqrt{\tilde{z}_j\tau_1^{1+2Q}} \odot X$, and $X = (\tau^1_{T-Q}, (Y_Q^{T-1})', ..., (Y_1^{T-Q})')$, where τ_a^b is a vector of ones with dimension $a \times b$. $\tilde{X}_j = \sqrt{(\tilde{z}_j \odot \tilde{n}_j)\tau_1^{1+2Q}} \odot X$, $\tilde{z}_j = (\tilde{z}_{Q+1,j}, ..., \tilde{z}_{T,j})'$, $\tilde{n}_j = (\tilde{n}_{Q+1,j}, ...\tilde{n}_{T,j})'$, $\bar{Y}_j = \sqrt{\tilde{z}_j\tau_1^2} \odot (Y_{Q+1}^T)'$, and $X'_{t-1} = (1, Y'_{t-1}, ..., Y'_{t-Q})$. The operator $\odot$ represents Hadamard product.

The iterated rules for α_j and Σ_j remain similar to these in Section 3.1 with only minor changes. Note that A_j has an iterated rule that resembles the format of the maximum likelihood estimates for a vector autoregressive model (VAR). When truncation is not in presence, it becomes $A_j^{l+1} = (\bar{X}_j'\bar{Y}_j)'(\bar{X}_j'\bar{X}_j)^{-1}$. Therefore, (3.24) can be viewed as applying VAR to the pseudo complete sample.

3.4 Asymptotic theory

In this section, we discuss the asymptotic properties of the ML estimator. The following theorem shows that under some regular conditions, the MLE is consistent. We

begin by imposing the following assumptions:

Assumption 1. $\{Y_t\}$ are generated from (3.1), and are strictly stationary and ergodic.

Assumption 2. Ψ_0 is the true parameter set, and Ψ_0 is an interior point of Ξ, where Ξ is a compact subset of $\{\Psi \in (0,1)^{P-1} \times \mathbb{R}^{(5+4Q)P} : \Sigma_j \text{ are positive definite } \forall j\}$.

Assumption 3. $E(\|Y_t\|^2) < \infty$, where $\|.\|$ is the Euclidean norm.

These assumptions are fairly regular in the literature. It may be challenging to verify Assumption 1 as the model is nonlinear. The necessary and sufficient conditions for stationarity and ergodicity that are imposed on parameters remain for future research. Notice that for the Gaussian MTD and MAR models, the sufficient and necessary conditions for first-order and second-order stationarity have been derived (see e.g., Le et. al. 1996, Wong and Li 2000). Assumption 2 and Assumption 3 are sufficient to ensure the uniform convergence of the likelihood function.

The following theorem establishes the strong consistency of ML estimator and the proof can be found in Appendix B.7.

Theorem 1. *Under Assumption 1,2 and 3, the maximum likelihood estimator* $\hat{\Psi} = \underset{\Psi \in \Xi}{argmax} L(\Psi)$ *is strongly consistent, that is* $\hat{\Psi} \to \Psi_0$ *a.s.*

3.5 Monte Carlo Simulation

In this section, we perform Monte Carlo simulation to evaluate the finite sample performance of the proposed EM algorithm on the TMT model. Experiments are designed for both unconditional and conditional cases.

3.5.1 Unconditional case experiments

We consider two cases with the number of components being $P = 2$ (DGP 1) and $P = 3$ (DGP 2). The data generating process is as follow. First, we set the parameters according to the configurations in Table 3.1 and Table 3.2. Second, we calculate η_j for all j, which represents the corresponding component weight for each component before the truncation is imposed. The relationship between α_j and η_j can be described as: $\alpha_j = \frac{\eta_j F_j}{\sum_{j=1}^{P} \eta_j F_j}$. Third, a large enough sample is drawn from the bivariate normal mixture distribution (component weight η_j). Finally, only the observations that satisfy the constraint $x_t \geq y_t$ are kept.[4]

The initial values of parameters are estimated using K-means [5], from where the EM algorithm iterates until convergence to find the MLE. [6] We consider two sample sizes ($T = 200$ and $T = 1000$). The number of Monte Carlo replications is 100.

In Table 3.1 and Table 3.2, we report the means and standard errors of the estimated parameters across replications. The biases of parameters are small in both DGPs. As the sample size increases, the estimates get closer to the true values and the standard errors become smaller. One should bear in mind that in all the DGPs, it is not necessary to impose constraints on μ (e.g., $w'\mu \geq 0$) since μ is not the mean of the truncated normal distribution.

[4] From these observations that satisfy the constraint, start collecting from the 101^{th} observation (the initial 100 observations are discarded, known as the burn-in period) until the desired sample size is reached.

[5] K-means provides bias estimates because it doesn't account for the truncation. It treats the sample as if it comes from a bivariate normal mixture distribution. Nevertheless, in our experiments, these initial values are usually good enough for the EM algorithm to converge to the true parameters.

[6] The stopping criterium is set such that either 200 iterations are reached or the increase in log-likelihood (3.8) is less than e^{-10}.

Two components	α	μ	Σ	
True	0.4	8	1	0.5
		7	0.5	1
	0.6	4	2	0.3
		3	0.3	2
EM (T=200)	0.3995	7.9666	1.0137	0.5303
	(0.0374)	(0.1653)	(0.2103)	(0.2237)
		7.0037	0.5303	1.0735
		(0.1961)	(0.2237)	(0.3297)
	0.6005	3.9657	2.0160	0.3161
	(0.0374)	(0.3028)	(0.4444)	(0.3399)
		2.9665	0.3161	1.9763
		(0.2932)	(0.3399)	(0.4059)
EM (T=1000)	0.3997	7.9961	1.0002	0.4988
	(0.0180)	(0.0730)	(0.1015)	(0.0774)
		7.0083	0.4988	1.0136
		(0.0754)	(0.0774)	(0.1138)
	0.6003	3.9994	1.9831	0.2975
	(0.0180)	(0.1122)	(0.1873)	(0.1224)
		3.0082	0.2975	1.9995
		(0.1301)	(0.1224)	(0.1994)

Note: the numbers in parentheses are standard errors.

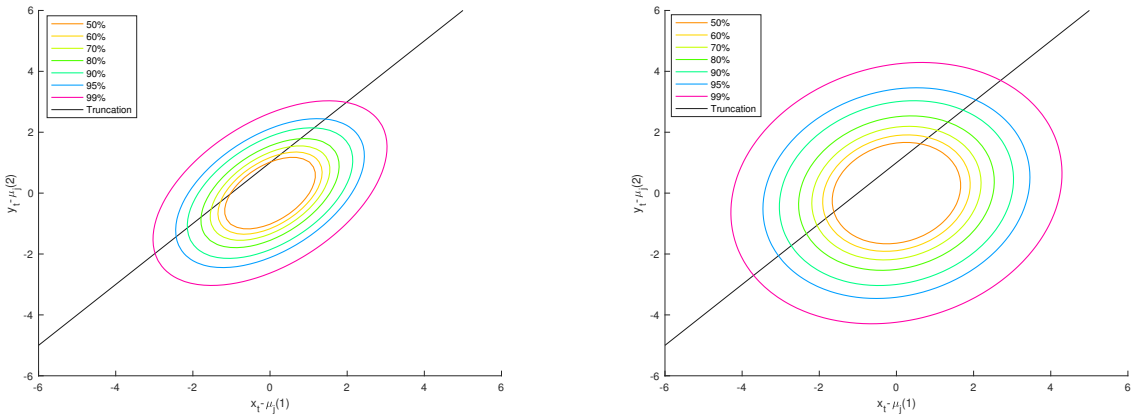
Table 3.1: Simulation results for DGP 1

Three components	α	μ	Σ	
True	0.2	10 9	1 0.5 0.5 1	
	0.3	2 2	3 1 1 3	
	0.5	−4 −6	5 2 2 5	
EM (T=200)	0.1971 (0.0280)	9.7305 (1.4018) 8.7987 (1.2179)	1.0930 (0.4596) 0.5093 (0.2354)	0.5093 (0.2354) 1.0291 (0.4231)
	0.2966 (0.0350)	2.2173 (1.5182) 2.1865 (1.3117)	2.9945 (0.9741) 1.0580 (0.5984)	1.0580 (0.5984) 2.9086 (1.1476)
	0.5063 (0.0349)	−4.1199 (0.4724) −5.9924 (0.3916)	5.2792 (1.3622) 2.0684 (0.7550)	2.0684 (0.7550) 4.8277 (0.9366)
EM (T=1000)	0.1991 (0.0144)	10.0053 (0.1045) 9.0070 (0.1022)	0.9981 (0.1107) 0.5074 (0.0882)	0.5074 (0.0882) 1.0219 (0.1183)
	0.3010 (0.0159)	2.0231 (0.2787) 2.0295 (0.2621)	2.9561 (0.4035) 0.9945 (0.2765)	0.9945 (0.2765) 3.0806 (0.5105)
	0.4999 (0.0176)	−3.9821 (0.1661) −5.9978 (0.1804)	4.9736 (0.4139) 2.0095 (0.3685)	2.0095 (0.3685) 4.9974 (0.5284)

Note: the numbers in parentheses are standard errors.

Table 3.2: Simulation results for DGP 2

To visualize the truncations on the mixture distribution, we plot in Figure 3.1 the truncations for two component in DGP 1. For a better comparison, each component is re-centered at the origin (shifted by μ_j) together with the truncation lines.

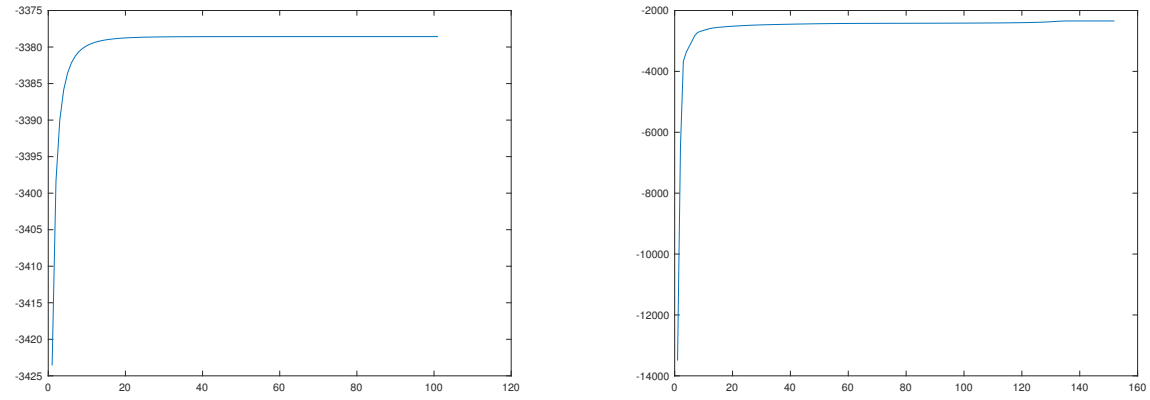


(a) Component 1 (b) Component 2

Note: $\mu_j(i)$ is the i^{th} element of μ_j.

Figure 3.1: Truncation of component density

In Figure 3.2(a), we plot the likelihood (3.8) for a one-time implementation of the EM algorithm in DGP 1. It provides the evidence that monotonicity in likelihood holds for the new EM algorithm. Moreover, the speed of convergence is fast with convergence achieved in about 20 iterations.



(a) Unconditional experiment (b) Conditional experiment

Figure 3.2: Log-likelihood

3.5.2 Conditional case experiments

Unlike the unconditional case where each observation is temporally independent, the conditional case carries time dependence in each observation. Hence, the data generating process is slightly different. First, we set the parameters as in Table 3.3. Second, at time t, we calculate $\eta_{t,j}$ for each j, where the subscript t comes from $F_{t,j}$ as μ_j is now replaced with $\mu_{t,j}$. Notice that α_j is fixed while $\eta_{t,j}$ changes with time. Third, independent random draws (e.g. 1000 draws) are made from the bivariate normal mixture distribution (component weight $\eta_{t,j}$). Fourth, we keep the draws that satisfy the constraint $x_t \geq y_t$, from which one is selected randomly as the observation at time t. Repeat the above steps until a sample with desired sample size is generated.[7]

In Table 3.3, we design three cases (DGP 3 and DGP 4 are $TMT(2,1)$, and DGP 5 is $TMT(3,1)$). Specifically, DGP 3 considers two components with the constraint ($x_t \geq y_t$) binding for one but not the other. [8] DGP 4 focuses on the case where the constraint is binding for both components. DGP 5 is a combination with the restriction not binding, binding with low persistency, and binding with high persistency. To visualize the constraint, we plot in Figure 3.3 the truncations for DGP 5. As the truncation is time varying, it cuts the density at different locations after re-centering (shifted by $\mu_{t,j}$ for each t and each j). The variation in truncations is smaller for the low persistency component because the location of truncation is more likely to be dominated by the constant C_j.

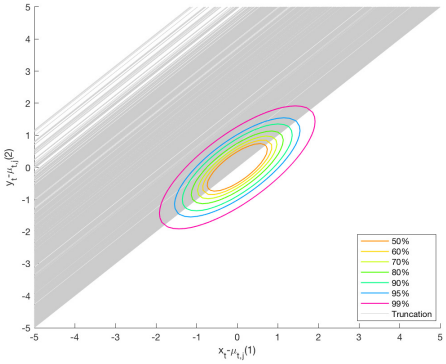
[7] Similar to section 5.1, the first 100 observations are discarded.

[8] The constraint will not be binding if $w'\mu_{t,j} = w'(C_j + B_{j,1}Y_{t-1} + ... + B_{j,Q}Y_{t-Q}) \gg 0$. In our simulation, we fix B and manipulate C to allow the restriction to be binding or not.

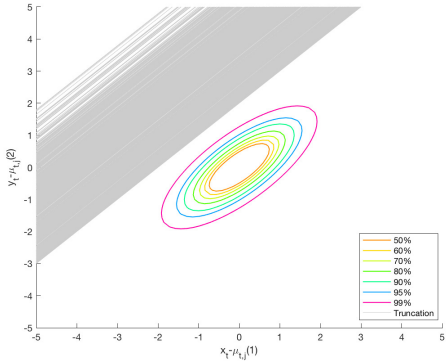
DGP		α	C	B		Σ	
3	NB	0.4	2	0.1	−0.8	0.4	0.3
			0	−0.8	0.1	0.3	0.4
	B	0.6	−2	0.7	−0.1	0.4	0.3
			−2	−0.1	0.7	0.3	0.4
4	B	0.4	0	0.1	−0.8	0.4	0.3
			0	−0.8	0.1	0.3	0.4
	B	0.6	2	0.2	−0.1	0.4	0.3
			2	−0.1	0.2	0.3	0.4
5	B	0.5	2	0.1	−0.8	0.4	0.3
			2	−0.8	0.1	0.3	0.4
	NB	0.3	2	0.3	−0.4	0.4	0.3
			0	−0.4	0.3	0.3	0.4
	B	0.2	−2	0.2	−0.1	0.4	0.3
			−2	−0.1	0.2	0.3	0.4

Note: B and NB denote binding and not binding respectively.

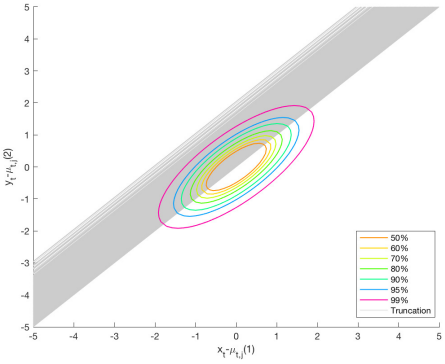
Table 3.3: Data Generating Process (DGP 3 - DGP 5)



(a) Binding (high persistency)



(b) Not Binding



(c) Binding (low persistency)

Figure 3.3: Truncations of DGP 5

We initialize the EM algorithm by randomly choosing 50 points from the parameter space.[9] Each point runs EM algorithm separately. The one that achieves the highest likelihood is chosen. We consider two sample sizes ($T = 200$ and $T = 1000$). The number of Monte Carlo replications is 100.

We summarize the average results across replications from Table 3.4 to Table 3.6. Standard errors are calculated over replications. In all cases, the EM algorithm performs satisfactory in both small and large sample experiments. The standard error shrinks towards zero as the sample size increases. Last but not least, we can see in Figure 3.2(b) that the monotonicity of EM algorithm is preserved for the likelihood (3.7).

DGP 3	α	C	B		Σ	
True	0.4	2	0.1	−0.8	0.4	0.3
		0	−0.8	0.1	0.3	0.4
	0.6	−2	0.7	−0.1	0.4	0.3
		−2	−0.1	0.7	0.3	0.4
EM (T=200)	0.3964	1.9385	0.1054	−0.7978	0.4177	0.3006
	(0.0319)	(0.7890)	(0.0801)	(0.0383)	(0.2974)	(0.0986)
		0.0510	−0.7941	0.1185	0.3006	0.4096
		(0.4026)	(0.0632)	(0.1738)	(0.0986)	(0.1867)
	0.6036	−1.9644	0.6939	−0.1061	0.3957	0.2997
	(0.0319)	(0.4446)	(0.0644)	(0.0766)	(0.0560)	(0.0476)
		−2.0041	−0.1023	0.6891	0.2997	0.4015
		(0.3235)	(0.0730)	(0.0595)	(0.0476)	(0.0661)
EM (T=1000)	0.3989	2.0073	0.0983	−0.8009	0.3937	0.2931
	(0.0152)	(0.0734)	(0.0127)	(0.0163)	(0.0253)	(0.0230)
		0.0038	−0.8016	0.0989	0.2931	0.3916
		(0.0785)	(0.0133)	(0.0170)	(0.0230)	(0.0280)
	0.6011	−2.0037	0.6995	−0.1023	0.4011	0.3006
	(0.0152)	(0.0625)	(0.0099)	(0.0141)	(0.0234)	(0.0212)
		−2.0038	−0.1012	0.6985	0.3006	0.3987
		(0.0615)	(0.0102)	(0.0144)	(0.0212)	(0.0261)

Note: the numbers in parentheses are standard errors.

Table 3.4: Simulation results for DGP 3

[9]Elements of α is uniformly selected from $(0, 1)$ and sum up to one. Elements of B are uniformly selected from $(-1, 1)$. Elements of C and off-diagonal elements of L are uniformly selected from $(-3, 3)$, where L is the Cholesky decomposition lower triangle matrix of $\Sigma = LL'$. Diagonal elements of L are uniformly selected from $(0, 3)$. For DGP 9, 200 initial points are chosen to account for a higher dimensional parameter space.

DGP 4	α	C	B		Σ	
True	0.4	0	0.1	−0.8	0.4	0.3
		0	−0.8	0.1	0.3	0.4
	0.6	2	0.2	−0.1	0.4	0.3
		2	−0.1	0.2	0.3	0.4
EM (T=200)	0.3988 (0.0415)	−0.0130	0.1034	−0.8003	0.3748	0.2805
		(0.2122)	(0.2689)	(0.2610)	(0.0747)	(0.0718)
		0.0219	−0.7720	0.0607	0.2805	0.3878
		(0.2326)	(0.2643)	(0.2704)	(0.0718)	(0.0945)
	0.6012 (0.0415)	1.9462	0.2349	−0.1332	0.4023	0.2879
		(0.2226)	(0.1950)	(0.1854)	(0.0723)	(0.0568)
		2.0131	−0.0790	0.1753	0.2879	0.3944
		(0.2178)	(0.2044)	(0.1960)	(0.0568)	(0.0728)
EM (T=1000)	0.4010 (0.0177)	−0.0088	0.0967	−0.7971	0.3978	0.2940
		(0.1269)	(0.1268)	(0.1127)	(0.0386)	(0.0322)
		0.0208	−0.8237	0.1233	0.2940	0.3983
		(0.1076)	(0.1111)	(0.1003)	(0.0322)	(0.0462)
	0.5990 (0.0177)	1.9644	0.2187	−0.1178	0.4085	0.3002
		(0.1349)	(0.0863)	(0.0838)	(0.0353)	(0.0306)
		2.0605	−0.1280	0.2271	0.3002	0.4203
		(0.1935)	(0.1189)	(0.1173)	(0.0306)	(0.0523)

Note: the numbers in parentheses are standard errors.

Table 3.5: Simulation results for DGP 4

DGP 5	α	C	B		Σ	
True	0.5	2	0.1	−0.8	0.4	0.3
		2	−0.8	0.1	0.3	0.4
	0.3	2	0.3	−0.4	0.4	0.3
		0	−0.4	0.3	0.3	0.4
	0.2	−2	0.2	−0.1	0.4	0.3
		−2	−0.1	0.2	0.3	0.4
EM (T=200)	0.5078 (0.0427)	1.9985 (0.1535)	0.0978 (0.0551)	−0.8027 (0.0687)	0.3910 (0.0665)	0.2908 (0.0560)
		1.9867 (0.1433)	−0.7980 (0.0515)	0.0972 (0.0642)	0.2908 (0.0560)	0.3911 (0.0594)
	0.2932 (0.0378)	2.0114 (0.1505)	0.2991 (0.0640)	−0.3892 (0.0771)	0.3665 (0.0887)	0.2736 (0.0746)
		0.0055 (0.1390)	−0.4047 (0.0610)	0.3111 (0.0753)	0.2736 (0.0746)	0.3664 (0.0792)
	0.1990 (0.0294)	−2.0138 (0.2691)	0.2002 (0.0892)	−0.1047 (0.1025)	0.3873 (0.0944)	0.2917 (0.0820)
		−1.8540 (0.4928)	−0.1382 (0.1145)	0.2350 (0.1285)	0.2917 (0.0820)	0.4110 (0.1459)
EM (T=1000)	0.5000 (0.0178)	2.0026 (0.0583)	0.0990 (0.0223)	−0.8016 (0.0247)	0.3966 (0.0272)	0.2995 (0.0228)
		1.9920 (0.0574)	−0.7969 (0.0220)	0.0953 (0.0249)	0.2995 (0.0228)	0.3991 (0.0253)
	0.3001 (0.0167)	1.9968 (0.0710)	0.3008 (0.0273)	−0.3987 (0.0304)	0.3907 (0.0334)	0.2963 (0.0274)
		−0.0054 (0.0707)	−0.3983 (0.0263)	0.2990 (0.0308)	0.2963 (0.0274)	0.3986 (0.0334)
	0.1999 (0.0115)	−1.9983 (0.0954)	0.2003 (0.0298)	−0.0987 (0.0437)	0.3886 (0.0483)	0.2914 (0.0382)
		−2.0139 (0.1007)	−0.0960 (0.0310)	0.1996 (0.0427)	0.2914 (0.0382)	0.3906 (0.0493)

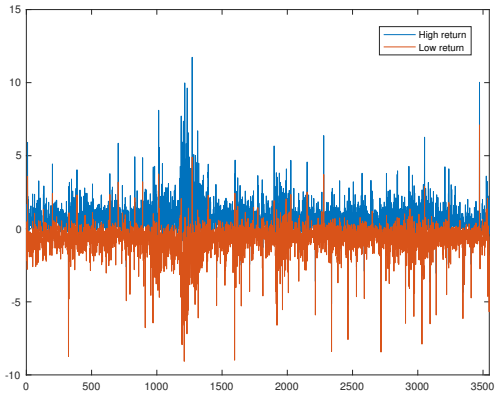
Note: the numbers in parentheses are standard errors.

Table 3.6: Simulation results for DGP 5

3.6 Empirical Application

We apply TMT to model the interval-valued IBM daily stock returns. The high/low return is calculated as the percentage change of the highest/lowest daily price with respect to the closing price of the previous day. For example, the high return at time t is: $r_{high,t} = 100(P_{high,t} - P_{close,t-1})/P_{close,t-1}$. The data is constructed as an interval-valued time series with $r_{high,t} \geq r_{low,t}$. To visualize the data, we plot a sample from 2004/1/1 to 2018/4/1 (3584 observations) in Figure 3.4. We can see that the volatility for the high and

low returns is high in some periods while remaining quiet in others, suggesting potentially the presence of multiple regimes in the variance of the system.



(a) Real data

Figure 3.4: Daily IBM High/Low Stock Returns (2004/1/1 to 2018/4/1)

We consider TMT model with up to seven components and four lags. That is, $P = \{2, ..., 7\}$, and $Q = \{1, 2, 3, 4\}$, with total 28 specifications. [10] The best fitted model selected by BIC is $TMT(4, 2)$. The estimation results are reported in Table 3.7.[11] It is interesting to see that the fourth component has high volatility (big Σ) while only happens with a small probability (small α). Figure 3.5 shows the truncations for each component across time after re-centering (shifted by $\mu_{t,j}$ for each t and each j). The truncations vary by component: the first and second components have truncations almost not binding while for the last two components the truncations are binding.

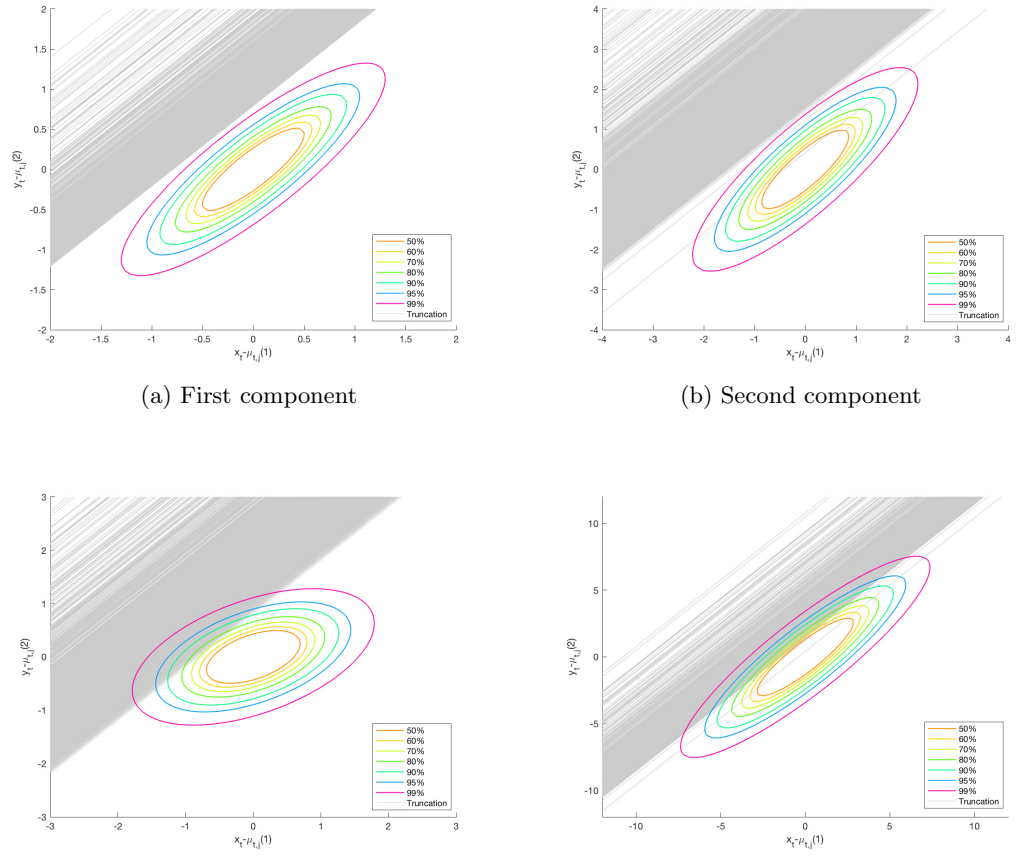
[10]The case when only one component is involved ($TMT(1, Q)$) turns out to be the same as GL, which, for a better comparison, will be discussed in the following separately.

[11]Standard errors are calculated using block bootstrap (Politis and White 2004)

Component	α	C	B_1		B_2		Σ	
1	0.4184	0.3916	0.0681	−0.1033	−0.0276	0.0327	0.1838	0.1600
	(0.0428)	(0.0535)	(0.0331)	(0.0412)	(0.0368)	(0.0370)	(0.0230)	(0.0195)
		−0.2864	−0.0683	0.0801	−0.1285	0.1480	0.1600	0.1909
		(0.0688)	(0.0411)	(0.0501)	(0.0402)	(0.0397)	(0.0195)	(0.0204)
2	0.3635	0.3678	0.1758	−0.1563	0.0152	−0.0857	0.5367	0.5165
	(0.0450)	(0.0859)	(0.0781)	(0.0829)	(0.0442)	(0.0587)	(0.0883)	(0.0832)
		−0.4786	−0.0843	0.1641	−0.2135	0.1674	0.5165	0.7006
		(0.0886)	(0.0819)	(0.1001)	(0.0531)	(0.0856)	(0.0832)	(0.0840)
3	0.1323	0.4125	0.6549	−0.5425	0.1157	−0.2460	0.3476	0.1228
	(0.0508)	(0.1946)	(0.1715)	(0.1354)	(0.1214)	(0.0968)	(0.0819)	(0.0606)
		−0.1677	−0.1510	0.1473	0.1101	−0.2316	0.1228	0.1778
		(0.1054)	(0.0973)	(0.0821)	(0.0632)	(0.0693)	(0.0606)	(0.0617)
4	0.0857	0.1484	0.1015	−0.1265	0.5265	−0.3146	5.9263	5.4043
	(0.0189)	(0.3580)	(0.1948)	(0.1856)	(0.2271)	(0.1736)	(0.8068)	(0.7199)
		−0.9836	−0.1358	0.3614	−0.0414	0.2858	5.4043	6.2028
		(0.4077)	(0.1980)	(0.1778)	(0.2525)	(0.1805)	(0.7199)	(0.8251)

Note: the numbers in parentheses are standard errors.

Table 3.7: Estimation results of $TMT(4, 2)$



(a) First component

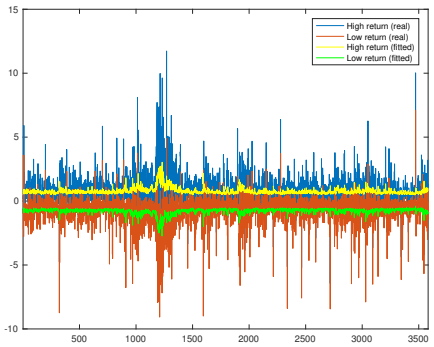
(b) Second component

(c) Third component

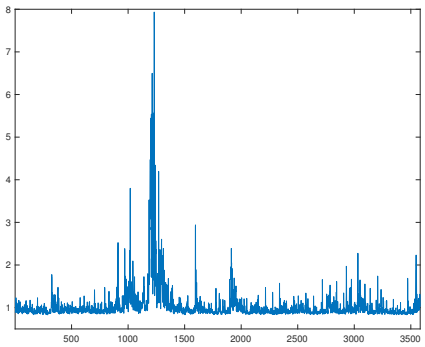
(d) Fourth component

Figure 3.5: Truncations for the fitted $TMT(4, 2)$

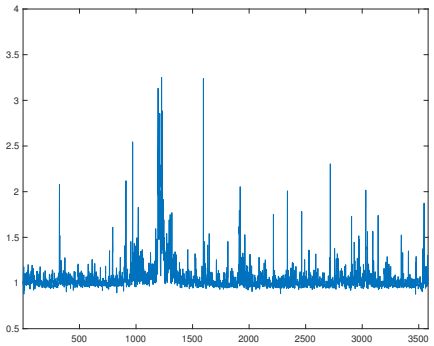
We plot in Figure 3.6 the fitted conditional means (3.3) together with the realized data. The persistency in the data seems to be well described. Figure 3.6 also shows the fitted conditional variances and correlation coefficients (3.5) of the high/low returns. The spikes in the fitted variances is aligned with the volatility clustering in the data. The contemporaneous conditional correlations stay at a relatively high level most of the time while drop toward zero during the volatile periods. It aligns with the observation that the ranges (gaps between upper and lower bounds) tend to be larger in these periods. In Figure 3.7, we plot some fitted conditional densities to illustrate the flexibility of the truncated normal mixture distribution.



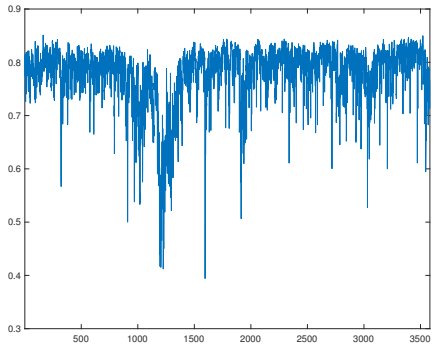
(a) Fitted Conditional Mean



(b) Fitted Variance (High Return)

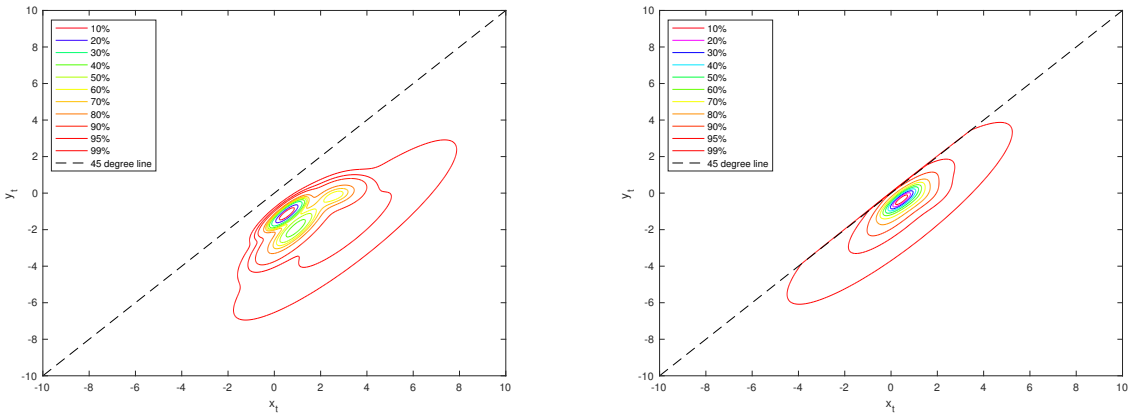


(c) Fitted Variance (Low Return)



(d) Fitted Correlation

Figure 3.6: Fitted Conditional Mean, Variance and Correlation of Daily IBM High/Low Stock Returns (2004/1/1 to 2018/4/1)



(a) 12/18/2008 (b) 12/29/2017

Figure 3.7: Fitted Conditional Density Contours

Given that not all the parameters in Table 3.7 are significant, and to account for the increase in parameters when the number of components grows larger, we also consider a restricted version of the model, $RTMT(P)$, where the restriction is imposed such that each component contains only one lag from the information set. For instance, $\mu_{t,j} = C_j + B_{j,j}Y_{t-j}$ and $B_{j,r} = 0$ for $r \neq j$. We consider up to seven components ($P = \{1, ..., 7\}$) for $RTMT$. Finally, we compare the TMT and $RTMT$ models with four other models. The number of lags for these models is selected using BIC. The linear vector autoregressive model serves as a benchmark. Two multivariate $GARCH$ models are considered to account for conditional heteroskedasticity in the data. See Bauwens et. al. (2006) for a review of multivariate $GARCH$ models. We also implement GL. Notice that, however, VAR and $VAR - MGARCH$ models cannot preserve the natural order of the ITS. A detailed comparison of the six models is summarized in Table 3.8.

Model for the mean	Model for the variance	Log-likelihood	Number of parameters	BIC
$VAR(7)$	_	-8604	30	-17,454
$VAR(7)$	MGARCH(1,1)-SBEKK	-8175	35	-16,064
$VAR(7)$	MGARCH(1,1)-DCC	-8155	39	-15,991
$GL(7)$	_	-8486	33	-17,243
$RTMT(5)$	_	-6975	49	-14,352
$TMT(4,2)$	_	-6833	55	-14,117

Table 3.8: Evaluation of models

TMT achieves the highest BIC and likelihood while using the most parameters. $RTMT$ trades the likelihood and BIC for a smaller number of parameters. VAR uses the smallest number of parameters and ends up having the smallest likelihood and BIC. After accounting for time-varying conditional variance, the $VAR-MGARCH$ models improve the performance over VAR significantly, implying that the data is conditional heteroskedastic. In terms of all criteria, GL lies in between VAR and $VAR-MGARCH$. This suggests that although GL preserves the natural oder of the interval data, it has a limited ability accommodating conditional heteroskedasticity.

3.7 Conclusions

We propose a truncated mixture transition model for the interval-valued time series. The natural order of the data (upper bound greater than lower bound) is guaranteed in our model using truncated normal distributions. The model enjoys great flexibility in terms of both parameter and density specifications. However, the standard EM algorithm to estimate mixture models fails since no closed-form solutions can be obtained in M step. Therefore, a new EM algorithm is proposed, which brings the pseudo data generating process to a higher level and encloses a closed-form solution in M step. We prove the consistency of the maximum likelihood estimator. Simulation results show that the new EM algorithm

performs well. Last but not least, we illustrate the performance of the model with an application to the IBM daily high/low stock returns and it ourperforms other competing models.

Bibliography

[1] Amemiya, T. (1973), "Regression analysis when the dependent variable is truncated normal". *Econometrica*, Vol. 41, pp. 997–1016.

[2] Barndorff-Nielsen, O. (1965), "Identifiability of mixtures of exponential families," *Journal of Mathematical Analysis and Applications*, Vol. 12, pp. 115-121.

[3] Bauwens, et. al. (2006), "Multivariate GARCH models: a survey", *Journal of Applied Econometrics*, Vol. 21, pp. 79-109.

[4] Berchtold, André and Raftery, Adrian (2002), "The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series", *Statist. Sci.*, Vol. 17, pp. 328-356.

[5] Chang, S.H., et. al. (2011), "Chernoff-Type Bounds for the Gaussian Error Function", *IEEE Trans. Commun.*, Vol. 59, pp. 2939–2944.

[6] Dempster, A., et. al. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, pp. 1-38.

[7] González-Rivera, Gloria and Lin, Wei (2013), "Constrained Regression for Interval-valued Data," *Journal of Business and Economic Statistics*, Vol. 31, pp. 473–490.

[8] Hamilton, James D (1990), "Analysis of time series subject to changes in regime". *Journal of Econometrics*, Vol. 45, pp. 39-70.

[9] Hassan, M.Y. and Lii, Keh-Shin (2006), "Modeling marked point processes via bivariate mixture transition distribution models", *Journal of the American Statistical Association*, Vol. 101, pp. 1241-1252.

[10] Kalliovirta, Leena, et. al. (2016), "Gaussian mixture vector autoregression", *Journal of Econometrics,*Vol. 192, pp. 485-498.

[11] Krengel, U. (1985), *Ergodic Theorems*. de Gruyter, Berlin.

[12] Nhu D. Le, et. al. (1996), "Modeling flat stretches, bursts and outliers in time series using mixture transition distribution models", *Journal of the American Statistical Association*, Vol. 91, No. 436.

[13] Lee, Gyemin and Scott, Clayton (2012), "EM algorithms for multivariate Gaussian mixture models with truncated and censored data", *Computational Statistics & Data Analysis*, Vol. 56, pp. 2816-2829.

[14] Lima Neto, E., and de Carvalho, F. (2010), "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics and Data Analysis*, Vol. 54, pp. 333-347.

[15] McNeil and Frey (2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance*, Vol. 7, pp.271-300.

[16] Nath, G. Baikunth (1972), "Moments of a linearly truncated bivariate normal distribution", *Austral. J. Statist.*, Vol. 14, pp. 97-102.

[17] Potscher, B.M. and Prucha, I.R. (1991), "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part i: consistency and approximation concepts", *Econometric Reviews*, Vol. 10, pp.125-216.

[18] Rao, R. Ranga (1962), "Relations between Weak and Uniform Convergence of Measures with Applications", *Ann. Math. Statist,* Vol. 33, pp.659-680.

[19] Straumann, D. and Mikosch, T. (2006), "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach", *Ann. Statist*, Vol. 34, pp. 2449-2495.

[20] Teicher, Henry (1961), "Identifiability of Mixtures", *Ann. Math. Statist.*, Vol. 32, 244-248.

[21] Tu, Y. and Y. Wang (2016), "Center and log range models for interval-valued data with an application to forecast stock returns", *working paper*.

[22] Wong, C., & Li, W. (2000). "On a Mixture Autoregressive Model", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 62, pp. 95-115.

[23] Wu, C. F. Jeff. (1983), "On the Convergence Properties of the EM Algorithm", *Ann. Statist.*, Vol. 11, 95-103.

Appendix A

Appendix for Chapter 1

A.1 Bias correction for the forecast of range (Guerrero, 1993)

A second order Taylor expansion of the log-range, i.e., $y_{r,T+h}$, around $E_T(\exp(y_{r,T+h}))$

$$\begin{aligned} y_{r,T+h} \simeq \quad & \log(E_T(\exp(y_{r,T+h}))) + \\ & + \tfrac{1}{E_T(\exp(y_{r,T+h}))}(\exp(y_{r,T+h}) - E_T(\exp(y_{r,T+h}))) + \\ & - \tfrac{1}{2(E_T(\exp(y_{r,T+h})))^2}(\exp(y_{r,T+h}) - E_T(\exp(y_{r,T+h})))^2 \end{aligned}$$

Take conditional expectation,

$$\begin{aligned} \hat{y}_{r,T+h|T} \equiv E_T(y_{r,T+h}) \simeq \quad & \log(E_T(\exp(y_{r,T+h}))) \\ - \frac{1}{2(E_T(\exp(y_{r,T+h})))^2} & E_T(\exp(y_{r,T+h}) - E_T(\exp(y_{r,T+h})))^2 \end{aligned}$$

By approximating the conditional variance of the log-range $W_{h,22} \equiv Var_T(y_{r,T+h})$ by

$$W_{h,22} \simeq \frac{1}{(E_T(\exp(y_{r,T+h})))^2} E_T(\exp(y_{r,T+h}) - E_T(\exp(y_{r,T+h})))^2$$

we can write $\hat{y}_{r,T+h|T} \equiv E_T(y_{r,T+h}) \simeq \log(E_T(\exp(y_{r,T+h}))) - \frac{1}{2}W_{h,22}$. It follows that

$$y^*_{r,T+h} \simeq \exp(\hat{y}_{r,T+h|T}) \exp(\frac{W_{h,22}}{2})$$

If the log-range is in fact normally distributed, the above expression is exact. We take a Taylor's expansion of the range, i.e. $\exp(y_{r,T+h})$, around the conditional mean $E_T(y_{r,T+h})$, i.e.,

$$\begin{aligned}
\exp(y_{r,T+h}) = \quad & \exp(E_T(y_{r,T+h})) \quad + \exp(E_T(y_{r,T+h}))[y_{r,T+h} - E_T(y_{r,T+h})] + \\
& + \frac{1}{2}\exp(E_T(y_{r,T+h}))[y_{r,T+h} - E_T(y_{r,T+h})]^2 + \\
& + \frac{1}{3!}\exp(E_T(y_{r,T+h}))[y_{r,T+h} - E_T(y_{r,T+h})]^3 + \\
& + \frac{1}{4!}\exp(E_T(y_{r,T+h}))[y_{r,T+h} - E_T(y_{r,T+h})]^4 + ...
\end{aligned}$$

By taking conditional expectation and plugging in the conditional moments of a normal variate, we have

$$\begin{aligned}
E_T[\exp(y_{r,T+h})] &= \exp(E_T(y_{r,T+h}))[1 + \frac{W_{h,22}}{2} + \frac{1}{4!}3W^2_{h,22} + \frac{1}{6!}15W^3_{h,22} + \cdots] \\
&= \exp(E_T(y_{r,T+h})) \exp(\frac{W_{h,22}}{2})
\end{aligned}$$

which is exactly the Guerrero's bias-corrected forecast.

Appendix B

Appendix for Chapter 2

Tables and Figures: Modeling SP500 Daily Low/High Return Interval

	SP500 Low/High Returns		
	Center	Range	log-Range
Mean	-0.002	1.154	-0.060
Median	0.005	0.928	-0.075
Standard Deviation	0.642	0.826	0.631
Excess Kurtosis	4.021	8.781	-0.215
Skewness	-0.337	2.295	0.154
Minimum	-4.219	0.146	-1.925
Maximum	3.812	8.731	2.167
Correlation Coefficient (wrt center)	1	-0.122	-0.096
Q (10) (p value)	13.817 (0.18)	7103 (0)	7296 (0)
Q (15) (p value)	18.201 (0.25)	9500 (0)	9808 (0)
Q (20) (p value)	34.606 (0.02)	11569 (0)	12037 (0)

Table B.1: Descriptive Statistics for Center, Range and log-Range of Daily SP500 Low/High Return Intervals (Jan.2, 2009-Apr.20, 2018)

SP500 Low/High Return. Restricted VAR(6) for Center and log-Range System

	Center			log-Range		
	Coeff. estimate	SE (robust)	t-statistic	Coeff. estimate	SE (robust)	t-statistic
Constant	-0.0069	0.0148	-0.4662	-0.0013	0.0092	-0.1435
C(-1)				-0.1739	0.0134	-12.9768
C(-2)				-0.0900	0.0143	-6.3002
C(-3)				-0.0594	0.0136	-4.3840
C(-4)				-0.0363	0.0137	-2.6450
C(-5)				0.0130	0.0140	0.9326
C(-6)				-0.0087	0.0142	-0.6100
log-R(-1)				0.1651	0.0226	7.2994
log-R(-2)				0.2172	0.0225	9.6581
log-R(-3)				0.1599	0.0226	7.0875
log-R(-4)				0.0922	0.0227	4.0583
log-R(-5)				0.0986	0.0233	4.2373
log-R(-6)				0.1080	0.0217	4.9768
Adj. R-squared	0			0.5229		

SP500 Low/High Return. Restricted VAR(6) Residuals

	Center	log-Range
Mean	0.000	0.000
Median	0.008	-0.009
Standard Deviation	0.666	0.409
Excess Kurtosis	3.808	0.023
Skewness	-0.294	0.149
Minimum	-4.212	-1.274
Maximum	3.819	1.575
Correlation (Center/Log-range)	-0.1689	-0.1689
Q (10) (p value)	14.232 (0.16)	17.081 (0.07)
Q (15) (p value)	19.400 (0.20)	20.626 (0.15)
Q (20) (p value)	34.380 (0.02)	28.846 (0.09)

Table B.2: Estimation of VAR for Center and log-Range System (Jan. 1, 2009-Dec. 31, 2016)

SP500 Low/High Return. Tests on VAR(6) residuals

	t-statistics ($t_{k,\alpha}$)				
α	lag k				
	1	2	3	4	5
0.01	3.94	3.32	3.52	3.73	3.72
0.05	3.54	2.48	3.09	2.74	2.29
0.1	-0.08	-1.37	-1.62	-2.12	-1.57
0.2	-3.62	-4.20	-3.95	-5.23	-4.72
0.3	-5.21	-6.12	-5.91	-5.81	-6.27
0.4	-5.75	-6.21	-6.43	-6.24	-6.26
0.5	-6.52	-6.34	-6.76	-6.74	-6.56
0.6	-6.29	-6.37	-6.43	-6.35	-6.57
0.7	-4.41	-4.72	-4.64	-4.63	-4.65
0.8	-3.34	-3.41	-3.56	-3.48	-3.55
0.9	-0.51	-0.67	-0.61	-0.71	-0.65
0.95	0.99	0.63	0.78	0.79	0.64
0.99	3.64	3.48	3.33	3.01	2.70
C-statistic (C_k)	114.60	107.23	118.33	118.34	108.55

C_k aggregates all 13 autocontours for a given lag k; its 5% critical value is 22.36

Table B.3: Generalized-AutoContouR (G-ACR) tests for conditional bivariate normality (González-Rivera and Sun, 2015) for SP500 Low/High Return Intervals.

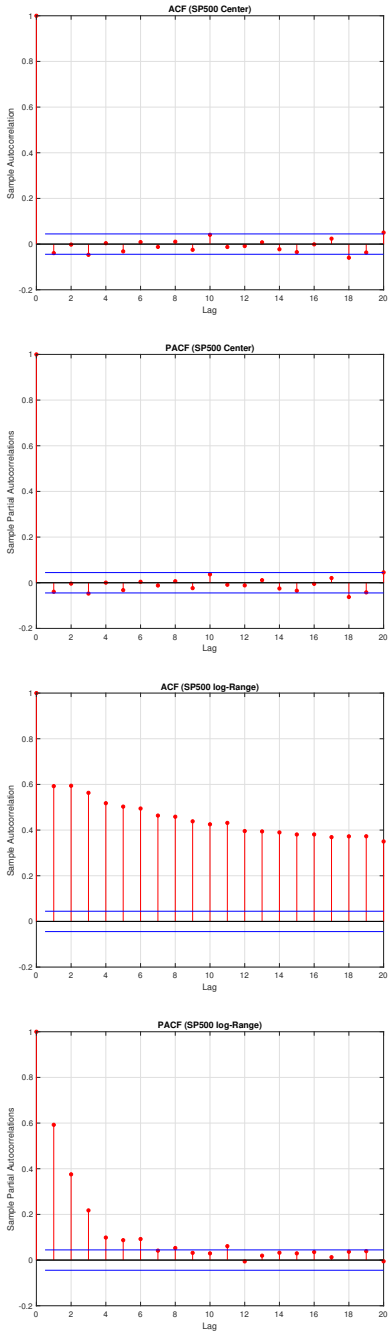


Figure B.1: SP500 low/high return interval. Autocorrelograms of center and log-range.

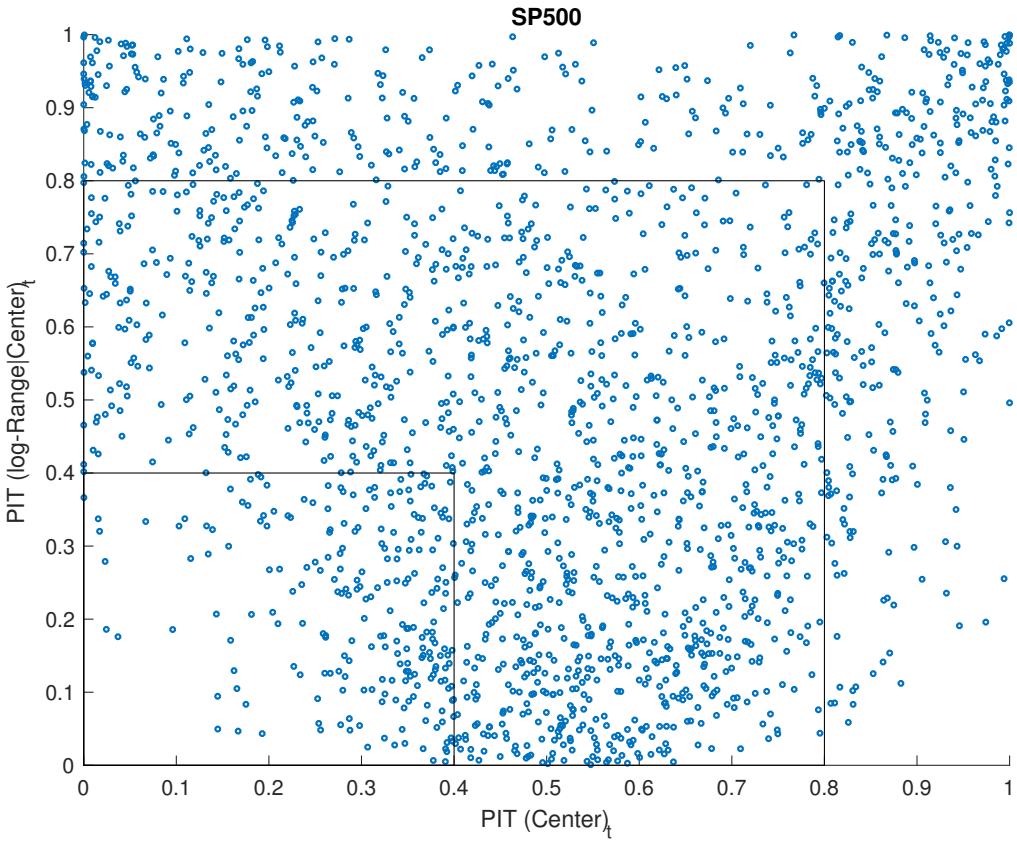


Figure B.2: SP500 low/high return interval. G-ACR specification tests for conditional bivariate normality of center and log-range. PITs are not uniformly distributed in the $[0,1]x[0,1]$ square.

Appendix C

Appendix for Chapter 3

C.1 Proof of $w'E(Y_t|\mathcal{F}^{t-1})) \geq 0$

It is sufficient to show that $w'M^1_{o,t,j} + w'\mu_{t,j} \geq 0$ for all j. Thus, it suffices to prove that

$$\frac{\phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})} \quad \geq \quad \frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}$$

Let $\lambda = \frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}$. When $\lambda \leq 0$, the above inequality obviously holds.

When $\lambda > 0$, we know that $1-\Phi(\lambda) = \frac{1}{2}erfc(\frac{\lambda}{\sqrt{2}})$, where erfc is the complementary error function defined as $erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty exp(-t^2)dt$. Also, we have $\phi(\lambda) = \frac{1}{\sqrt{2\pi}}exp(-\frac{\lambda^2}{2})$. The inequality becomes

$$\frac{1}{\sqrt{2\pi}}exp(-\frac{\lambda^2}{2}) \quad \geq \quad \frac{1}{2}erfc(\frac{\lambda}{\sqrt{2}})\lambda$$

Using the property of erfc function: $erfc(z) \leq \frac{2}{\sqrt{\pi}}\frac{exp(-z^2)}{z+\sqrt{z^2+\frac{4}{\pi}}}$, when $z > 0$, we have

$$\frac{1}{\sqrt{\pi}}\frac{exp(-\frac{\lambda^2}{2})\lambda}{\frac{\lambda}{\sqrt{2}}+\sqrt{\frac{\lambda^2}{2}+\frac{4}{\pi}}} \geq \frac{1}{2}erfc(\frac{\lambda}{\sqrt{2}})\lambda$$

Hence, we found the upper bound of $\frac{1}{2}erfc(\frac{\lambda}{\sqrt{2}})\lambda$, and it suffices to show that

$$\frac{1}{\sqrt{2\pi}}exp(-\frac{\lambda^2}{2}) \geq \frac{1}{\sqrt{\pi}}\frac{exp(-\frac{\lambda^2}{2})\lambda}{\frac{\lambda}{\sqrt{2}}+\sqrt{\frac{\lambda^2}{2}+\frac{4}{\pi}}}$$

$$\Longleftrightarrow 1 \geq \frac{1}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{2}{\pi\lambda^2}}}$$

which obviously holds when $\lambda > 0$.

C.2 The EM algorithm for normal mixture model (unconditional case)

This section reviews the EM algorithm when the component density $f_j(Y_t|\mu_j, \Sigma_j)$ in (3.8) is a bivariate normal distribution. EM algorithm transforms the problem into a missing data framework and constructs a pseudo complete data generating process. It starts by assuming that each observation comes from one of the P components, and there is a latent variable indicating which component the observation truly comes from. Let

$z_{tj} \in \{0, 1\}$ be the indicator variable such that $z_{tj} = 1$ if Y_t is generated from component j and 0 otherwise. The objective is to maximize the pseudo complete likelihood of $\{Y, z\}$. Denote $z_t = \{z_{t1}, ..., z_{tP}\}$. To construct the complete likelihood, the latent variable z_{tj} is specified to follow a multinomial distribution:

$$g(z_t|\Theta) = \prod_{j=1}^{P} \alpha_j^{z_{tj}} \tag{C.1}$$

The conditional density of Y_t on z_t is

$$h(Y_t|z_t, \Theta) = \prod_{j=1}^{P} \Big[f_j(Y_t|\mu_j, \Sigma_j) \Big]^{z_{tj}} \tag{C.2}$$

The complete density function becomes

$$\begin{aligned} l(Y_t, z_t|\Theta) &= g(z_t|\Theta) h(Y_t|z_t, \Theta) \\ &= \prod_{j=1}^{P} \Big[\alpha f_j(Y_t|\mu_j, \Sigma_j) \Big]^{z_{tj}} \end{aligned} \tag{C.3}$$

Therefore, the complete log-likelihood function for Θ can be written as

$$L^C(\Theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{P} z_{tj} log \alpha_j + \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{P} z_{tj} log f_j(Y_t|\mu_j, \Sigma_j) \tag{C.4}$$

where T is the sample size. The EM algorithm begins by initializing the parameter set, Θ^0, followed by the E and M steps.

E Step: Because z is not observed, $L^C(\Theta)$ is replaced with its conditional expectation ($Q(\Theta|\Theta^l)$) conditional on the the observed data (Y) and the parameter set from the previous iteration (Θ^l).

$$Q(\Theta|\Theta^l) = E(L^C(\Theta)|Y, \Theta^l) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj} log\alpha_j + \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj} log f_j(Y_t|\mu_j, \Sigma_j) \quad \text{(C.5)}$$

$$\begin{aligned}
\tilde{z}_{tj} &\equiv E(z_{tj}|Y_t, \Theta^l) \\
&= P(z_{tj}|Y_t, \Theta^l) \\
&= \frac{P(z_{tj}, Y_t, \Theta^l)}{P(Y_t, \Theta^l)} \\
&= \frac{\alpha_j^l f_j(Y_t|\mu_j^l, \Sigma_j^l)}{\sum_{k=1}^{P}\alpha_k^l f_k(Y_t|\mu_j^l, \Sigma_j^l)}
\end{aligned} \quad \text{(C.6)}$$

M Step: The updated parameter set is obtained by $\Theta^{l+1} = \underset{\Theta}{argmax} Q(\Theta|\Theta^l)$:

$$\alpha_j^{l+1} = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}}{T} \quad \text{(C.7)}$$

$$\mu_j^{l+1} = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}Y_t}{\sum_{t=1}^{T}\tilde{z}_{tj}} \quad \text{(C.8)}$$

$$\Sigma_j^{l+1} = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}(Y_t - \mu_j^{l+1})(Y_t - \mu_j^{l+1})'}{\sum_{t=1}^{T}\tilde{z}_{tj}} \quad \text{(C.9)}$$

Iterate E step and M step until convergence. Dempster et. al. (1977) pointed out that the likelihood (3.8) is closely related to the feasible pseudo complete likelihood (C.5): $L(\Theta^l) = Q(\Theta^l|\Theta^l) \leq Q(\Theta^{l+1}|\Theta^l) \leq L(\Theta^{l+1})$. Therefore, as $Q(\Theta|\Theta^l)$ is maximized in each iteration (which implies $Q(\Theta^l|\Theta^l) \leq Q(\Theta^{l+1}|\Theta^l)$), the likelihood (3.8) increases monotonically ($L(\Theta^{l+1}) \geq L(\Theta^l)$).

C.3 The EM algorithm for truncated normal mixture model (unconditional case)

Lee and Scott (2010) apply the EM algorithm to the multivariate truncated normal mixture model with each component truncated by a rectangle, e.g., $s \leq Y \leq k$, where s and k are vectors with the same dimension as Y. We adapt their arguments to derive the EM algorithm as below:

E Step: Following the same steps as Appendix A.1, the expression for $\tilde{z}_{tj}$ is the same as (C.6). However, $f_j(Y_t|\mu_j^l, \Sigma_j^l)$ is now a truncated bivariate normal distribution.

M Step:

$$\alpha_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj}}{T} \tag{C.10}$$

$$\mu_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj} Y_t}{\sum_{t=1}^T \tilde{z}_{tj}} - v_j(\mu_j^{l+1}, \Sigma_j^{l+1}) \tag{C.11}$$

$$\Sigma_j^{l+1} = \frac{\sum_{t=1}^T \tilde{z}_{tj}(Y_t - \mu_j^{l+1})(Y_t - \mu_j^{l+1})'}{\sum_{t=1}^T \tilde{z}_{tj}} + I_j(\mu_j^{l+1}, \Sigma_j^{l+1}) \tag{C.12}$$

where $v_j(\mu_j^{l+1}, \Sigma_j^{l+1})$ and $I_j(\mu_j^{l+1}, \Sigma_j^{l+1})$ are nonlinear functions of μ_j^{l+1} and Σ_j^{l+1}. Details are discussed in appendix A.3.1.

C.3.1 Derivation of the EM algorithm

Let Y follows a truncated bivariate normal distribution:

$$f(Y) = \frac{1}{2\pi|\Sigma|[1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})]} exp[-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)] \tag{C.13}$$

Denote $Y^o = Y - \mu$, and its first and second moments are given as (Nath 1972):

$$M_o^1 = \frac{\Sigma w}{\sqrt{w'\Sigma w}} \frac{\phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})}{1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})}$$

$$M_o^2 = \Sigma + \frac{\Sigma w w'\Sigma}{w'\Sigma w} \frac{-w'\mu}{\sqrt{w'\Sigma w}} \frac{\phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})}{1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})}$$

In E step, we can write down the conditional expectation of the complete log-likelihood function:

$$Q(\Theta|\Theta^l) = E(L^C(\Theta)|Y,\Theta^l) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj}\Big[log\alpha_j - log2\pi - \frac{1}{2}log|\Sigma_j| - \frac{1}{2}(Y_t-\mu_j)'\Sigma_j^{-1}(Y_t-\mu_j) - log(1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}))\Big]$$

where $1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}) = \frac{1}{\sqrt{\pi}}\int_{\frac{-w'\mu_j}{\sqrt{2w'\Sigma_j w}}}^{\infty} exp(-t^2)dt.$

First, we take derivative of $log(1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}))$ with respect to μ_j

$$\begin{aligned}
\frac{\partial}{\partial \mu_j}\left[log(1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}))\right] &= \frac{1}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}\left\{\frac{1}{\sqrt{\pi}}\frac{w}{\sqrt{2}\sqrt{w'\Sigma_j w}}exp(-(\frac{-w'\mu_j}{\sqrt{2}\sqrt{w'\Sigma_j w}})^2)\right\} \\
&= \frac{\phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}\frac{w}{\sqrt{w'\Sigma_j w}} \\
&= \frac{ww'M^1_{o,j}}{w'\Sigma_j w}
\end{aligned}$$

where $M^1_{o,j}$ is M^1_o with $\mu = \mu_j$ and $\Sigma = \Sigma_j$.

Next, take the derivative of $Q(\Theta|\Theta^l)$ with respect to μ_j

$$\frac{\partial}{\partial \mu_j}[Q(\Theta|\Theta^l)] = \frac{1}{T}\sum_{t=1}^{T}\tilde{z}_{tj}\left[\Sigma_j^{-1}Y_t - \Sigma_j^{-1}\mu_j - \frac{ww'M^1_{o,j}}{w'\Sigma_j w}\right] = 0$$

We can get:

$$\mu_j = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}Y_t}{\sum_{t=1}^{T}\tilde{z}_{tj}} \text{-} v_j(\mu_j,\Sigma_j)$$

where $v_j(\mu_j, \Sigma_j) = \frac{\Sigma_j ww'M^1_{o,j}}{w'\Sigma_j w}$.

Now, we take derivative of $Q(\Theta|\Theta^l)$ with respect to Σ_j.

First, we can get

$$w'M_{o,j}^2 w = w'\Sigma_j w + w'\Sigma_j w(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})[\frac{\phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}]$$

where $M_{o,j}^2$ is M_o^2 with $\mu = \mu_j$ and $\Sigma = \Sigma_j$.

Next, take derivative of $log(1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}))$ with respect to Σ_j

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_j}[log(1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}}))] &= \frac{1}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}\left\{\frac{1}{\sqrt{\pi}}[\frac{w'\mu_j}{2\sqrt{2}(w'\Sigma_j w)^{\frac{3}{2}}}ww'exp(-(\frac{-w'\mu_j}{\sqrt{2}\sqrt{w'\Sigma_j w}})^2)]\right\} \\
&= \frac{1}{2}(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})(\frac{\phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})}{1-\Phi(\frac{-w'\mu_j}{\sqrt{w'\Sigma_j w}})})(\frac{-ww'}{w'\Sigma_j w}) \\
&= \frac{1}{2}\frac{w'M_{o,j}^2 w - w'\Sigma_j w}{w'\Sigma_j w}(\frac{-ww'}{w'\Sigma_j w}) \\
&= \frac{1}{2}w[\frac{1}{w'\Sigma_j w} - \frac{w'M_{o,j}^2 w}{(w'\Sigma_j w)^2}]w'
\end{aligned}$$

Then, we take the derivative of $Q(\Theta|\Theta^l)$ with respect to Σ_j

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_j}[Q(\Theta|\Theta^l)] &= \frac{1}{T}\sum_{t=1}^{T}\tilde{z}_{tj}\left\{-\frac{1}{2}\Sigma_j^{-1} + \frac{1}{2}\Sigma_j^{-1}(Y_t-\mu_j)(Y_t-\mu_j)'\Sigma_j^{-1}\right. \\
&\quad \left. -\frac{1}{2}w[\frac{1}{w'\Sigma_j w} - \frac{w'M_{o,j}^2 w}{(w'\Sigma_j w)^2}]w'\right\} \\
&= 0
\end{aligned}$$

Some linear algebra properties were used: $\frac{\partial log|A|}{\partial A} = (A')^{-1}$ and $\frac{\partial x'A^{-1}x}{\partial A} = -A^{-1}xx'A^{-1}$.

Finally, we can get:

$$\Sigma_j \quad = \quad \frac{\sum_{t=1}^{T} \tilde{z}_{tj}(Y_t - \mu_j)(Y_t - \mu_j)'}{\sum_{t=1}^{T} \tilde{z}_{tj}} + I_j(\mu_j, \Sigma_j)$$

where $I_j(\mu_j, \Sigma_j) = \Sigma_j w[\frac{1}{w'\Sigma_j w} - \frac{w' M_{o,j}^2 w}{(w'\Sigma_j w)^2}]w'\Sigma_j$.

C.4 E step of the new EM algorithm

$$
\begin{aligned}
&E[L^C(\Theta)|Y,\Theta^l] \\
=&E_{z,n|Y,\Theta^l}\{E[L^C(\Theta)|z,n,Y,\Theta^l]\} \\
=&E_{z,n|Y,\Theta^l}\{E[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} z_{tj}(log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \sum_{k=1}^{n_t} logf_j^N(Y_{t,k}))|z,n,Y,\Theta^l]\} \\
=&E_{z,n|Y,\Theta^l}\{\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} z_{tj}(log\alpha_j + logf_j^N(Y_{t,n_t+1}) + n_t E[logf_j^N(Y_{t,k})|z,n,Y,\Theta^l])\} \\
=&E_{z|Y,\Theta^l}\{\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} z_{tj}(log\alpha_j + logf_j^N(Y_{t,n_t+1}) + E(n_t|z,Y,\Theta^l)E[logf_j^N(Y_{t,k})|z,n,Y,\Theta^l])\} \\
=&E_{z|Y,\Theta^l}\{\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} z_{tj}(log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \\
&(\sum_{n_t=0}^{\infty} n_t \prod_{h=1}^{P}[(1-F_h^l)^{n_t}F_h^l]^{z_{th}})(\int logf_j^N(Y_{t,k}) \prod_{m=1}^{P}(\frac{f_m^{N,l}(Y_{t,k})}{1-F_m^l})^{z_{tm}} dY_{t,k}))\} \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} E_{z|Y,\Theta^l}\{z_{tj}(log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \\
&(\sum_{n_t=0}^{\infty} n_t \prod_{h=1}^{P}[(1-F_h^l)^{n_t}F_h^l]^{z_{th}})(\int logf_j^N(Y_{t,k}) \prod_{m=1}^{P}(\frac{f_m^{N,l}(Y_{t,k})}{1-F_m^l})^{z_{tm}} dY_{t,k}))\} \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} P(z_{tj}|Y,\Theta^l)[log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \\
&\frac{1-F_j^l}{F_j^l}(\int logf_j^N(Y_{t,k})(\frac{f_j^{N,l}(Y_{t,k})}{1-F_j^l})dY_{t,k})] \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{P} \tilde{z}_{tj}[log\alpha_j + logf_j^N(Y_{t,n_t+1}) + \tilde{n}_{t,j}(\int logf_j^N(Y_{t,k})(\frac{f_j^{N,l}(Y_{t,k})}{1-F_j^l})dY_{t,k})]
\end{aligned}
$$

where $E_{z,n|Y,\Theta^l}(.)$ takes the joint expectation of z and n conditional on Y and Θ^l.

Law of iterated expectation $E(Y|X) = E[E(Y|Z,X)|X]$ was used.

C.5 M step of the new EM algorithm

To begin with, we derive the first two moments for Y coming from the invalid truncation area ($x < y$), whose density of the has the following form:

$$f(Y,\mu,\Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}[1-\Phi(\frac{w'\mu}{\sqrt{w'\Sigma w}})]} exp[-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)] \qquad \text{(C.14)}$$

Let $Y^d = Y - \mu$. Then, the first and second moments of $Y^d = \begin{pmatrix} x^d \\ y^d \end{pmatrix}$ are:

$$M_d^1 = \frac{-\Sigma w}{\sqrt{w'\Sigma w}} \frac{\phi(\frac{w'\mu}{\sqrt{w'\Sigma w}})}{1-\Phi(\frac{w'\mu}{\sqrt{w'\Sigma w}})}$$
$$M_d^2 = \Sigma + \frac{\Sigma w w'\Sigma}{w'\Sigma w} \frac{w'\mu}{\sqrt{w'\Sigma w}} \frac{\phi(\frac{w'\mu}{\sqrt{w'\Sigma w}})}{1-\Phi(\frac{w'\mu}{\sqrt{w'\Sigma w}})}$$

- Take derivative of (3.14) with respect to μ_j.

$$\frac{\partial Q(\Theta|\Theta^l)}{\partial \mu_j} = \frac{1}{T}\sum_{t=1}^{T} \tilde{z}_{tj}[\Sigma_j^{-1}Y_t - \Sigma_j^{-1}\mu_j + \tilde{n}_{t,j}\int(\Sigma_j^{-1}Y_{t,k} - \Sigma_j^{-1}\mu_j)(\frac{f_j^{N,l}(Y_{t,k})}{1-F_j^l})dY_{t,k}] = 0$$
$$\Rightarrow \sum_{t=1}^{T}\tilde{z}_{tj}Y_t - \mu_j\sum_{t=1}^{T}\tilde{z}_{tj} + \sum_{t=1}^{T}\tilde{z}_{tj}\tilde{n}_{t,j}(M_{d,j}^{1,l} + \mu_j^l) - \mu_j\sum_{t=1}^{T}\tilde{z}_{tj}\tilde{n}_{t,j} = 0$$
$$\Rightarrow \mu_j^{l+1} = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}(Y_t + \tilde{n}_{t,j}(M_{d,j}^{1,l} + \mu_j^l))}{\sum_{t=1}^{T}\tilde{z}_{tj}(1+\tilde{n}_{t,j})}$$

where $M_{d,j}^{1,l}$ is M_d^1 with $\mu = \mu_j^l$, $\Sigma = \Sigma_j^l$.

- Take derivative of (3.14) with respect to Σ_j^{-1}.

$$\frac{\partial Q(\Theta|\Theta^l)}{\partial \Sigma_j^{-1}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{z}_{tj}[\frac{1}{2}\Sigma_j - \frac{1}{2}(Y_t-\mu_j^{l+1})(Y_t-\mu_j^{l+1})' +$$

$$\tilde{n}_{t,j}\int(\frac{1}{2}\Sigma_j - \frac{1}{2}(Y_{t,k}-\mu_j^{l+1})(Y_{t,k}-\mu_j^{l+1})')(\frac{f_j^{N,l}(Y_{t,k})}{1-F_j^l})dY_{t,k}] = 0$$

$$\Rightarrow \sum_{t=1}^{T}\tilde{z}_{tj}\Sigma_j - \sum_{t=1}^{T}\tilde{z}_{tj}(Y_t-\mu_j^{l+1})(Y_t-\mu_j^{l+1})' + \sum_{t=1}^{T}\tilde{z}_{tj}\tilde{n}_{t,j}\Sigma_j - \sum_{t=1}^{T}\tilde{z}_{tj}\tilde{n}_{t,j}M_{d',j}^2 = 0$$

$$\Rightarrow \Sigma_j^{l+1} = \frac{\sum_{t=1}^{T}\tilde{z}_{tj}[(Y_t-\mu_j^{l+1})(Y_t-\mu_j^{l+1})' + \tilde{n}_{t,j}M_{d',j}^2]}{\sum_{t=1}^{T}\tilde{z}_{tj}(1+\tilde{n}_{t,j})}$$

where $M_{d',j}^2 = M_{d,j}^{2,l} + (\mu_j^l - \mu_j^{l+1})(M_{d,j}^{1,l})' + (M_{d,j}^{1,l})(\mu_j^l - \mu_j^{l+1})' + (\mu_j^l - \mu_j^{l+1})(\mu_j^l - \mu_j^{l+1})'$, and $M_{d,j}^{2,l}$ is M_d^2 with $\mu = \mu_j^l$, $\Sigma = \Sigma_j^l$.

C.6 M step of the new EM algorithm (conditional case)

The closed-form solution for α_j and Σ_j can be easily derived similar to the unconditional case. Here we focus on A_j. Notice that maximizing $Q(\Psi|\Psi^l)$ is equivalent to minimizing the following expression for the purpose of taking derivative with respect to A_j:

$$\begin{aligned} L(A) &= \sum_{t=P+1}^{T}\sum_{j=1}^{P}\tilde{z}_{tj}[(Y_t - A_jX_{t-1})'\Sigma_j^{-1}(Y_t - A_jX_{t-1}) + \\ &\quad \tilde{n}_{t,j}\int^{Tr}((Y_{t,k} - A_jX_{t-1})'\Sigma_j^{-1}(Y_{t,k} - A_jX_{t-1}))(\frac{f_{t,j}^{N,l}(Y_{t,k})}{1-F_{t,j}^l})dY_{t,k}] \\ &= \sum_{j=1}^{P}\{[vec(\bar{Y}_j) - (I_2\otimes\bar{X}_j)vec(A_j')]'(\Sigma_j^{-1}\otimes I_{T-Q})[vec(\bar{Y}_j) - (I_2\otimes\bar{X}_j)vec(A_j')] + \\ &\quad \int[vec(\tilde{Y}_j) - (I_2\otimes\tilde{X}_j)vec(A_j')]'(\Sigma_j^{-1}\otimes I_{T-Q})[vec(\tilde{Y}_j) - (I_2\otimes\tilde{X}_j)vec(A_j')]f_j^l(\tilde{Y}_j)d\tilde{Y}_j\} \end{aligned}$$

where $\tilde{Y}_j = \sqrt{(\tilde{z}_j \odot \tilde{n}_j)\tau} \odot Y_k$, and $Y_k = (Y_{Q+1,k}, ..., Y_{T,k})'$. Take derivative of $L(A)$ with respect to $vec(A_j')$:

$$\frac{\partial L(A)}{\partial vec(A_j')}$$

$$\begin{aligned}
&= -2(I_2 \otimes \bar{X}_j)(\Sigma_j^{-1} \otimes I_{T-Q})vec(\bar{Y}_j) + 2(I_2 \otimes \bar{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})(I_2 \otimes \bar{X}_j)vec(A_j') + \\
&\int [-2(I_2 \otimes \tilde{X}_j)(\Sigma_j^{-1} \otimes I_{T-Q})vec(\tilde{Y}_j) + 2(I_2 \otimes \tilde{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})(I_2 \otimes \tilde{X}_j)vec(A_j')]f(\tilde{Y}_j)d\tilde{Y}_j \\
&= -(I_2 \otimes \bar{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})vec(\bar{Y}_j) + (I_2 \otimes \bar{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})(I_2 \otimes \bar{X}_j)vec(A_j') - \\
&(I_2 \otimes \tilde{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})vec(\tilde{M}^1_{d',\bar{T},j}) + (I_2 \otimes \tilde{X}_j)'(\Sigma_j^{-1} \otimes I_{T-Q})(I_2 \otimes \tilde{X}_j)vec(A_j') \\
&= -[(\Sigma_j^{-1} \otimes \tilde{X}_j')vec(\tilde{M}^1_{d',\bar{T},j}) + (\Sigma_j^{-1} \otimes \bar{X}_j')vec(\bar{Y}_j)] + [(\Sigma_j^{-1} \otimes \bar{X}_j'\bar{X}_j) + (\Sigma_j^{-1} \otimes \tilde{X}_j'\tilde{X}_j)]vec(A_j') \\
&= -[vec(\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j}\Sigma_j^{-1}) + vec(\bar{X}_j'\bar{Y}_j\Sigma_j^{-1})] + [\Sigma_j^{-1} \otimes (\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)]vec(A_j') \\
&= -(\Sigma_j^{-1} \otimes I_2)vec(\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j} + \bar{X}_j'\bar{Y}_j) + [\Sigma_j^{-1} \otimes (\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)]vec(A_j') \\
&= 0
\end{aligned}$$

Then, we can write down $vec(A_j')$ as:

$$\begin{aligned}
&vec(A_j') \\
&= [\Sigma_j^{-1} \otimes (\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)]^{-1}(\Sigma_j^{-1} \otimes I_2)vec(\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j} + \bar{X}_j'\bar{Y}_j) \\
&= (I_2 \otimes (\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)^{-1})vec(\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j} + \bar{X}_j'\bar{Y}_j) \\
&= vec[(\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)^{-1}(\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j} + \bar{X}_j'\bar{Y}_j)]
\end{aligned}$$

Therefore, we have

$$A_j^{l+1} = (\tilde{X}_j'\tilde{M}^1_{d',\bar{T},j} + \bar{X}_j'\bar{Y}_j)'(\bar{X}_j'\bar{X}_j + \tilde{X}_j'\tilde{X}_j)^{-1}$$

C.7 Proof of Theorem 1

First, we introduce a lemma that shows the mixture truncated normal distribution is identifiable.

Lemma 1. *Let* $\nu = (\mu, \Sigma)$*, and suppose that* $\Lambda = \{F(Y,\nu); \nu \in \mathbb{R}^6, Y \in \mathbb{R}^2\}$ *is the family of distributions* whose *density is given by*

$$f(Y,\nu) = \frac{1}{2\pi\sqrt{|\Sigma|}[1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})]} exp[-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)] \tag{C.15}$$

Then $\psi_\gamma(Y) = \sum_{j=1}^P \alpha_j F(Y,\nu_j)$*, the class of finite mixtures of* Λ*, is identifiable.* $\gamma = \{\alpha_j, \nu_j | \forall j\}$*,* $\alpha_j > 0$*, and* $\sum_{j=1}^P \alpha_j = 1$*. In other words,* $\psi_\gamma(Y) = \psi_{\gamma^*}(Y) \Rightarrow \gamma = \gamma^*$.

Proof of Lemma 1:

We first define the exponential family.

If, for some $\sigma-$finite measure μ,

$$dF(Y,\tau) = a(\tau)b(Y)exp[\tau' h(Y)]d\mu(Y) \tag{C.16}$$

for $Y \in \mathbb{R}^n$, $\tau(m \times 1)$, and $h(Y)$ $(m \times 1)$, where $a(\tau) > 0, b(Y) \geq 0$ and a, b, h_j, for $j = 1, 2, \ldots, m$ are all measurable, then F is called an exponential family member.

Barndorff-Nielsen (1965) proves that the class ψ is identifiable if all of the following hold: (a) F belongs to the exponential family, (b) μ is n-dimensional Lebesgue measure, (c) functions h_j, $j = 1, 2, \ldots, m$, are all continuous, and (d) the set $\{y : y = h(Y), b(Y) > 0, Y \in \mathbb{R}^n\}$ contains a nonempty open set.

First, we show that the distribution with density given by (C.15) belongs to exponential family as it can be written as:

$$\begin{aligned} \frac{dF(Y,\tau)}{d\mu(Y)} &= \frac{1}{2\pi\sqrt{|\Sigma|}[1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})]} exp[-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)] \\ &= a(\tau)b(Y)exp[\tau' h(Y)] \end{aligned}$$

where μ is two-dimensional Lebesgue measure. $\tau = \left(\Sigma^{-1}\mu, \quad -\frac{1}{2}vec(\Sigma^{-1})\right)$, $a(\tau) = \left\{\sqrt{|\Sigma|}[1-\Phi(\frac{-w'\mu}{\sqrt{w'\Sigma w}})]exp(\frac{1}{2}\mu'\Sigma^{-1}\mu)\right\}^{-1}$, $b(Y) = \frac{1}{2\pi}$, and $h(Y) = \left(Y, \quad vec(YY')\right)'$.

The image of the mapping h: $\mathbb{R}^2 \to \mathbb{R}^6$, for $x \geq y$ is the set $\Omega = \{h(Y), x \geq y\}$, which contains an open set $\Omega' = \{h(Y), x > y\}$. In addition, the map from τ to ν is unique. Lemma 1 follows. □

Now, we can proceed to prove Theorem 1. It is straightforward to see that $L(\Psi)$ is a measurable function of data for each $\Psi \in \Xi$, and continuous in Ψ. Therefore, it suffices to show that (a) the log-likelihood follows a uniform strong law of large numbers: $\sup_{\Psi \in \Xi} |$ $L(\Psi) - E[L(\Psi)] \mid \to 0$ a.s. as $T \to \infty$; (b) the identification condition: $E[L(\Psi)] \leq E[L(\Psi_0)]$, and $E[L(\Psi)] = E[L(\Psi_0)]$ implies $\Psi = \Psi_0$. (see Amemiya (1973, Lemma 3)).

Let $L(\Psi) = \frac{1}{T-P}\sum_t l(\Psi)$. By Assumption 1 and continuity of $l(\Psi)$, $l(\Psi)$ is stationary and ergodic (see Krengel (1985, Proposition 4.3)), and hence $E[L(\Psi)] = E[l(\Psi)]$. To verify (a), it suffices to show that $E[\underset{\Psi\in\Xi}{sup} \mid l(\Psi) \mid] < \infty$ (see Rao (1962) or Straumann and Mikosch (2006 Theorem 2.7)). Kalliovirta et.al. (2016) prove the the above inequality holds for the likelihood in their model one side at a time. We are going to adapt similar similar procedures here. Specifically, we know that

$$l(\Psi) = log\{\sum_{j=1}^{P} \alpha_j (2\pi)^{-1} |\Sigma_j|^{-1/2}$$

$$exp[-\frac{1}{2}(Y_t - A_j X_{t-1})'\Sigma_j^{-1}(Y_t - A_j X_{t-1})]/[\frac{1}{2}erfc(-w'A_j X_{t-1}/\sqrt{2w'\Sigma_j w})]\}$$

where $w = (1, -1)'$. Assumption 2 implies that, $\Delta \geq |\Sigma_j| \geq \delta$, $\forall j$ for some $\delta > 0$, and $\Delta < \infty$, and that $w'\Sigma_j w \geq \gamma$, $\forall j$ for some $\gamma > 0$. We also know that $exp[-\frac{1}{2}(Y_t - A_j X_{t-1})'\Sigma_j^{-1}(Y_t - A_j X_{t-1})] \leq 1$. In addition, when $-w'A_j X_{t-1}/\sqrt{2w'\Sigma_j w} \leq 0$, $erfc(-w'A_j X_{t-1}/\sqrt{2w'\Sigma_j w}) \geq 1$, and thus we can see that $l(\Psi) \leq log(\pi^{-1}\delta^{-1/2})$. When $-w'A_j X_{t-1}/\sqrt{2w'\Sigma_j w} > 0$, we apply the inequality $erfc(x) \geq \frac{1}{2}exp(-2x^2)$ (see Chang et. al. (2011, Theorem 2)), thus

$$
\begin{aligned}
erfc(-w'A_jX_{t-1}/\sqrt{2w'\Sigma_j w}) &\geq \frac{1}{2}exp(-w'A_jX_{t-1}X'_{t-1}A'_jw/w'\Sigma_j w) \\
&\geq \frac{1}{2}exp[-\frac{1}{\gamma}tr(X_{t-1}X'_{t-1}A'_jww'A_j)] \\
&\geq \frac{1}{2}exp[-\frac{1}{\gamma}tr(X_{t-1}X'_{t-1})tr(A'_jww'A_j)] \\
&\geq \frac{1}{2}exp[-\frac{\kappa}{\gamma}X'_{t-1}X_{t-1}]
\end{aligned}
$$

where the last inequality holds by compactness of Ξ (Assumption 2). That is, $tr(A'_jww'A_j) \leq \kappa$, $\forall j$ for some $0 < \kappa < \infty$. Now, it can be seen that

$$
\begin{aligned}
l(\Psi) &\leq log\{\sum_{j=1}^{P}\alpha_j(2\pi)^{-1}\delta^{-1/2}4exp[\frac{\kappa}{\gamma}X'_{t-1}X_{t-1}]\} \\
&= log(2\pi^{-1}\delta^{-1/2}) + \frac{\kappa}{\gamma}X'_{t-1}X_{t-1}
\end{aligned}
$$

Therefore, regardless of the value of $-w'A_jX_{t-1}/\sqrt{2w'\Sigma_j w}$, we have $l(\Psi) \leq log(2\pi^{-1}\delta^{-1/2}) + \frac{\kappa}{\gamma}X'_{t-1}X_{t-1}$.

On the other hand, it can be seen that

$$
\begin{aligned}
&(Y_t - A_j X_{t-1})' \Sigma_j^{-1} (Y_t - A_j X_{t-1}) \\
=&tr[(Y_t - A_j X_{t-1})(Y_t - A_j X_{t-1})' \Sigma_j^{-1}] \\
\leq&tr[(Y_t - A_j X_{t-1})(Y_t - A_j X_{t-1})'] tr(\Sigma_j^{-1}) \\
=&(Y_t - A_j X_{t-1})'(Y_t - A_j X_{t-1}) tr(\Sigma_j^{-1}) \\
\leq&(1 + Y_t' Y_t + X_{t-1}' X_{t-1})\rho
\end{aligned}
$$

where the first inequality holds because both $(Y_t - A_j X_{t-1})(Y_t - A_j X_{t-1})'$ and Σ_j^{-1} are positive semi-definite. The second last inequality is implied by Cauchy-Schwarz inequality and Assumption 2 ($tr(\Sigma_j^{-1}) \leq \rho$, $\forall j$ for some $0 < \rho < \infty$). Furthermore, $erfc(-w' A_j X_{t-1} / \sqrt{2w' \Sigma_j w}) \leq 2$, thus

$$
\begin{aligned}
l(\Psi) &\geq log\{\sum_{j=1}^{P} \alpha_j (2\pi)^{-1} \Delta^{-1/2} exp[-\frac{1}{2}(1 + Y_t' Y_t + X_{t-1}' X_{t-1})\rho]\} \\
&= G_1 - \frac{1}{2}\rho(1 + Y_t' Y_t + X_{t-1}' X_{t-1})
\end{aligned}
$$

for some finite G_1. Overall, we have $G_1 - \frac{1}{2}\rho(1 + Y_t' Y_t + X_{t-1}' X_{t-1}) \leq l(\Psi) \leq log(2\pi^{-1}\delta^{-1/2}) + \frac{\kappa}{\gamma} X_{t-1}' X_{t-1}$, from which $E[\sup_{\Psi \in \Xi} \mid l(\Psi) \mid] < \infty$ holds because $X_{t-1}' X_{t-1} = 1 + Y_{t-1}' Y_{t-1} + ... + Y_{t-Q}' Y_{t-Q}$, and $E(Y_t' Y_t) < \infty$ for all t by Assumption 3.

Now, we verify (b). Let $s(Y_{t-Q}^{t-1}, \Psi_0)$ be the stationary distribution of Y_{t-Q}^{t-1} as , then

$$
\begin{aligned}
&E[L(\Psi)] - E[L(\Psi_0)] \\
&= \iint s(Y_{t-Q}^{t-1}, \Psi_0)[\sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})] log \frac{\sum_{j=1}^{P} \alpha_j f_j(Y_t|Y_{t-Q}^{t-1}, A_j, \Sigma_j)}{\sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})} dY_t dY_{t-Q}^{t-1} \\
&= \int s(Y_{t-Q}^{t-1}, \Psi_0)\{\int [\sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})] log \frac{\sum_{j=1}^{P} \alpha_j f_j(Y_t|Y_{t-Q}^{t-1}, A_j, \Sigma_j)}{\sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})} dY_t\} dY_{t-Q}^{t-1}
\end{aligned}
$$

where the inner integral is the negative Kullback-Leibler divergence between two mixture densities: $\sum_{j=1}^{P} \alpha_j f_j(Y_t|Y_{t-Q}^{t-1}, A_j, \Sigma_j)$ and $\sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})$. Therefore, $E[L(\Psi)] - E[L(\Psi_0)] \leq 0$ and the equality holds if and only if

$$
\sum_{j=1}^{P} \alpha_j f_j(Y_t|Y_{t-Q}^{t-1}, A_j, \Sigma_j) = \sum_{j=1}^{P} \alpha_{j,0} f_j(Y_t|Y_{t-Q}^{t-1}, A_{j,0}, \Sigma_{j,0})
$$

By the identification result from Lemma 1, we have that $\alpha_j = \alpha_{j,0}$, $\Sigma_j = \Sigma_{j,0}$ and $A_j X_{t-1} = A_{j,0} X_{t-1}$ for all j, where $A_j X_{t-1} = A_{j,0} X_{t-1}$ implies either that $A_j = A_{j,0}$ or that X_{t-1} takes values only on a $2(Q-1)$ dimensional hyperplane. The latter is impossible as $\{X_{t-1}\}$ takes values on $H \subset \mathbb{R}^{2Q}$, where H has positive Lebesque measure. Therefore, $\alpha_j = \alpha_{j,0}$, $\Sigma_j = \Sigma_{j,0}$ and $A_j = A_{j,0}$ for all j.