Title
The Econometric Analysis of Interval-valued Data and Adaptive Regression Splines

Permalink
https://escholarship.org/uc/item/3j693877

Author
Lin, Wei

Publication Date
2013

Peer reviewed|Thesis/dissertation
The Econometric Analysis of Interval-valued Data
and
Adaptive Regression Splines

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

Doctor of Philosophy

in

Economics

by

Wei Lin

June 2013

Dissertation Committee:
Dr. Goloria González-Rivera, Co-Chairperson
Dr. Aman Ullah, Co-Chairperson
Dr. Todd Sorensen
The Dissertation of Wei Lin is approved:

______________________________

______________________________
Committee Co-Chairperson

______________________________
Committee Co-Chairperson

University of California, Riverside
Acknowledgments

I would like to gratefully and sincerely thank Dr. Gloria González-Rivera and Dr. Aman Ullah for their guidance, understanding, patience, and most importantly, their encouragement during my graduate studies at UC Riverside. Their mentorship was paramount in providing a well rounded experience which is consistent with my long-term career goals. I am truly fortunate to have had the opportunity to work with them. Without their help, I would not have been here.

I would also like to thank my committee members, Dr. Todd Sorensen for his thought provoking suggestions and the general collegiality offered to me over the years. The remote server of 12-core Mac Pro in his office saved me a lot of time in performing numerous simulations and bootstraps. In addition, I'd like to recognize all the other professors who taught me, as each of them made to my intellectual growth during my years of study at the University of California Riverside.

Finally, I would like to thank Yingying Sun, Mingming Jiang and other friends in Riverside who helped me a lot in my life, and turned my 5 years stay in Riverside into an enjoyable and immemorial experience.
This dissertation is dedicated to my parents and my brother, who suered from substantial pain and stress ever since 1984 when I was born as the second child in the family, for their courage and devotion.
ABSTRACT OF THE DISSERTATION

The Econometric Analysis of Interval-valued Data
and
Adaptive Regression Splines

by

Wei Lin

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2013
Dr. Goloria González-Rivera, Co-Chairperson
Dr. Aman Ullah, Co-Chairperson

Chapter 1, 3 and 4 focus on the analysis of interval-valued data (joint with Professor González-Rivera). In Chapter 1, we propose a constrained regression model that preserves the natural order of the interval in all instances. Within the framework of interval time series, we specify a general dynamic bivariate system for the upper and lower bounds of the intervals, and propose a (modified) two-step estimator. Monte Carlo simulations show good finite sample properties of the proposed estimators. We model the daily interval of low/high SP500 returns before and after 2007, and find that truncation is very severe during and after the financial crisis of 2008, so that a modified two-step procedure should be implemented. In Chapters 3 and 4, we adopt an alternative modelling approach for interval-valued data that exploits the extreme property of lower/upper bounds of interval, which is ignored in the existing literature. Specifically, Chapter 3 and 4 propose two different models and estimation strategies (ML and semiparametric estimation) that combines the knowledge of order statistics and extreme
value theory with interval-valued data respectively.

As a separate strand of research, in Chapter 2 (joint with Professor Ullah), we propose an adaptive spline estimator based on Friedman (1991)’s multivariate adaptive regression splines. The model takes the form of an expansion in the cross product spline bases, where the numbers of spline functions, the degree of tensor product and knot locations are automatically selected adaptively by using generalized cross validation. Our estimator is more tractable not only in computational implementations but in theoretical deductions as well. We establish the asymptotic normality of our adaptive estimator, and obtain the optimal convergence rate that it can possibly achieve. The optimal convergence rate depends on the order ratio of the number of selected spline basis functions to the total potential ones. The Monte Carlo simulation, comparing the adaptive estimator with classical regression splines given various DGP settings, shows that our estimator has more significant improvement upon classical regression splines by producing smaller AMSE given the DGP with multivariate covariates. We also apply our adaptive estimator to the study of the effect of public capital stock on the gross state product using the pooled panel data set in Baltagi and Pinnoi (1995).
## Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>1 Constrained Regression for Interval-valued Data</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 General Framework and Basic Assumptions</td>
<td>7</td>
</tr>
<tr>
<td>1.3 Estimation</td>
<td>12</td>
</tr>
<tr>
<td>1.3.1 Conditional log-likelihood function</td>
<td>13</td>
</tr>
<tr>
<td>1.3.2 Two-Step Estimation: General Remarks</td>
<td>15</td>
</tr>
<tr>
<td>1.3.3 Two-step Estimation: The First Step</td>
<td>18</td>
</tr>
<tr>
<td>1.3.4 Two-step Estimation: The Second Step</td>
<td>22</td>
</tr>
<tr>
<td>1.3.5 Two-step Estimation: Implementation Issues</td>
<td>26</td>
</tr>
<tr>
<td>1.3.6 Two-step Estimation: A Modified Two-step Estimator</td>
<td>27</td>
</tr>
<tr>
<td>1.4 Simulation</td>
<td>32</td>
</tr>
<tr>
<td>1.5 The SP500 Low/High Return Interval</td>
<td>45</td>
</tr>
<tr>
<td>1.6 Comparison with Existing Approaches</td>
<td>61</td>
</tr>
<tr>
<td>1.6.1 In-Sample Evaluation Criteria: Loss Functions</td>
<td>64</td>
</tr>
<tr>
<td>1.6.2 In-Sample Evaluation Criteria: Mean Estimates, Bias, and MSE</td>
<td>73</td>
</tr>
<tr>
<td>1.6.3 In-Sample Evaluation Criteria: S&amp;P 500 Daily Low/High Interval Returns</td>
<td>80</td>
</tr>
<tr>
<td>1.7 Conclusion</td>
<td>82</td>
</tr>
<tr>
<td>Bibliography</td>
<td>85</td>
</tr>
</tbody>
</table>

2 Modified Multivariate Adaptive Regression Splines | 87 |
| 2.1 Introduction | 87 |
| 2.2 Model | 90 |
| 2.2.1 General Setting | 90 |
| 2.2.2 Adaptive Selection Procedure | 92 |
| 2.3 Asymptotic Properties | 95 |
| 2.4 Monte Carlo | 100 |
| 2.4.1 Data Generating Processes | 100 |
| 2.4.2 Evaluation | 101 |
| 2.5 Empirical Application | 106 |
| 2.5.1 Data | 106 |
3 Maximum Likelihood Estimation of Interval-valued Returns and Trading Intensity

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
3.2 General Framework . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116
3.3 A Special Case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119
  3.3.1 Estimation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119
  3.3.2 Forecasting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 121
3.4 Empirical Application . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
3.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 129

Bibliography 132

4 Semiparametric Estimation of Interval-valued Time Series Using Extreme Value Approach

4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
4.2 Basic Assumptions and Methodology . . . . . . . . . . . . . . . . . . . . 138
4.3 Estimation Procedure . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149
4.4 The Five-minute Low/High Stock Returns of Wells Fargo . . . . . . . . 155
4.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 170

Bibliography 172

Appendix A Estimation of (σ_l, σ_u, ρ) in Two-step Estimation 174

Appendix B Proof of Theorem 1.3.2 176

Appendix C Proof of Proposition 1.3.1 181

Appendix D Proof of Theorem 4.3.1 185

Appendix E Primitive Conditions for Theorem 4.3.1 187
List of Figures

1.1 Truncated Distribution of the Error Term .................................................. 9
1.2 Inverse of the Mill’s Ratio ........................................................................... 26
1.3 Minimum Distance Estimator ......................................................................... 31
1.4 Daily SP500 Low/High Returns: Stable Period (2004/1/1-2007/1/1) .......... 47
1.5 Daily SP500 Low/High Returns: Unstable Period (2007/1/1-2011/4/29) ... 48
1.6 Estimated Inverse Mill’s Ratio: Stable Period (2004/1/1-2007/1/1) ........... 51
1.7 Estimated Inverse Mill’s Ratio: Unstable Period (2007/1/1-2011/4/29) ....... 51
1.8 Observability Restriction: Stable Period (2004/1/1-2007/1/1) ...................... 60
1.9 Observability Restriction: Unstable Period (2007/1/1-2011/4/29) ............... 60

2.1 Histograms of MSE for Splines with Uniformity in DGPs ....................... 104
2.2 Histograms of MSE for Splines with Mixed Normality in DGPs ................. 105
2.3 Average Marginal Effects of Public Capital Stocks on GSP for Each States (the United States) ................................................................. 110
2.4 Average Marginal Effects of Public Capital Stocks on GSP for Each Year (California) ................................................................. 111

3.1 Time Series Plots of Daily Prices/Returns and Number of Trades ............ 124
3.2 ACF and PACF for Number of Trades .......................................................... 127
3.3 Time Series Plots of Daily Prices/Returns and Number of Trades ............ 130
4.1 Time Series Plots of 5-min Low/High Returns and Number of Trades ........ 156
4.2 Fitted Returns from Competing Models ...................................................... 167
4.3 Fitted Returns from Proposed Models (P1 and P2) ...................................... 168
4.4 Fitted Returns from Proposed Models (P3 and P4) ...................................... 169
List of Tables

1.1 Specification of Data Generating Process ........................................ 34
1.2 Simulation Results for Case B-1 .................................................. 36
1.3 Simulation Results for Case B-2 .................................................. 37
1.4 Simulation Results for Case B-3 .................................................. 38
1.5 Simulation Results for Case B-4 .................................................. 39
1.6 Simulation Results for Case NB-1 ................................................ 40
1.7 Simulation Results for Case NB-2 ................................................ 41
1.8 Simulation Results for Case NB-3 ................................................ 42
1.9 Simulation Results for Case NB-4 ................................................ 43
1.10 Descriptive Statistics for Stable Period ....................................... 47
1.11 Descriptive Statistics for Unstable Period .................................... 48
1.12 First Step Estimation for Stable Period ...................................... 49
1.13 First Step Estimation for Unstable Period .................................... 50
1.14 Block Sizes for Stationary Block Bootstrapping ............................ 53
1.15 Lower Bound Regression Results of Three Models (Stable Period) ... 54
1.16 Upper Bound Regression Results of Three Models (Stable Period) ... 55
1.17 Lower Bound Regression Results of Three Models (Unstable Period) .. 57
1.18 Upper Bound Regression Results of Three Models (Unstable Period) .. 58
1.19 Estimation Methodologies ........................................................... 64
1.20 Data Generating Processes .......................................................... 64
1.21 Methodology Evaluation for DGP1 (HIGH persistence and BINDING observability restriction) .............................................. 68
1.22 Methodology Evaluation for DGP2 (LOW persistence and BINDING observability restriction) .............................................. 69
1.23 Methodology Evaluation for DGP3 (HIGH persistence and NON-BINDING observability restriction) ...................................... 70
1.24 Methodology Evaluation for DGP4 (LOW persistence and NON-BINDING observability restriction) ...................................... 71
1.25 Simulation Results of DGP1 and GDP2 with Multivariate Normal Errors 74
1.26 Simulation Results of DGP3 and DGP4 with Multivariate Normal Errors 75
1.27 Simulation Results of DGP1 and DGP2 with Multivariate Student-t Errors 77
1.28 Simulation Results of DGP3 and DGP4 with Multivariate Student-t Errors 78
1.29 Methodology Evaluation for S&P500 Daily Low/High Interval Returns 81
2.1 Model Comparison ................................................................. 102
Chapter 1

Constrained Regression for

Interval-valued Data

1.1 Introduction

With the advent of sophisticated information systems, data collection has become less costly and, as a result, massive data sets have been generated in many disciplines. Economics and business are not exceptions. For instance, financial data is available at very high frequencies for almost every asset that is transacted in a public market providing data sets with millions of observations. Marketing data sets offer high granularity about consumers and products characteristics. Environmental stations produce data sets that contain high and low frequency records of temperatures, atmospheric conditions, pollutants, etc. across many regions. Statistical institutes, like the Census Bureau, collect socioeconomic information about all individuals in a nation. These massive information data sets tend to be released in an aggregated format, either because of confidential-
ity reasons or because the interest of study is not the individual unit but a collective of units. In these cases, the researcher does not face classical data sets, i.e. \( \{ y_i \} \) for \( i = 1, \cdots, n \) or \( \{ y_t \} \) for \( t = 1, \cdots, T \) where \( y_i \) or \( y_t \) are single values in the real line, but data aggregated in some fashion, like interval data \([y_l, y_u]\) that offers information on the lower and upper bound of the variable of interest. For example, information about income or networth comes very often in interval format, or low and high prices of an asset in a given day, or daily temperature intervals, or low/high prices of electronic devices for several stores, etc.

Interval-valued data sets are also considered symbolic data sets. Within the symbolic approach pioneered by Billard and Diday (2003, 2006), there are a variety of proposals to fit a regression model to interval data. A review is provided in Arroyo, González-Rivera and Mate (2011). The simplest approach (Billard and Diday, 2000) is to fit a regression model to the centers of the intervals of the dependent variable and of the regressors. Further approaches consider two separate regressions, one for the lower bound and another for the upper bound of the intervals, either with no constraints in the regression coefficients (Billard and Diday, 2002) or by constraining both regressions to share the same regression coefficients (Brito, 2007). In a similar line, Lima Neto and de Carvalho (2008) propose running two different regressions, one for the center and another for the range of the intervals, with no constraints. None of these approaches guarantees that the fitted values from the regressions will satisfy the natural order of an interval, i.e. \( \hat{y}_l \leq \hat{y}_u \), for all observations in the sample. Recently, Lima Neto and de Carvalho (2010) impose non-negativity constraints on the regression coefficients of the model for the range and solve a quadratic programming problem to find the least squares solution. However, for these constraints to be effective, the range regression
must entertain only non-negative regressors (for instance regressing the range of the dependent variable on the ranges of the regressors), which limit the usefulness of the model.

In this paper we propose a regression model, either for cross-sectional or time series data, that guarantees the natural order of the fitted interval bounds for all the observations in the sample, and for any potential interval forecast based on the model. Within the framework of interval time series (ITS), we specify a bivariate system for the lower and upper bounds of the time series. The observability restriction \( y_{l,t} \leq y_{u,t} \) implies that the conditional probability density function of the errors is truncated. Under the assumption of bivariate normal errors, the amount of truncation will depend on the variance-covariance matrix of the errors and it will be time-varying because the truncation is a function of the difference between the conditional means specified for the lower and upper bounds. When the observability restriction is severe, i.e., the truncation of the bivariate density is substantial, not only the conditional expectations of the errors are different from zero but also the errors are correlated with the regressors, thus any least-squares estimation (linear or non-linear) will fail to deliver consistent estimators of the parameters of the model. We propose a two-step estimation procedure, combining maximum likelihood and least squares estimation, that will deliver consistent estimators. The first step consists of modeling the range of the interval, which is distributed as a truncated normal density, to obtain by maximum likelihood estimates of the inverse of the Mill’s ratio \( \hat{\lambda}_{t-1} \) that embodies the severity of the observability restriction. Only when the restriction is severe, the second step is necessary. This step consists of introducing \( \hat{\lambda}_{t-1} \) in a least-squares regression to correct the selection bias imposed by the restriction. However, the estimation in the second step may be plagued with multi-
collinearity problems because in some instances $\hat{\lambda}_{t-1}$ is an almost linear function of the regressors. Since multicollinearity cannot be resolved by dropping some of the regressors, we propose a modified second step by implementing a minimum distance estimator that delivers consistent estimates of all parameters in the model. The advantage of the modified second step is that even when the observability restriction is not severe ($\hat{\lambda}_t \approx 0$ for most $t$), we are able to identify all parameters of the model without much loss in efficiency.

As an illustration of the methods that we propose, we model the interval of daily low/high returns to the SP500 index before and after 2007. Before 2007, the daily interval exhibits very little volatility, but after 2007, volatility is the dominant characteristic due to the events of the financial crisis of 2008. These two periods have very different dynamics. We implement the modified two-step estimator and we find that in the stable period the observability restriction is not severe, so that simple OLS will suffice to estimate a dynamic system for the lower and upper bounds of the interval. In contrast, in the high volatility period the restriction is very severe, thus simple OLS estimates should not be trusted and the second step is necessary to guarantee the consistency of the estimators.

The modeling of the low/high interval is interesting in itself for several reasons. For instance, in technical analysis, several trading strategies are based on the dynamics of a popular object, known as the ”candlestick”, which is composed of two intervals, the low/high and the open/close. In financial econometrics, the low/high interval also provides estimators of the volatility of asset returns, see Parkinson (1980), Yang and Zhang (2000), Alizadeh, Brandt, and Diebold (2002) among others. However, the most important reason for our interest in estimation and forecasting with interval-valued data...
lies on the fact that the only format available for some data sets is the interval format. Financial data sets are exceptional; in general they are very rich and information come in many formats, for instance, available databases contain records of stock prices for every transaction in the market so that we could analyze prices at the highest and the lowest frequencies; there is an almost continuous measurement in the transaction price. But this is not always the case in other areas within economics or in other sciences. Some examples follow.

The US Energy Information Administration gather electricity prices for each state in US. Since there are so many factors affecting the prices of electricity, there is substantial variation across states and across localities in the same state. This agency provides average retail price at the state level in interval format, i.e. min/max price, which is more informative of the realities of this market. The US Department of Agriculture provides livestock prices also in interval format. The Livestock Marketing Information Center (Iowa State University) reports interval prices of several items, for instance, min/max daily beef prices. Though they compute a weighted price, this is not the price of a given transaction, so that the interval min/max contains more valuable information to the participants in the market. In the appraisal industry, the objective is to find a "fair market price" for items, such as real estate, for which the market value cannot be observed directly unless the item is sold. It is standard practice in this industry to record min/max prices of similar items that have had a recent transaction so that the "fair" market price, though non-observable, must be contained within such an interval. Even with financial data sets, it is interesting to note though that bond market data is not as transparent as stock data and bond traders mostly report the bid/ask interval of the transaction, in which the price is contained. In other fields different form economics, for
instance in medicine, we have databases with patient data recorded in interval format, the most indicative is blood pressure measurements i.e., diastolic and systolic pressure (low and high numbers respectively). In earth sciences, temperature records across locations also come in interval format, i.e. min/max temperature for a given location.

These examples show that the low/high interval of a variable is a common format that provides additional information beyond an average measurement, and in some cases, it is the only format available to the researcher. It should be noted that estimating and forecasting with low/high interval-valued data is different from estimating and forecasting two quantiles. The low/high bounds are extremes. In quantile regression, the loss function requires fixing the probability $\alpha$ associated with the quantile. If we wish to approximate the low/high interval with quantile regression, it seems natural to fix $\alpha = 0$ for estimation of the lower bound and $\alpha = 1$ for the estimation of the upper bound, but if the variables of analysis are defined in the domain $(-\infty, +\infty)$, these are also the values of the corresponding $(0, 1)$ quantiles. If our interest is any other quantile, e.g. the interquartile range $[Q_{0.25}, Q_{0.75}]$, and the data is available in a classic point-valued format, then quantile regression could be implemented.

The organization of the paper is as follows. In section 2, we provide the general framework and the basic assumptions. In section 3, we present the two-step estimation procedure and develop the asymptotic properties of the estimators in the first and second steps. In section 4, we conduct extensive Monte Carlo simulations that show the finite sample properties of the two-step and modified two-step estimators. In section 5, we model the daily interval of low/high SP500 returns. In section 6, we provide an extensive comparison of the proposed estimation methods with those existing in the literature, and in section 7, we conclude.
1.2 General Framework and Basic Assumptions

In this section we introduce the general regression framework for interval-valued time series data. The objective is the estimation of a parametric specification of the conditional mean of an interval-valued stochastic process.

Generally, an interval is defined as follows:

**Definition 1.** An interval $[Y]$ over a set $(R, \leq)$ is an ordered pair $[Y_l, Y_u]$ where $Y_l, Y_u \in R$ are the lower and upper bounds of the interval such that $Y_l \leq Y_u$.

We can also define an interval random variable on a probability space $(\Omega, F, P)$ as the mapping $Y : F \rightarrow [Y_l, Y_u] \subset R$. In a time series framework, we further define an interval-valued stochastic process as a collection of interval random variables indexed by time, i.e. $\{Y_t\}$ for $t \in T$; and an interval-valued time series (ITS) as a realization $\{[y_l, y_u]\}$ for $t = 1, 2, \ldots, T$ of an interval-valued stochastic process.

We are interested in modeling the dynamics of the process $\{Y_t\} = \{[Y_l, Y_u]\}$ as a function of an information set that potentially includes not only the past history of the process, i.e. $Y^{t-1} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_0\}$ but also any other exogenous random variables $X_t = \{X_t, X_{t-1}, \ldots, X_0\}$ where $X_t = \{X_{lt}, X_{2t}, \ldots, X_{pt}\}$. In this context, we focus the modeling exercise on establishing the joint dynamics of the lower $\{Y_{lt}\}$ and upper $\{Y_{ut}\}$ bounds taking into account the natural ordering of the interval. Thus, a general data generating process is written as

$$\begin{bmatrix}
Y_t \\
Y_{ut}
\end{bmatrix} =
\begin{bmatrix}
G_l(Y^{t-1}, X_t' \beta_l) \\
G_u(Y^{t-1}, X_t' \beta_u)
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{lt} \\
\varepsilon_{ut}
\end{bmatrix}$$

(1.2.1)
such that

\[ Y_{lt} \leq Y_{ut}, \quad (1.2.2) \]

where \( G_l(\cdot) \) and \( G_u(\cdot) \) are two differentiable functions, \( \beta_l \) and \( \beta_u \) are two \( J \times 1 \) vectors of parameters, and \( \epsilon_t \equiv (\epsilon_{lt}, \epsilon_{ut})' \) is the vector of error terms. We call (1.2.2) the observability restriction to be imposed on the data generating process.

The observability restriction (1.2.2) is the key feature of the specification because it generates two important issues for the estimation of the model (1.2.1). First, the restriction \( Y_{lt} \leq Y_{ut} \) implies a restriction on the distribution of the error vector. By imposing (1.2.2) into (1.2.1), the errors are restricted as follows

\[
G_l(Y^{t-1}, X^t; \beta_l) + \epsilon_{lt} \leq G_u(Y^{t-1}, X^t; \beta_u) + \epsilon_{ut},
\]

\[
\epsilon_{ut} - \epsilon_{lt} \geq G_l(Y^{t-1}, X^t; \beta_l) - G_u(Y^{t-1}, X^t; \beta_u). \quad (1.2.3)
\]

The transformed observability restriction (1.2.3) implies that, conditioning on the information set \( \mathcal{I}_{t-1} \equiv (Y^{t-1}, X^t) \), the joint distribution of \( (\epsilon_{lt}, \epsilon_{ut}) \) is truncated from below.

In Figure 1.1 we describe graphically a truncated joint density of the errors. In the plane formed by the variables \( (\epsilon_{lt}, \epsilon_{ut}) \), the ellipse represents a contour of the joint density, and the 45° degree line \( \epsilon_{ut} = \epsilon_{lt} + (G_l - G_u) \) is the truncation line, which separates the shaded area, where \( Y_{lt} \leq Y_{ut} \) holds, from the area where the restriction is violated.

From Figure 1.1 it is easy to observe that the feasible support for the errors will depend on the variance-covariance matrix of the errors as well as any other parameters affecting the shape and extension of the contours, and on the position of the truncation line,
which is a function of the difference between the two conditional mean functions. Small
dispersion of the errors together with a large difference, i.e. $G_l << G_u$ tend to mitigate
the severity of the observability restriction because it reduces the probability of the
errors falling below the truncation line to the point that the restriction might not longer
be binding and it could be safely removed from the model.

However, if the restriction is binding, it cannot be ignored in the estimation of the model
because, in one hand, it may generated predicted values of $Y_{lt}$ and $Y_{ut}$ that do not follow
the natural order of an interval, and on the other, it will affect the asymptotic properties
of the estimators as we see next. By taking conditional expectations with respect to
$\mathcal{F}_{t-1}$ in (1.2.1) and imposing (1.2.2), we have

$$E_{t-1} (Y_{lt} | Y_{lt} \leq Y_{ut}) = G_l(Y^{t-1}, X^t; \beta_l) + E_{t-1} (\varepsilon_{lt} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$$

$$E_{t-1} (Y_{ut} | Y_{lt} \leq Y_{ut}) = G_u(Y^{t-1}, X^t; \beta_u) + E_{t-1} (\varepsilon_{ut} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$$

When the observability restriction is binding, the conditional expectations of the errors,
i.e. $E_{t-1} (\varepsilon_{lt} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$ and $E_{t-1} (\varepsilon_{ut} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$, will not be zero
and furthermore, they will depend on the regressors of the model through the functions
Thus, any least-squares estimation (linear or nonlinear) will fail to deliver consistent estimators of the parameters of the model.

Before introducing our estimation procedures, we need to state some basic assumptions on (1.2.1).

**Assumption 1. (Weak Stationarity)**

The interval-valued stochastic process \( \{Y_t\} = \{Y_{lt}, Y_{ut}\} \) is covariance stationary, which means that the lower \( \{Y_{lt}\} \) and upper \( \{Y_{ut}\} \) processes are themselves covariance-stationary.

We also require covariance stationarity in the regressors \( X_t \equiv (X_{1t}, \ldots, X_{pt})' \).

This assumption allows estimators with standard asymptotic properties. However, the methods proposed in the following sections will also apply to non-stationary data but the properties of the estimators will be non-standard.

**Assumption 2. (Exogeneity)**

The regressors \( (Y^{t-1}, X^t) \) are strictly exogenous variables i.e. \( E(\varepsilon_t | Y^{t-1}, X^t) = 0 \)

This assumption is standard in regression analysis to protect the estimators against endogeneity bias. In our context, the objective is to analyze the dynamics of \( \{Y_{ut}\} \) and \( \{Y_{lt}\} \) as a system. For instance, in a VAR system, the right hand side of the system will have lags of \( \{Y_{ut}\} \) and \( \{Y_{lt}\} \). If we were to introduce additional regressors \( X_t \), we could proceed in several ways, either expanding the VAR system to include \( X_t \) as another element of the system, or considering only predetermined regressors, i.e. \( X_{t-1}, X_{t-2}, \ldots \), or requiring the weak exogeneity of \( X_t \). By proceeding in either way, we will focus exclusively on the endogeneity generated by the binding observability restriction, that is, when \( E_{t-1}(\varepsilon_{lt} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u) \) and \( E_{t-1}(\varepsilon_{ut} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u) \), will not be zero.
**Assumption 3.** *(Conditional Independence)*

The regressors \((X_T, ..., X_{t+1})\) and \(Y^t\) are conditional independent given \(X^t\), i.e. \((X_T, ..., X_{t+1}) \perp Y^t \mid X^t\)

This assumption relates to the previous one in the sense that it opens the system of \(\{Y_{ut}\}\) and \(\{Y_{lt}\}\) to the effect of other regressors which are not explicitly modeled within the system. For instance, in a VAR framework, if we were to model jointly \(\{Y_{ut}\}\), \(\{Y_{lt}\}\), and \(X_t\), this assumption will not be needed. But because we focus only on the dynamics of \(\{Y_{ut}\}\) and \(\{Y_{lt}\}\), we need to assume that \(Y_t\) does not Granger-cause \(X_t\) to avoid biased and potentially inconsistent estimators.

**Assumption 4.** *(Normality)*

The error terms \(\epsilon_t = (\epsilon_{lt}, \epsilon_{ut})\) are independent and identically distributed (i.i.d.) bivariate normal random variables

\[
f(\epsilon_t) = (2\pi)^{-1/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} \epsilon_t^\prime \Sigma^{-1} \epsilon_t\right\}
\]

with variance-covariance matrix

\[
\Sigma = \begin{bmatrix}
\sigma_l^2 & \rho \sigma_l \sigma_u \\
\rho \sigma_l \sigma_u & \sigma_u^2 
\end{bmatrix}.
\]

This assumption may seem restrictive but it provides at least a quasi-maximum likelihood approach to the estimation of the system \(\{Y_{ut}\}\) and \(\{Y_{lt}\}\). If the observability restriction is not binding, then the estimation of the system, either by maximum likelihood under normality or by least squares, will produce consistent estimators though
they will not be the most efficient. If heteroscedasticity is present, the estimators are still consistent but we would need to implement a heteroscedasticity-consistent estimator of the variance to produce the right inference. If we were to assume any other density, and again running the risk of a false assumption, we will not be sure whether QMLE results hold, see Newey and Steigerwald (1997). If the observability restriction is binding, bivariate normality implies that the distribution of the errors is conditionally truncated normal with conditional heteroscedasticity. Our estimation procedures takes care of the heteroscedasticity, and since we are modeling extremes, low and high, the density of these variables cannot be symmetric, thus the truncation takes care of the asymmetry. Furthermore, the simulations presented in Section 6 of the paper show that our proposed estimators are very robust to misspecification of the density when there are relevant dynamics in the conditional means of \( \{Y_{ut}\} \) and \( \{Y_{lt}\} \). The potential misspecification of the regressor \( \lambda_{t-1} \) seems to affect mainly the estimation of the constant but we will show that the overall estimation of the system generates very good fitted intervals with substantially smaller losses than other competing methods.

1.3 Estimation

Given the implications of the observability restriction for a least squares estimator of the parameters in (1.2.1), it is natural to think that a full information estimator, like maximum likelihood (ML), will be better suited to guarantee consistency. In this section, we will introduce the conditional log-likelihood function of a sample \( y^T \) in order to underline the contribution of the restriction to the estimation. However, our main objective is to develop a two-step estimation procedure that delivers consistent estimators but it is easier to implement and it overcomes some of the limitations of the ML
1.3.1 Conditional log-likelihood function

For a sample of $T$ observations, $y^T \equiv (y_T, y_{T-1}, \cdots, y_1)$ and $x^T \equiv (x_T, x_{T-1}, \cdots, x_1)$, and for a fixed initial value $y^0$, let $f_Y(y^T | x^T; \theta)$ be the joint conditional density of $y^T$, where the parameter vector $\theta \in \Theta$ is an open subset of $R^K$. The conditional likelihood $l(y, \theta)$ of the sample $y^T$ is such that

$$l(y, \theta) = \begin{cases} 0, & \text{if } y_{lt} > y_{ut} \\ f_{Y^T}(y^T | x^T; \theta) & \text{if } y_{lt} \leq y_{ut}. \end{cases}$$

Then, it follows that

$$l(y, \theta) = f_{Y^T}(y^T | x^T; \theta)$$

$$= f_{Y^T}(y^T | y_t \leq y_{ut}, x^T; \theta) \times \Pr(y_t \leq y_{ut} | x^T; \theta)$$

$$= \prod_{t=1}^T f_{Y_t}(y_t | y_{lt} \leq y_{ut}, y_{t-1}^l, x^t; \theta)$$

$$= \prod_{t=1}^T \frac{f_{Y_t}(y_t | y_{t-1}^l, x^t; \theta)}{\Pr(y_t \leq y_{ut} | y_{t-1}^l, x^t; \theta)}$$

(1.3.4)

where $f_{Y_t}$ is the density of $Y_t$ conditional on the information $(Y_{t-1}^l, X^t)$. In (1.3.4), we have also called assumption 3 regarding the conditional independence of $X^T$ and $Y^{T-1}$ given $x^{T-1}$.

Under the normality assumption 4 and for each observation $t$, the numerator of the
conditional likelihood \( (1.3.4) \) is written as

\[
f_{Y_t}(y_t|y_{t-1}, x^t; \theta) = (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} [y_t - G(y_{t-1}, x^t; \beta)]' \Sigma^{-1} [y_t - G(y_{t-1}, x^t; \beta)] \right\}
\]

where \( \theta = (\beta, \Sigma) \), and the denominator as

\[
R_t(y_{t-1}, x^t; \theta) \equiv \Pr(y_{lt} \leq y_{ut}|y_{t-1}, x^t; \theta) = 1 - \Phi \left( \frac{G_l(y_{t-1}, x^t; \beta_l) - G_u(y_{t-1}, x^t; \beta_u)}{\sqrt{\sigma^2_u + \sigma^2_l - 2 \rho \sigma_u \sigma_l}} \right)
\]

where \( \Phi(\cdot) \) is the standard normal cumulative distribution function.

Consequently, the conditional log-likelihood function of a sample \( y^T \) is written as

\[
L(y, \theta) = -\log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2T} \sum_{t=1}^{T} [y_t - G(y_{t-1}, x^t; \beta)]' \Sigma^{-1} [y_t - G(y_{t-1}, x^t; \beta)]
\]

\[
-\frac{1}{T} \sum_{t=1}^{T} \log R_t(y_{t-1}, x^t; \theta)
\]

(1.3.5)

The maximum likelihood estimator \( \hat{\theta}_{ML} \) is the maximizer of (1.3.5). Observe that this estimator will be highly nonlinear, even for the case of a linear process in (1.2.1), because of the contribution of the observability restriction term \( R_t(y_{t-1}, x^t; \theta) \) to the log-likelihood function. The term \( R_t(y_{t-1}, x^t; \theta) \) provides the probability mass that is left in the joint density after the truncation takes place. It is easily seen that \( 0 \leq R_t(y_{t-1}, x^t; \theta) \leq 1. \)

If the restriction is not binding, \( R_t(y_{t-1}, x^t; \theta) = 1 \) for all \( t \), and its contribution to the
log-likelihood function is zero. In this case the restriction is redundant and it can be removed from the specification of the model. On the other hand, if the observability restriction is binding i.e. $R_t(y^{t-1}, x^t; \theta) < 1$ for some $t$, it must be taken into account in the estimation of the model. Ignoring the restriction will result in the inconsistency of ML estimator. In theory, the ML estimator has obvious advantage. If the true distribution of $\varepsilon_t$ is normal as in assumption 1, under certain regularity conditions, the ML estimator $\hat{\theta}_{ML}$ is consistent and asymptotically normal. However in practice, given the nonlinearity of the ML estimator induced by the observability restriction, we should expect multiple local maxima in the log-likelihood function leading to multiple solutions and non-trivial convergence problems in the maximization algorithm. Therefore, the consistency of ML estimator will depend on a good guess of the initial value of the parameters. For these reasons, we propose an alternative two-step procedure that combines maximum likelihood and least squares estimation, that it is easy to implement and will deliver consistent estimators of the parameters of the model.

1.3.2 Two-Step Estimation: General Remarks

Given the popularity of time series VAR models, we will consider the general process \[(1.2.1)\] to be a linear autoregressive specification of order $p$. However, the two-step procedure to be described next will be also applicable to nonlinear specifications by properly choosing a nonlinear estimation technique in the second step.

1 A sufficient and necessary condition for a non-binding restriction is $\frac{G_l(y^{t-1}, x^t; \beta_l) - G_u(y^{t-1}, x^t; \beta_u)}{\sqrt{\sigma^2 + \sigma^2 - 2\sigma\sigma}} \ll 0$ for all $t$.

2 The regularity conditions that guarantee the consistency and asymptotic normality of ML estimator $\hat{\theta}_{ML}$ are summarized in Theorem 4.1.1, Theorem 4.1.3 in Amemiya (1985) and Theorem 4.6 in White (1994).
The interval autoregressive model, IAR(p), is described as follows

\[
\begin{bmatrix}
y_{l,t} \\
y_{u,t}
\end{bmatrix} = \begin{bmatrix}
\beta_{lc} \\
\beta_{uc}
\end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix}
\beta_{11}^{(j)} & \beta_{12}^{(j)} \\
\beta_{21}^{(j)} & \beta_{22}^{(j)}
\end{bmatrix} \begin{bmatrix}
y_{l,t-j} \\
y_{u,t-j}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{lt} \\
\varepsilon_{ut}
\end{bmatrix}
\]

with observability restriction \( y_{lt} \leq y_{ut} \), and an error term \( \epsilon_t \) that is bivariate normal i.i.d.

Let us call \( \mathcal{I}_{t-1} = (y_{t-1}, ..., y_{t-p}, ...) \) the information set. Then, conditioning on the information set \( \mathcal{I}_{t-1} \), the conditional mean of the IAR(p) process is

\[
E_{t-1}(y_{lt}|y_{ut} \geq y_{lt}) = \beta_{lc} + \sum_{j=1}^{p} \beta_{11}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{12}^{(j)} y_{u,t-j} + E_{t-1}(\varepsilon_{lt}|y_{ut} \geq y_{lt})
\]

\[
E_{t-1}(y_{ut}|y_{ut} \geq y_{lt}) = \beta_{uc} + \sum_{j=1}^{p} \beta_{21}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{22}^{(j)} y_{u,t-j} + E_{t-1}(\varepsilon_{ut}|y_{ut} \geq y_{lt})
\]

Under the normality assumption 4, we easily derive the conditional expectation of the error term (see Appendix A),

\[
E_{t-1}(\varepsilon_{lt}|y_{ut} \geq y_{lt}) = C_l \lambda_{t-1}
\]

\[
E_{t-1}(\varepsilon_{ut}|y_{ut} \geq y_{lt}) = C_u \lambda_{t-1}
\]

where

\[
C_l = -\frac{\sigma_l^2 - \rho \sigma_u \sigma_l}{\sigma_m},
\]

\[
C_u = \frac{\sigma_u^2 - \rho \sigma_u \sigma_l}{\sigma_m}
\]

\[
\sigma_m^2 = \sigma_u^2 + \sigma_l^2 - 2 \rho \sigma_l \sigma_u,
\]
\[ \lambda_{t-1} = \frac{\phi(\Delta(y^{t-1}, \Delta \beta)/\sigma_m)}{1 - \Phi(\Delta(y^{t-1}, \Delta \beta)/\sigma_m)} \]  
\[ (1.3.6) \]

\[ \Delta(y^{t-1}, \Delta \beta) \equiv G_l - G_u = \Delta \beta_c + \sum_{j=1}^{p} \Delta \beta^{(j)}_1 y_{l,t-j} + \sum_{j=1}^{p} \Delta \beta^{(j)}_2 y_{u,t-j} \]

\[ \Delta \beta = (\beta_{lc} - \beta_{uc}, \beta^{(1)}_{11} - \beta^{(1)}_{21}, \beta^{(1)}_{12} - \beta^{(1)}_{22}, \cdots, \beta^{(p)}_{11} - \beta^{(p)}_{21}, \beta^{(p)}_{12} - \beta^{(p)}_{22}) \]  
\[ (1.3.7) \]

Therefore, the regression models can be explicitly written as

\[ y_{lt} = \beta_{lc} + \sum_{j=1}^{p} \beta^{(j)}_{11} y_{l,t-j} + \sum_{j=1}^{p} \beta^{(j)}_{12} y_{u,t-j} + C_l \lambda_{t-1} + v_{lt} \]  
\[ (1.3.8) \]

\[ y_{ut} = \beta_{uc} + \sum_{j=1}^{p} \beta^{(j)}_{21} y_{l,t-j} + \sum_{j=1}^{p} \beta^{(j)}_{22} y_{u,t-j} + C_u \lambda_{t-1} + v_{ut} \]  
\[ (1.3.9) \]

where now \( v_{lt} = \varepsilon_{lt} - C_l \lambda_{t-1} \) and \( v_{ut} = \varepsilon_{ut} - C_u \lambda_{t-1} \) are martingale difference sequences with respect to \( \mathcal{F}_{t-1} \), i.e. \( E_{t-1}(v_{lt}|y_{ut} \geq y_{lt}) = 0 \) and \( E_{t-1}(v_{ut}|y_{ut} \geq y_{lt}) = 0 \).

Two remarks are in order. First, observe that \( C_u - C_l = \sigma_m \), and we need \( \sigma_m > 0 \) to be strictly positive for the constants \( C_u \) and \( C_l \) to be well defined. This implies that the specific case for which \( \sigma_u^2 = \sigma_l^2 \) and \( \rho = 1 \) must be ruled out. This could happen when the interval \([\varepsilon_{lt}, \varepsilon_{ut}]\) is degenerated and collapses to a single value. Secondly, \( \lambda_{t-1} \) is the inverse of the Mill’s ratio and embodies the severity of the observability restriction. When the restriction is non-binding we have seen in the previous section that \( R_t(y^{t-1}, x^t; \theta) = 1 \) for all \( t \), which implies that \( \lambda_{t-1} = 0 \) for all \( t \).

Based on the regressions (1.3.8) and (1.3.9), the two-step estimation strategy consists of estimating \( \lambda_{t-1} \) in the first step, and assessing how binding the observability restriction is. The second step is only meaningful when the restriction is binding. In this case, we proceed to plug in the estimate \( \hat{\lambda}_{t-1} \) in (1.3.8) and (1.3.9) and perform least squares estimation. The proposed two-step estimation strategy resembles Heckman’s (1979) two-
step estimation procedure for sample selection models. However, there are important conceptual differences. In Heckman’s strategy, the selection mechanism (the first step) includes the full sample of observations, e.g. women who participate and who do not in the labor market, and the regression model (the second step) includes a partial sample, those for which the dependent variable of interest is observed, e.g. the wage of those women who work. In our strategy, we carry the same sample in both steps because those observations that violate the observability restriction are impossible to be included in any sample. Hence, from the start, our first step will focus on a truncated normal regression model, which arises very naturally when we model the dynamics of the range of the interval. We will show that from the estimation of this model, we will be able to estimate $\lambda_{t-1}$. Our second step is analogous to Heckman’s in that the objective is to correct the selection bias of the least squares estimator in the regression of interest. However, Heckman’s bias is inconsequential when the error terms of the selection equation and of the regression of interest are uncorrelated. In our second step, even if the error terms of the lower and upper bound regressions are uncorrelated, the bias and inconsistency of the least squares estimator will remain when the observability condition is binding and it is not considered in the second-step regression model.

1.3.3 Two-step Estimation: The First Step

Our objective is to estimate $\lambda_{t-1}$. To this end, we model the range of the interval $\Delta y_t = y_{ut} - y_{lt}$, which according to the IAR(p) model will exhibit the following dynamics

$$y_{ut} - y_{lt} = -\left(\Delta \beta_c + \sum_{j=1}^{p} \Delta \beta_1^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \Delta \beta_2^{(j)} y_{u,t-j} \right) + \Delta \varepsilon_t$$  \hspace{1cm} (1.3.10)
Under the normality assumption 4, and imposing the observability restriction, the difference of the two error terms, $\Delta \varepsilon_t$, follows a truncated normal distribution. Therefore, the conditional density of the range $\Delta y_t$ is,

$$
\begin{align*}
    f(\Delta y_t | \Delta y_t \geq 0, y^{t-1}; \Delta \beta, \sigma_m) &= \frac{f(\Delta y_t | y^{t-1}; \Delta \beta, \sigma_m)}{\Pr(\Delta y_t \geq 0 | y^{t-1}; \Delta \beta, \sigma_m)} \\
    &= \frac{1}{\sigma_m} \frac{\phi(\Delta y_t / \sigma_m + \Delta(y^{t-1}, \Delta \beta) / \sigma_m)}{1 - \Phi(\Delta(y^{t-1}, \Delta \beta) / \sigma_m)}
\end{align*}
$$

(1.3.11)

Based on (1.3.11), we can construct the log-likelihood function of a sample of $T$ observations $\Delta y$

$$
T^{-1} L(\Delta y; \Delta \beta, \sigma_m) = \frac{1}{T} \sum_{t=1}^{T} \log f(\Delta y_t | \Delta y_t \geq 0, y^{t-1}; \Delta \beta, \sigma_m)
$$

(1.3.12)

to obtain the maximum likelihood estimators $\hat{\Delta \beta}$ and $\hat{\sigma}_m$ as the

$$
\arg \max_{\Delta \beta, \sigma_m} [T^{-1} L(\Delta y; \Delta \beta, \sigma_m)].
$$

The ML estimators will be plugged in (1.3.6) to finally obtain $\hat{\lambda}_{t-1}$.

There are two advantages in modeling the range of the interval in the first step. First, the number of parameters to be estimated is reduced from $2(1 + 2p) + 3$ in the full ML estimation (1.3.5) to $1 + 2p + 1$ in (1.3.12). Second, and more importantly, for the truncated normal regression model, there is a unique solution to the maximization problem so that the ML estimator is the global maximizer of the likelihood function. In addition, standard asymptotic properties on the consistency and asymptotic normality of the ML estimators and $\hat{\lambda}_{t-1}$ are easily established. We add the following assumption

**Assumption 5.** (Mixing Conditions)
The interval-valued stochastic process \( \{ Y_t \} = \{ Y_{lt}, Y_{ut} \} \) is either (a) \( \phi \)-mixing of size \(-r/(2r-1), r \geq 1\) or (b) \( \alpha \)-mixing of size \(-r/(r-1)\), such that \( E|Y_{lt}|^{r+\delta} < \Delta < \infty \) and \( E|Y_{ut}|^{r+\delta} < \Delta < \infty \) for some \( \delta > 0 \) for all \( t \).

**Theorem 1.3.1. Consistency and Asymptotic Normality of the first-step ML Estimator**

Let \( \theta^* = (\Delta \beta/\sigma_m, \sigma_m) \equiv (\Delta \beta^*, \sigma_m) \) be a \( 1 \times (2p + 2) \) parameter vector corresponding to model (1.3.10). Under assumptions 1 – 5, the maximum likelihood estimator \( \hat{\theta}^* \) has the following properties,

(a) \( \hat{\theta}^*_ML \) converges to the true value \( \theta^*_0 \) in probability, i.e.

\[
\lim_{T \to \infty} \hat{\theta}^*_ML = \theta^*_0;
\]

(b) \( \hat{\theta}^*_ML \) is asymptotically normally distributed, i.e.

\[
\sqrt{T}(\hat{\theta}^*_ML - \theta^*_0) \overset{d}{\to} N(0, V)
\]

where

\[
V = - \left[ \lim_{T \to \infty} E \left( \frac{\partial^2 L}{\partial \theta^* \partial \theta^{*\prime}} \bigg| \theta^*_0 \right) \right]^{-1}.
\]

The truncated normal regression model has been extensively studied for cross-sectional data. Tobin (1958) was first to consider the problem and he proposed the maximum likelihood estimator. Amemiya (1973) proved strong consistency and asymptotic normality of the ML estimator on a compact parameter space. Orme (1989), Orme and Ruud (2002) further proved that the solution to the likelihood equations for the truncated normal regression is unique and that there is a global maximizer of the log-likelihood function. The proofs of the asymptotic properties in Amemiya (1973) are directly ap-
pllicable to a time series framework by strengthening the moment conditions. With assumption 4, the Kolmogorov’s strong law of large numbers and Liapounov’s central limit theorem for non-identically distributed random variables used in Amemiya (1973), are replaced by McLeish(1974)’s strong law of large numbers (Theorem 2.10) and Wooldridge-White (1988)’s central limit theorem for mixing processes (Corollary 3.1) to guarantee that Theorem 1.3.1 holds.

The asymptotic properties of the estimator of the inverse of the Mill’s ratio follow as a corollary of Theorem 1.3.1 because \( \lambda(\cdot) \) is a continuous and differentiable function with respect to \( \theta^* \).

**Corollary 1. Consistency and Asymptotic Normality of the Inverse of the Mill’s Ratio**

The estimator of the inverse of the Mill’s ratio \( \hat{\Lambda} \equiv (\hat{\lambda}_0, \cdots, \hat{\lambda}_{T-1}) \) has the following properties

(a) \( \lambda(y^t, \hat{\Delta}\beta_{ML}^*) \) converges to the true value \( \lambda(y^t, \Delta\beta_0^*) \), i.e.

\[
\text{plim}_{T \to \infty} \lambda(y^t, \hat{\Delta}\beta_{ML}^*) = \lambda(y^t, \Delta\beta_0^*)
\]

(b) \( \hat{\Lambda} \) is asymptotically normally distributed, i.e.

\[
\sqrt{T}(\hat{\Lambda} - \Lambda) \xrightarrow{d} N(0, S_0)
\]

where

\[
S_0 = J(\Delta\beta_0^*)V_{\Delta\beta_0^*}J(\Delta\beta_0^*)'
\]

and \( V_{\Delta\beta_0^*} \) is the asymptotic variance matrix of \( \sqrt{T}(\hat{\Delta}\beta_{ML}^* - \Delta\beta_0^*) \), a leading prin-
cipal minor of matrix $V$, and the $t$-th row of matrix $J(\Delta \beta_0^*)$ is,

$$j_t = \lambda(y^{t-1}, \Delta \beta_0^*)\lambda(y^{t-1}, \Delta \beta_0^*) - z_{t-1} \Delta \beta_0^* | z_{t-1}.$$

where $z_{t-1} = (1, y_{1,t-1}, y_{2,t-1}, \cdots, y_{l,t-1}, y_{u,t-1}).$

1.3.4 Two-step Estimation: The Second Step

Given the estimate $\lambda_{t-1}$, we plug it in the regression equations (1.3.8) and (1.3.9) to obtain the feasible model

$$y_{lt} = \beta_{lc} + \sum_{j=1}^{p} \beta_{11}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{12}^{(j)} y_{u,t-j} + C_l \lambda_{t-1} + u_{lt}$$

(1.3.14)

$$y_{ut} = \beta_{uc} + \sum_{j=1}^{p} \beta_{21}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{22}^{(j)} y_{u,t-j} + C_u \lambda_{t-1} + u_{ut}$$

(1.3.15)

where now the error term of the feasible regression has two sources of variation, one coming from the $\lambda$ estimator, and the other coming from the error term in the infeasible regression, i.e. $u_{lt} = C_l (\lambda_{t-1} - \lambda_{t-1}) + v_{lt}$ and $u_{ut} = C_u (\lambda_{t-1} - \lambda_{t-1}) + v_{ut}$. As a result, the error term of the feasible regression will be heteroscedastic.

Writing the feasible regressions in matrix form,

$$y_{l} = \hat{H} \gamma_{l} + u_{l}$$

(1.3.16)

$$y_{u} = \hat{H} \gamma_{u} + u_{u}$$

(1.3.17)
where

\[ y_l = (y_{l,1}, \ldots, y_{l,T})' \]
\[ y_u = (y_{u,1}, \ldots, y_{u,T})' \]
\[ \gamma_l = (\beta_l, C_l) \]
\[ \gamma_u = (\beta_u, C_u) \]
\[ \hat{\lambda} = (\hat{\lambda}_0, \ldots, \hat{\lambda}_{T-1})' \]
\[ u_l = C_l(\Lambda - \hat{\Lambda}) + v_l \]
\[ u_u = C_u(\Lambda - \hat{\Lambda}) + v_u \]
\[ \hat{H} = (Z, \hat{\lambda}) \]

The least squares estimators of the parameters \( \gamma_l \) and \( \gamma_u \) are

\[
\hat{\gamma}_l = \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}'y_l \quad (1.3.18)
\]
\[
\hat{\gamma}_u = \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}'y_u \quad (1.3.19)
\]

The following theorem establishes the consistency and asymptotic normality of the two-step estimators \( \hat{\gamma}_l \) and \( \hat{\gamma}_u \).

**Theorem 1.3.2.** Consistency and asymptotic normality of the second step OLS estimator
Under the following assumptions,

(i) $\text{plim}_{T \to \infty} \frac{1}{T} \mathbf{H}' \mathbf{H} = \mathbf{B}^{-1}$, which is nonsingular;

(ii) $\frac{1}{T} \mathbf{H}' \mathbf{J}(\Delta \beta^*)$ converges uniformly in probability to the matrix function $\mathbf{Q}(\Delta \beta^*)$; and $\frac{1}{T} \mathbf{J}'(\Delta \beta^*) \mathbf{J}(\Delta \beta^*)$ is bounded uniformly in probability at least in a neighborhood of the true value $\Delta \beta^*_0$;

(iii) $E|h_{t-1,i}v_t|^2 < \infty$, $E|h_{t-1,i}v_{ul}|^2 < \infty$, and $E|j_{t-1,i}v_{lt}|^2 < \infty$ for $t = 1, \cdots, T$ and $i = 1, \cdots, 2p + 2$;

(iv) $\Psi_{l,T} \equiv \text{var}(T^{-1/2} \mathbf{H}' \mathbf{v}_l) \to \Psi_l$ and $\Psi_{u,T} \equiv \text{var}(T^{-1/2} \mathbf{H}' \mathbf{v}_u) \to \Psi_u$, and $\Psi_l$, $\Psi_u$ are finite and positive definite;

Then, the two-step estimators $\hat{\gamma}_l$ and $\hat{\gamma}_u$

(a) converge to their true values in probability,

(b) with asymptotic normal distributions

\[
\sqrt{T}(\hat{\gamma}_l - \gamma_l) \overset{d}{\to} N(0, \mathbf{B}_l\mathbf{\Xi}_l\mathbf{B}_l'),
\]
\[
\sqrt{T}(\hat{\gamma}_u - \gamma_u) \overset{d}{\to} N(0, \mathbf{B}_u\mathbf{\Xi}_u\mathbf{B}_u').
\]

where

\[
\mathbf{B} = \text{plim}_{T \to \infty} \left( \frac{\mathbf{H}' \mathbf{H}}{T} \right)^{-1} = \text{plim}_{T \to \infty} \left( \frac{\mathbf{H}' \mathbf{H}}{T} \right)^{-1}
\]

\[
\mathbf{\Xi}_l = \Psi_l + C_l^2 \mathbf{Q}_l' \mathbf{S}_0 \mathbf{Q}_0 + \mathbf{M}_l + \mathbf{M}_l'
\]

\[
\mathbf{\Xi}_u = \Psi_u + C_u^2 \mathbf{Q}_u' \mathbf{S}_0 \mathbf{Q}_0 + \mathbf{M}_u + \mathbf{M}_u'
\]
In which

\[
Q_0 = \lim_{T \to \infty} \frac{1}{T} H'J(\Delta \beta_0^*),
\]

\[
S_0 = J(\Delta \beta_0^*)V_{\Delta \beta_0^*}J(\Delta \beta_0^*)',
\]

\[
M_{l0} = \lim_{T \to \infty} \frac{1}{T} E \left( H'v_l(\Lambda - \hat{\Lambda})'HC_l \right),
\]

\[
M_{u0} = \lim_{T \to \infty} \frac{1}{T} E \left( H'v_u(\Lambda - \hat{\Lambda})'HC_u \right).
\]

In equation \([1.3.20]\) and \([1.3.21]\), the first terms \(\Psi_l\) and \(\Psi_u\) are the variance-covariance matrices of the errors \(v_{lt}\) and \(v_{ut}\) respectively, if \(\Lambda\) were observable. The second term \(Q_0' S_0 Q_0\) is capturing the uncertainty induced by the estimates of the inverse of the Mill’s ratio \(\hat{\Lambda}\). The last two terms, \(M_{l0}\) and \(M_{u0}\), capture the covariances between the error terms \(v_{lt}\) and \(v_{ut}\) with \(\hat{\Lambda}\). Although \(v_{lt}\) and \(v_{ut}\) are martingale difference sequences, they are correlated with \(\lambda_{t+i}\) for \(i = 0, 1, \cdots, T - t\). This is a further difference with Heckman’s two-step estimator. In Heckman’s covariance matrix, the matrix \(M_0\) is zero because in a cross-sectional setting the error \(v\) is uncorrelated with the inverse of the Mill’s ratio. Since the asymptotic variance-covariance matrices in \([1.3.20]\) and \([1.3.21]\) capture the heteroscedasticity induced by the observability restriction together with the time dependence induced by \(\hat{\Lambda}\), Newey and West (1994)’s HAC variance-covariance matrix estimator should suffice to estimate \(B\Xi_l B\) and \(B\Xi_u B\) consistently. In addition, we also estimate the unconditional variances \(\sigma_l^2\) and \(\sigma_u^2\) of the respective errors \(\varepsilon_{lt}\) and \(\varepsilon_{ut}\) and their correlation coefficient \(\rho\) by implementing a simple method of moments, which is described in Appendix A.
1.3.5 Two-step Estimation: Implementation Issues

The implementation of the two-step estimator may be subject to multicollinearity, and consequently the parameters $\gamma_l$ and $\gamma_u$ in the second step, equations (1.3.16) and (1.3.17), may not be precisely estimated or, in extreme cases, they may not be identified at all. There are two reasons why multicollinearity arises. First, the functional form (1.3.6) of the inverse of the Mill’s ratio $\lambda(\cdot)$ is approximately linear over a wide range of its argument $\Delta(y^{t-1}, \Delta \beta)/\sigma_m$ (see Figure 1.2) so that the estimated regressor $\hat{\Lambda}$ is almost collinear with the regressors in $Z$. This multicollinearity issues cannot be resolved by just dropping some of the regressors because the inclusion of $\hat{\Lambda}$ is necessary to guarantee the consistency of the estimators $\hat{\beta}_l$ and $\hat{\beta}_u$.

![Figure 1.2: Inverse of the Mill’s Ratio](image)

The second reason pertains to those cases in which the observability condition is not binding. As we have seen in the previous sections, when the observability condition
is not binding, the population value of \( \lambda(\cdot) \) is zero. Within a sample, we will observe values close to zero and a very small variance in \( \hat{\lambda}_t \). The direct consequence is that \( C_l \) and \( C_u \) are not identifiable. In the simulation section, we will discuss several cases in which this problem is severe.

For these two reasons, we propose a modified second step estimator that overcomes the identification problem of \( C_l \) and \( C_u \), and in addition, provides a direct identification of the unconditional variances \( \sigma^2 \) and \( \sigma^2 \) of the respective structural errors \( \varepsilon_{lt} \) and \( \varepsilon_{ut} \) and their correlation coefficient \( \rho \).

### 1.3.6 Two-step Estimation: A Modified Two-step Estimator

The first step of the estimation is identical to the first step explained in section 3.3, from which we obtain the estimates \( \hat{\Lambda} \) and \( \hat{\sigma}_m \). In the second step, we exploit the relationships among \( C_l, C_u, \sigma^2_u, \) and \( \sigma^2 \), that is,

\[
\begin{align*}
C_u + C_l &= \frac{\sigma^2 - \sigma^2}{\hat{\sigma}_m} \\
C_u - C_l &= \sigma_m.
\end{align*}
\]  

(1.3.22)

If \( \sigma^2 \), \( \sigma^2_u \) and \( \sigma_m \) were known, the system of equations (1.3.22) would have a unique solution so that \( C_l \) and \( C_u \) would be uniquely identified. Then, by writing \( \sigma^2_u \) and \( \sigma^2 \) as functions of \( C_l \) and \( C_u \), i.e. \( \sigma^2_u(C_u) \) and \( \sigma^2(C_l) \), we propose the following minimum distance estimator that will permit the identification of \( C_l \) and \( C_u \),

\[
(\tilde{C}_l, \tilde{C}_u) = \arg \min_{(C_l, C_u)} (C_u + C_l - \frac{\sigma^2_u(C_u) - \sigma^2(C_l)}{\hat{\sigma}_m})^2
\]  

(1.3.23)

such that \( C_u - C_l = \hat{\sigma}_m \)
or equivalently

\[(\tilde{C}_l, \tilde{C}_u) = \arg\min_{\{C_l, C_u\}} (C_u^2 - C_l^2 - \sigma_u^2(C_u) + \sigma_l^2(C_l))^2\]  

(1.3.24)

such that \( C_u - C_l = \tilde{\sigma}_m \)

Our first task is to find \( \sigma_u^2(C_u) \) and \( \sigma_l^2(C_l) \). In order to do so, observe that the *unconditional* variance \( \sigma_u^2 \) and \( \sigma_l^2 \) of the error terms \( \varepsilon_{ut} \) and \( \varepsilon_{lt} \) can be written as follows

\[
\sigma_l^2 = \text{var}(\varepsilon_{lt}) = \text{var}(E(\varepsilon_{lt} | \Delta \varepsilon_t \geq \Delta(y_t^{l-1}; \Delta \beta))) + E(\text{var}(\varepsilon_{lt} | \Delta \varepsilon_t \geq \Delta(y_t^{l-1}; \Delta \beta)))
\]

\[
= \text{var}(C_l \lambda(y_t^{l-1}, \beta)) + E(\text{var}(v_{lt}|y_t^{l-1}))
\]

\[
= C_l^2 \text{var}(\lambda_{l-1}) + E(\text{var}(v_{lt}|y_t^{l-1})) \tag{1.3.25}
\]

similarly for \( \sigma_u^2 \)

\[
\sigma_u^2 = C_u^2 \text{var}(\lambda_{t-1}) + E(\text{var}(v_{ut}|y_t^{l-1})) \tag{1.3.26}
\]

and for \( \sigma_m^2 = \text{var}(\Delta \varepsilon_t) \)

\[
\sigma_m^2 = \text{var}(E(\Delta \varepsilon_t | \Delta \varepsilon_t \geq \Delta(y_t^{l-1}; \Delta \beta))) + E(\text{var}(\Delta \varepsilon_t | \Delta \varepsilon_t \geq \Delta(y_t^{l-1}; \Delta \beta)))
\]

\[
= \sigma_m^2 \text{var}(\lambda_{t-1}) + E(\text{var}(\Delta v_t|y_t^{l-1})), \tag{1.3.27}
\]

where

\[
\Delta v_t = v_{ut} - v_{lt} = \Delta y + z_{l-1} \Delta \beta - \sigma_m \lambda_{t-1} \tag{1.3.28}
\]
by subtracting equations [1.3.8] and [1.3.9], and $\Delta \beta$ defined in [1.3.7]. From [1.3.27],

$$\text{var}(\lambda_{t-1}) = 1 - \frac{E(\text{var}(\Delta v_{t}|y^{t-1}))}{\sigma_m^2}.$$  

(1.3.29)

so that all is needed is a consistent estimator of the population moments $E(\text{var}(v_{lt}|y^{t-1}))$, $E(\text{var}(v_{ut}|y^{t-1}))$ and $E(\text{var}(\Delta v_{t}|y^{t-1}))$ to obtain $\sigma_u^2(C_u)$ and $\sigma_l^2(C_l)$ as functions of sample information. The following proposition 1.3.1 guarantees that this is the case.

First, let us call

$$\hat{\Delta} v_t = \Delta y_t + z_{t-1} \bar{\Delta} \beta - \tilde{\sigma}_m \hat{\lambda}_{t-1},$$  

(1.3.30)

$$\hat{u}_{lt} = y_{lt} - z_{t-1} \beta_l(C_l) - C_l \hat{\lambda}_{t-1},$$  

(1.3.31)

$$\hat{u}_{ut} = y_{ut} - z_{t-1} \beta_u(C_u) - C_u \hat{\lambda}_{t-1},$$  

(1.3.32)

where $\bar{\Delta} \beta$ and $\hat{\lambda}_{t-1}$ are the estimates from the first step, and $\beta_l(C_l)$ and $\beta_u(C_u)$ are the concentrated OLS estimates of $\beta$ in [1.3.14] and [1.3.15], i.e.

$$\beta_l(C_l) = (Z'Z)^{-1}Z'(Y_l - C_l \hat{A}),$$  

(1.3.33)

$$\beta_u(C_u) = (Z'Z)^{-1}Z'(Y_u - C_u \hat{A}),$$  

(1.3.34)

**Proposition 1.3.1.** Under assumptions 1 to 5 and for $\phi$- or $\alpha$-mixing sequences $v_{lt}^2$ and $v_{ut}^2$ with at least finite second moments, we have that

$$\frac{1}{T} \sum_{t=1}^{T} \hat{\Delta} v_t^2 \xrightarrow{P} E(\text{var}(\Delta v_{t}|y^{t-1})),$$  

(1.3.35)

$$\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{lt}^2 \xrightarrow{P} E(\text{var}(v_{lt}|y^{t-1})).$$  

(1.3.36)
\[
\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{ut}^2 \xrightarrow{p} E(\text{var}(v_{ut}|y^{t-1})) \tag{1.3.37}
\]

and therefore,

\[
\hat{\sigma}_l^2(C_l) \xrightarrow{p} \sigma_l^2, \tag{1.3.38}
\]
\[
\hat{\sigma}_u^2(C_u) \xrightarrow{p} \sigma_u^2. \tag{1.3.39}
\]

where

\[
\hat{\sigma}_l^2(C_l) \equiv C_l^2 \left(1 - \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{\Delta} v_t^2}{\hat{\sigma}_m^2}\right) + \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{lt}^2,
\]
\[
\hat{\sigma}_u^2(C_u) \equiv C_u^2 \left(1 - \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{\Delta} v_t^2}{\hat{\sigma}_m^2}\right) + \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{ut}^2.
\]

The proof is in Appendix C.

The implementation of the minimum distance estimator in (1.3.23) is described in Figure 1.3.

We proceed as follows:

1. pick any point \((C_l^*, C_u^*)\) on the line \(C_u = \hat{\sigma}_m + C_l\);
2. compute the corresponding concentrated \(\beta_l(C_l^*)\) and \(\beta_u(C_u^*)\) as in (1.3.33) and (1.3.34);
3. compute the corresponding residuals \(\hat{u}_{lt}, \hat{u}_{ut},\) and \(\hat{\Delta} v_t\) as in (1.3.31), (1.3.32), and (1.3.30) respectively;
4. calculate the intercept \((\sigma_u^2(C_u^*) - \sigma_l^2(C_l^*) / \hat{\sigma}_m\) to obtain the point \((C_l^*, C_u^{**})\) on the
Given the optimal solution \((\tilde{C}_l, \tilde{C}_u)\), the estimators of the parameters \(\beta\) of the original model are readily available as well as the variance-covariance matrix of the errors \(\varepsilon_{lt}\) and \(\varepsilon_{ut}\), i.e.

\[
\tilde{\beta}_l = \beta_l(\tilde{C}_l) = (Z'Z)^{-1}Z'(Y_l - \tilde{C}_l\hat{\Lambda}) \tag{1.3.41}
\]

\[
\tilde{\beta}_u = \beta_u(\tilde{C}_u) = (Z'Z)^{-1}Z'(Y_u - \tilde{C}_u\hat{\Lambda}) \tag{1.3.42}
\]
\[ \tilde{\sigma}_l^2 = \sigma_l^2(C_l) \quad (1.3.43) \]
\[ \tilde{\sigma}_u^2 = \sigma_u^2(C_u) \quad (1.3.44) \]
\[ \tilde{\rho} = \frac{\tilde{\sigma}_m^2 - \tilde{\sigma}_l^2 - \tilde{\sigma}_u^2}{-2\tilde{\sigma}_l\tilde{\sigma}_u}. \quad (1.3.45) \]

**Theorem 1.3.3. Consistency of Modified Two-step Estimator**

The modified two-step estimator \((\tilde{C}_l, \tilde{C}_u)\) and those defined in (1.3.41) – (1.3.45) converge in probability to the true values of the parameters.

In order to prove Theorem 1.3.3 which states the consistency of estimates \(\tilde{C}_l\) and \(\tilde{C}_u\) in (1.3.23) or (1.3.24), we only need to verify the assumptions stated in Theorem 7.3.2 in Mittelhammer et. al. (2000) that guarantee the consistency of extremum estimators.\(^3\)

Proposition 1.3.1 guarantees that the restricted objective function in (1.3.24) converges in probability to that provided in (1.3.22). In addition, since the system of equations (1.3.22) has a unique solution and the restricted objective function (1.3.24) is a continuous and convex function in \(C_l\) and \(C_u\), it is uniquely minimized at the true values of \(C_l\) and \(C_u\).

### 1.4 Simulation

We perform Monte Carlo simulations to assess the finite sample performance of the two proposed estimation strategies: the two-step estimator and the modified two-step estimator; and compare these estimators with a naive OLS estimator, which does not take into account the relevance of the observability restriction.

\(^3\)See Newey and MacFadden (1994, pp. 2133-34) for the proof. The four assumptions are (a) \(m(\theta, Y, X)\) converges uniformly in probability to a function of \(\theta\), say \(m_0(\theta)\); (b) \(m_0(\theta)\) is continuous in \(\theta\); (c) \(m_0(\theta)\) is uniquely maximized at the true value \(\theta_0\); and (d) the parameter space \(\Omega\) is compact.
The data generating process (DGP) is specified as an IAR(1)

\[
\begin{bmatrix}
  y_{l,t} \\
  y_{u,t}
\end{bmatrix}
= \begin{bmatrix}
  \beta_{lc} \\
  \beta_{uc}
\end{bmatrix}
+ \begin{bmatrix}
  \beta_{11} & \beta_{12} \\
  \beta_{21} & \beta_{22}
\end{bmatrix}
\begin{bmatrix}
  y_{l,t-1} \\
  y_{u,t-1}
\end{bmatrix}
+ \begin{bmatrix}
  \varepsilon_{l,t} \\
  \varepsilon_{u,t}
\end{bmatrix}
\tag{1.4.46}
\]

such that

\[ y_{u,t} \geq y_{l,t} \]

and with an error term that is bivariate normally distributed

\[ \varepsilon_t \equiv (\varepsilon_{l,t}, \varepsilon_{u,t})' \sim N\left(0, \begin{pmatrix}
  \sigma^2_l & \rho \sigma_l \sigma_u \\
  \rho \sigma_l \sigma_u & \sigma^2_u
\end{pmatrix}\right). \]

The interval time series \{[y_{l,t}, y_{u,t}]\}_{t=1}^T is generated sequentially to guarantee that the bounds are not crossing each other i.e. \( y_{l,t} > y_{u,t} \). We proceed as follows. Given the interval datum \([y_{l,t-1}, y_{u,t-1}]\) at time \( t - 1 \), we draw error terms \( \varepsilon_t = (\varepsilon_{l,t}, \varepsilon_{u,t}) \) from the bivariate normal density and calculate \([y_{l,t}, y_{u,t}]\) for time \( t \). If a cross-over happens (i.e. \( y_{l,t} > y_{u,t} \)), we draw another pair of error terms until the observability restriction \( y_{l,t} \leq y_{u,t} \) is met. In doing so, we guarantee that the error terms \( \varepsilon_t \) are truncated bivariate normally distributed, and that the truncation varies across time because it will depend on the past interval-valued data \([y_{l-1}]\) as well as on the assumed parameters \( \beta \)'s in the IAR(1) DGP.

We have designed eight different specifications that are presented in Table 1.1.
Table 1.1: Specification of Data Generating Process

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Binding Cases</th>
<th>Non-binding Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-1</td>
<td>B-2</td>
</tr>
<tr>
<td>$\beta_{lc}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_{uc}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>$C_l$</td>
<td>$-1/\sqrt{2}$</td>
<td>$-1.4564$</td>
</tr>
<tr>
<td>$C_u$</td>
<td>$1/\sqrt{2}$</td>
<td>$-0.3479$</td>
</tr>
<tr>
<td>$\sigma^2_l$</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma^2_u$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Number of Simulation</td>
<td>1000</td>
<td></td>
</tr>
</tbody>
</table>
We have simulated a block of four DGPs where the observability restriction is binding and another block of four DGPs where it is not. Since the observability restriction for the IAR(1) model implies that

\[
\frac{\Delta \epsilon_t}{\sigma_m} \geq \Delta (y_{t-1}, \theta^*),
\tag{1.4.47}
\]

the right hand side of the inequality \eqref{1.4.47} will determine whether the observability restriction is binding or not. We guarantee that the observability restriction is not binding when

\[
\Delta (y_{t-1}, \theta^*) = \Delta \beta^*_c + \Delta \beta^*_1 y_{l,t-1} + \Delta \beta^*_2 y_{u,t-1} \ll 0
\]

otherwise, the restriction could be mildly or severely binding depending upon the choices of the parameters of the DGP. In our simulations, we fix the parameters in \(\Delta \beta^*_1\) and \(\Delta \beta^*_2\) and play with the intercept \(\Delta \beta^*_c\) to allow the restriction to be binding or not. For the four cases, B-1 to B-4, \(\beta_{lc} - \beta_{uc} = 0\), so that the observability restriction becomes binding; and for the four cases, NB-1 to NB-4, \(\beta_{lc} - \beta_{uc} = -4\), so that the restriction is not severely binding. Within each block, we simulate two IAR(1) DGPs, one with high persistence and another with low persistence; and for each one we assume two different variance-covariance matrix \(\Sigma\) for the errors, one with uncorrelated errors and another with highly correlated errors. For each DGP, we also run small and large sample experiments \((T = 250\text{ and } 2000)\) with 1000 replications per DGP.

We report our results in Tables \[1.2 - 1.9\].
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size (T = 250)</th>
<th>Large Sample Size (T = 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = 0$</td>
<td>$-0.4957$</td>
<td>$0.5061$</td>
</tr>
<tr>
<td>$\beta_{uc} = 0$</td>
<td>$0.4979$</td>
<td>$0.5087$</td>
</tr>
<tr>
<td>$\beta_{11} = 0.8$</td>
<td>$0.6645$</td>
<td>$0.1450$</td>
</tr>
<tr>
<td>$\beta_{12} = 0.1$</td>
<td>$0.2204$</td>
<td>$0.1316$</td>
</tr>
<tr>
<td>$\beta_{21} = 0.1$</td>
<td>$0.2191$</td>
<td>$0.1317$</td>
</tr>
<tr>
<td>$\beta_{22} = 0.8$</td>
<td>$0.6637$</td>
<td>$0.1468$</td>
</tr>
<tr>
<td>$C_l = -0.7071$</td>
<td>$-0.5086$</td>
<td>$1.1124$</td>
</tr>
<tr>
<td>$C_u = 0.7071$</td>
<td>$0.6050$</td>
<td>$1.0984$</td>
</tr>
<tr>
<td>$\sigma^2_l = 1$</td>
<td>$0.7918$</td>
<td>$0.2207$</td>
</tr>
<tr>
<td>$\sigma_u^2 = 1$</td>
<td>$0.7891$</td>
<td>$0.2232$</td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>$0.2575$</td>
<td>$0.2653$</td>
</tr>
</tbody>
</table>

Number of Simulation=1000, $1/\sqrt{2} = 0.7071$
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size ($T = 250$)</th>
<th>Large Sample Size ($T = 2000$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
</tr>
<tr>
<td>$\beta_{lc} = 0$</td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{uc} = 0$</td>
<td>-1.0645</td>
<td>1.0813</td>
</tr>
<tr>
<td>$\beta_{11} = 0.8$</td>
<td>0.4648</td>
<td>0.3531</td>
</tr>
<tr>
<td>$\beta_{12} = 0.1$</td>
<td>0.4188</td>
<td>0.3430</td>
</tr>
<tr>
<td>$\beta_{21} = 0.1$</td>
<td>0.0253</td>
<td>0.1000</td>
</tr>
<tr>
<td>$\beta_{22} = 0.8$</td>
<td>0.8602</td>
<td>0.0980</td>
</tr>
<tr>
<td>$C_l = -1.4564$</td>
<td>-1.3899</td>
<td>1.8628</td>
</tr>
<tr>
<td>$C_u = -0.3479$</td>
<td>-0.3609</td>
<td>1.3450</td>
</tr>
<tr>
<td>$\sigma^2_l = 3$</td>
<td>2.1280</td>
<td>0.8944</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>0.9424</td>
<td>0.1009</td>
</tr>
<tr>
<td>$\rho = 0.8$</td>
<td>0.8268</td>
<td>0.0336</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size (T = 250)</th>
<th>Large Sample Size (T = 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = 0$</td>
<td>-0.5715</td>
<td>0.5782</td>
</tr>
<tr>
<td>$\beta_{uc} = 0$</td>
<td>0.5642</td>
<td>0.5711</td>
</tr>
<tr>
<td>$\beta_{11} = 0.1$</td>
<td>0.0764</td>
<td>0.0744</td>
</tr>
<tr>
<td>$\beta_{12} = 0.05$</td>
<td>0.0663</td>
<td>0.0745</td>
</tr>
<tr>
<td>$\beta_{21} = 0.05$</td>
<td>0.0646</td>
<td>0.0733</td>
</tr>
<tr>
<td>$\beta_{22} = 0.1$</td>
<td>0.0782</td>
<td>0.0750</td>
</tr>
<tr>
<td>$C_l = -1/\sqrt{2}$</td>
<td>-3.0891</td>
<td>230.0</td>
</tr>
<tr>
<td>$C_u = 1/\sqrt{2}$</td>
<td>1.3011</td>
<td>182.6</td>
</tr>
<tr>
<td>$\sigma^2_l = 1$</td>
<td>0.6806</td>
<td>0.3251</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>0.6794</td>
<td>0.3262</td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>0.4579</td>
<td>0.4613</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size ($T = 250$)</th>
<th>Large Sample Size ($T = 2000$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = 0$</td>
<td>$-1.1702$</td>
<td>$1.1777$</td>
</tr>
<tr>
<td>$\beta_{uc} = 0$</td>
<td>$-0.2769$</td>
<td>$0.2948$</td>
</tr>
<tr>
<td>$\beta_{11} = 0$</td>
<td>$0.0498$</td>
<td>$0.1378$</td>
</tr>
<tr>
<td>$\beta_{12} = 0.05$</td>
<td>$0.0931$</td>
<td>$0.1799$</td>
</tr>
<tr>
<td>$\beta_{21} = 0.05$</td>
<td>$0.0399$</td>
<td>$0.0949$</td>
</tr>
<tr>
<td>$\beta_{22} = 0.1$</td>
<td>$0.1051$</td>
<td>$0.1260$</td>
</tr>
<tr>
<td>$C_l = -1.4564$</td>
<td>$124.2$</td>
<td>$7062$</td>
</tr>
<tr>
<td>$C_u = -0.3479$</td>
<td>$15.77$</td>
<td>$4069$</td>
</tr>
<tr>
<td>$\sigma^2_l = 3$</td>
<td>$1.6612$</td>
<td>$1.3476$</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>$0.9213$</td>
<td>$0.1150$</td>
</tr>
<tr>
<td>$\rho = 0.8$</td>
<td>$0.8607$</td>
<td>$0.0631$</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
Table 1.6: Simulation Results for Case NB-1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
<td>mean</td>
<td>rmse</td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = -2$</td>
<td>-2.0890</td>
<td>0.4074</td>
<td>-2.1056</td>
<td>0.4345</td>
<td>-2.0883</td>
<td>0.4076</td>
</tr>
<tr>
<td>$\beta_{uc} = 2$</td>
<td>2.0825</td>
<td>0.3816</td>
<td>2.0907</td>
<td>0.4118</td>
<td>2.0819</td>
<td>0.3819</td>
</tr>
<tr>
<td>$\beta_{11} = 0.8$</td>
<td>0.7842</td>
<td>0.0401</td>
<td>0.7830</td>
<td>0.0417</td>
<td>0.7843</td>
<td>0.0401</td>
</tr>
<tr>
<td>$\beta_{12} = 0.1$</td>
<td>0.0975</td>
<td>0.0385</td>
<td>0.0988</td>
<td>0.0402</td>
<td>0.0975</td>
<td>0.0385</td>
</tr>
<tr>
<td>$\beta_{21} = 0.1$</td>
<td>0.0983</td>
<td>0.0363</td>
<td>0.0989</td>
<td>0.0376</td>
<td>0.0983</td>
<td>0.0363</td>
</tr>
<tr>
<td>$\beta_{22} = 0.8$</td>
<td>0.7850</td>
<td>0.0390</td>
<td>0.7844</td>
<td>0.0410</td>
<td>0.7850</td>
<td>0.0390</td>
</tr>
<tr>
<td>$C_l = -1/\sqrt{2}$</td>
<td>-161.5</td>
<td>9243</td>
<td>-0.7026</td>
<td>0.0553</td>
<td>-0.7026</td>
<td>0.0553</td>
</tr>
<tr>
<td>$C_u = 1/\sqrt{2}$</td>
<td>-151.2</td>
<td>4418</td>
<td>0.7016</td>
<td>0.0541</td>
<td>0.7016</td>
<td>0.0541</td>
</tr>
<tr>
<td>$\sigma^2_l = 1$</td>
<td>0.9921</td>
<td>0.0899</td>
<td>0.9842</td>
<td>0.0906</td>
<td>0.9823</td>
<td>0.0903</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>0.9906</td>
<td>0.0885</td>
<td>0.9830</td>
<td>0.0893</td>
<td>0.9807</td>
<td>0.0891</td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>-0.0004</td>
<td>0.0630</td>
<td>-0.0001</td>
<td>0.0632</td>
<td>-0.0066</td>
<td>0.0642</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
Table 1.7: Simulation Results for Case NB-2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size $(T = 250)$</th>
<th>Large Sample Size $(T = 2000)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = -2$</td>
<td>-2.0208</td>
<td>0.9077</td>
</tr>
<tr>
<td>$\beta_{uc} = 2$</td>
<td>2.0813</td>
<td>0.5217</td>
</tr>
<tr>
<td>$\beta_{11} = 0.8$</td>
<td>0.7869</td>
<td>0.0632</td>
</tr>
<tr>
<td>$\beta_{12} = 0.1$</td>
<td>0.0910</td>
<td>0.0824</td>
</tr>
<tr>
<td>$\beta_{21} = 0.1$</td>
<td>0.0985</td>
<td>0.0351</td>
</tr>
<tr>
<td>$\beta_{22} = 0.8$</td>
<td>0.7869</td>
<td>0.0486</td>
</tr>
<tr>
<td>$C_l = -1.4564$</td>
<td>3.94E4</td>
<td>1.12E6</td>
</tr>
<tr>
<td>$C_u = -0.3479$</td>
<td>-2.70E4</td>
<td>6.08E5</td>
</tr>
<tr>
<td>$\sigma_l^2 = 3$</td>
<td>2.9536</td>
<td>0.2745</td>
</tr>
<tr>
<td>$\sigma_u^2 = 1$</td>
<td>0.9871</td>
<td>0.0911</td>
</tr>
<tr>
<td>$\rho = 0.8$</td>
<td>0.7987</td>
<td>0.0227</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
Table 1.8: Simulation Results for Case NB-3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Sample Size (T = 250)</th>
<th></th>
<th>Large Sample Size (T = 2000)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Two-step</td>
<td>Modified Two-step</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = -2$</td>
<td>-2.0165</td>
<td>0.1917</td>
<td>-1.7197</td>
<td>54.2693</td>
</tr>
<tr>
<td>$\beta_{uc} = 2$</td>
<td>2.0246</td>
<td>0.2021</td>
<td>2.6687</td>
<td>63.9572</td>
</tr>
<tr>
<td>$\beta_{11} = 0.1$</td>
<td>0.0945</td>
<td>0.0602</td>
<td>0.1137</td>
<td>0.6083</td>
</tr>
<tr>
<td>$\beta_{12} = 0.05$</td>
<td>0.0517</td>
<td>0.0650</td>
<td>0.0402</td>
<td>0.7072</td>
</tr>
<tr>
<td>$\beta_{21} = 0.05$</td>
<td>0.0498</td>
<td>0.0656</td>
<td>0.0385</td>
<td>0.8010</td>
</tr>
<tr>
<td>$\beta_{22} = 0.1$</td>
<td>0.0897</td>
<td>0.0649</td>
<td>0.0779</td>
<td>0.6570</td>
</tr>
<tr>
<td>$C_l = -1/\sqrt{2}$</td>
<td>225.5</td>
<td>12212</td>
<td>-0.7022</td>
<td>0.0548</td>
</tr>
<tr>
<td>$C_u = 1/\sqrt{2}$</td>
<td>-185.8</td>
<td>15901</td>
<td>0.7038</td>
<td>0.0555</td>
</tr>
<tr>
<td>$\sigma^2 = 1$</td>
<td>0.9843</td>
<td>0.0884</td>
<td>0.9839</td>
<td>0.0902</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>0.9866</td>
<td>0.0881</td>
<td>0.9865</td>
<td>0.0901</td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>0.0063</td>
<td>0.0598</td>
<td>-0.0014</td>
<td>0.0614</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
Table 1.9: Simulation Results for Case NB-4

<table>
<thead>
<tr>
<th>Parameters</th>
<th>OLS Small Sample Size ($T = 250$)</th>
<th>OLS Large Sample Size ($T = 2000$)</th>
<th>Two-step Modified Two-step</th>
<th>Two-step Modified Two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>rmse</td>
<td>mean</td>
<td>rmse</td>
</tr>
<tr>
<td>$\beta_{lc} = -2$</td>
<td>-2.0219</td>
<td>0.5431</td>
<td>-1.6596</td>
<td>14.0955</td>
</tr>
<tr>
<td>$\beta_{uc} = 2$</td>
<td>2.0008</td>
<td>0.3217</td>
<td>2.0827</td>
<td>7.5863</td>
</tr>
<tr>
<td>$\beta_{11} = 0.1$</td>
<td>0.0938</td>
<td>0.0996</td>
<td>0.0919</td>
<td>0.5700</td>
</tr>
<tr>
<td>$\beta_{12} = 0.05$</td>
<td>0.0542</td>
<td>0.1686</td>
<td>0.0474</td>
<td>0.9405</td>
</tr>
<tr>
<td>$\beta_{21} = 0.05$</td>
<td>0.0488</td>
<td>0.0582</td>
<td>0.0483</td>
<td>0.3187</td>
</tr>
<tr>
<td>$\beta_{22} = 0.1$</td>
<td>0.0978</td>
<td>0.0997</td>
<td>0.1015</td>
<td>0.5140</td>
</tr>
<tr>
<td>$C_l = -1.4564$</td>
<td>-98.20</td>
<td>64189</td>
<td>-1.4348</td>
<td>0.0907</td>
</tr>
<tr>
<td>$C_u = -0.3479$</td>
<td>160.6</td>
<td>34474</td>
<td>-0.3325</td>
<td>0.633</td>
</tr>
<tr>
<td>$\sigma^2_l = 3$</td>
<td>2.9703</td>
<td>0.2742</td>
<td>2.9494</td>
<td>0.2771</td>
</tr>
<tr>
<td>$\sigma^2_u = 1$</td>
<td>0.9886</td>
<td>0.0924</td>
<td>0.9807</td>
<td>0.0932</td>
</tr>
<tr>
<td>$\rho = 0.8$</td>
<td>0.7982</td>
<td>0.0236</td>
<td>0.7981</td>
<td>0.0238</td>
</tr>
</tbody>
</table>

Number of Simulation=1000
These are our findings:

1. When the observability restriction is binding (Cases B-1 to B-4), the mean values of the OLS estimates are quite far from the true values, as we expected. OLS estimators are not consistent due to the correlation of the regressors with the error terms. When the observability restriction is not severely binding (Cases NB-1 to NB-4), the mean values of the OLS estimates are very close to the true values. Since the observability restriction is not severely binding, $\lambda_{t-1}$ is very close to zero, so that the endogeneity problem does not arise.

2. When we implement the two-step estimation, the main issue that we face is identification of the model whether or not the restriction is binding. If the restriction is binding but $\lambda_{t-1}$ is almost linear in the regressors of the model, multicollinearity arises (Cases B-3 and B-4). The problem is more severe when there is low persistence in the model and the errors are correlated (Case B-4). Only when $\lambda_{t-1}$ exhibits substantial variation (Cases B-1 and B-2), we do not face a problem with the identification of the model and the mean values of the two-step estimates are very close to the true values. If the restriction is not binding, we expect severe multicollinearity. In Cases NB-1 to NB-4, the RMSE’s of $\hat{C}_l$ and $\hat{C}_u$ explode regardless of the persistence of the model and the sample size. When there is low persistence in the model (Cases NB-3 and NB-4), the RMSE’s of $\hat{\beta}_{l,c}$ and $\hat{\beta}_{u,c}$ also explode because the nearly-zero regressor $\lambda_{t-1}$ is highly collinear with the constant terms.

3. Modified two-step estimation resolves very nicely the identification problem whether the observability restriction is binding or not. If it is binding (Cases B-1 to B-4), the estimators are consistent whether there is low or high persistence and whether the errors
are or not correlated. The modified two-step estimates are very close to the true values and their standard errors are smaller than those of the two-step estimates, even in those cases where the model is well-identified (Cases B-1 and B-2). If the restriction is not binding (Cases NB-1 to NB-4) and thus redundant, the OLS estimator is consistent and efficient but the modified two-step estimator does not seem to be less efficient as the RMSE’s of the modified two-step estimates are very close to those of the OLS estimates.

In practice, we do not know \textit{a priori} whether the restriction is binding or not. In the first step, we assess the severity of the restriction by testing whether or not $\lambda_t = 0$. However, in the second step, we gather further information about the value of the restriction because when it is binding, the OLS estimates should be substantially different from the two-step estimates. In addition, the regressor $\hat{\lambda}_{t-1}$ should be statistically significant. Since multicollinearity affects the significance of $\hat{\lambda}_{t-1}$, we strongly recommend running the modified two-step estimator and assessing the differences with the OLS estimator.

1.5 The SP500 Low/High Return Interval

We model the interval time series of the daily low/high returns to the SP500 index. The returns are computed with respect to the closing price of the previous day, that is,

\begin{align*}
  r_{ht} &= \frac{P_{\text{high},t} - P_{\text{close},t-1}}{P_{\text{close},t-1}} \\
  r_{lt} &= \frac{P_{\text{low},t} - P_{\text{close},t-1}}{P_{\text{close},t-1}},
\end{align*}
where $P_{\text{high},t}$ and $P_{\text{low},t}$ are the highest and lowest price in the trading day $t$, and $P_{\text{close},t-1}$ is the closing price in the previous day $t-1$. Our sample runs from January 1st, 2004 to April 29th, 2011. We have split the sample into two periods that have very different dynamics so that we can showcase the role of the observability condition in the modeling exercise. The first period ranges from January 1st, 2004 to January 1st, 2007; we call it the 'stable period’ because is characterized by very low volatility. In contrast, the second period that goes from January 1st, 2007 to April 29th, 2011 is the ‘unstable period’ because of the high volatility associated with the great panic of the 2008 financial crisis. For both periods, we plot the time series of low/high returns interval in Figure 1.4 and Figure 1.5 and we report their descriptive statistics in Tables 1.10 and 1.11.
In the stable period, both low and high returns exhibit low volatilities (\( \sigma_l^2 = 0.1726 \) and \( \sigma_u^2 = 0.1609 \)), varying within a range of \(-2\% \sim 2\%\), whereas in the unstable period, the two time series vary within a wider range of \(-5\% \sim 5\%\), and exceptionally, in the last months of 2008, moving within a range of \(-10\% \sim 10\%\), thus producing a
Figure 1.5: Daily SP500 Low/High Returns: Unstable Period (2007/1/1-2011/4/29)

Table 1.11: Descriptive Statistics for Unstable Period

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Low Return</th>
<th>High Return</th>
<th>Center</th>
<th>Radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>−9.4210</td>
<td>−1.3010</td>
<td>−4.9190</td>
<td>0.1463</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>−1.3330</td>
<td>0.1769</td>
<td>−0.4698</td>
<td>0.4671</td>
</tr>
<tr>
<td>Median</td>
<td>−0.6204</td>
<td>0.5889</td>
<td>−0.0127</td>
<td>0.7106</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>−0.1498</td>
<td>1.1620</td>
<td>0.4170</td>
<td>1.1040</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.5050</td>
<td>11.9800</td>
<td>6.7410</td>
<td>5.6090</td>
</tr>
<tr>
<td>Mean</td>
<td>−0.9593</td>
<td>0.8593</td>
<td>−0.0500</td>
<td>0.9093</td>
</tr>
<tr>
<td>Variance</td>
<td>1.6539</td>
<td>1.3347</td>
<td>0.9687</td>
<td>0.5256</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.2982</td>
<td>−0.1118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>−2.6056</td>
<td>3.1601</td>
<td>−0.0715</td>
<td>2.7765</td>
</tr>
</tbody>
</table>

48
much higher volatile environment \((\sigma^2_l = 1.6539 \text{ and } \sigma^2_u = 1.3347)\). The stable period returns are characterized by much lower skewness and kurtosis than those in the unstable period. In addition, the correlation of low and high returns is 0.5797 in the stable period, which is larger than the correlation of 0.2982 in the unstable period. We also report the descriptive statistics of the center and radius of the intervals. In both periods, the center return has basically a zero mean, but the variance of the radius in the stable period is very small (0.0351) compared to that (0.5256) of the unstable period. In summary, the unstable period is dominated by a tremendous volatility shock, which is not present in the stable period.

In order to start with the estimation of the system, we need to choose the number of lags for the IAR\((p)\). We run an unrestricted IAR system and we proceed by selecting the optimal lag based on minimizing the BIC. In the stable period, the optimal number of lags is 2, and in the unstable period is 5.

We implement the first-step of the estimation by modeling the range of the interval time series \(\Delta r_t = r_u t - r_l t\) as in (1.3.10). By maximizing the log-likelihood function based on a truncated normal density (1.3.11), we obtain the estimates \(\hat{\theta}^*\) that are reported in Tables 1.12 and 1.13 for the stable and unstable periods respectively.

Table 1.12: First Step Estimation for Stable Period

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient ((\Delta \beta^*))</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{const})</td>
<td>-1.7655 (0.1407)</td>
<td></td>
</tr>
<tr>
<td>(r_{L,t-1})</td>
<td>0.4315 (0.1116)</td>
<td></td>
</tr>
<tr>
<td>(r_{U,t-1})</td>
<td>0.0201 (0.1172)</td>
<td></td>
</tr>
<tr>
<td>(r_{L,t-2})</td>
<td>0.7010 (0.1126)</td>
<td></td>
</tr>
<tr>
<td>(r_{U,t-2})</td>
<td>-0.2849 (0.1150)</td>
<td></td>
</tr>
<tr>
<td>(\sigma_m)</td>
<td>0.3699 (0.0107)</td>
<td></td>
</tr>
</tbody>
</table>

Time Span: 2004/1/1 – 2007/1/1
Number of Observations: 756
It is interesting to observe that the correlation between the range and lagged lower-bound returns is negative \(^4\), while the correlation between the range and lagged upper-bound returns is positive; however the magnitude of the effect of the lower-bound returns is dominant, which implies that, on average, the range will narrow when there is an upward movement in both bounds.

Based on the estimates \( \hat{\theta} \), we produce an estimate of the inverse of Mill’s ratio \( \lambda_{t-1} \), which characterizes the severity of the observability restriction. We plot the estimated time series \( \hat{\lambda}_t \) in Figures 1.6 and Figure 1.7 together with a 95% confidence interval.

In the stable period, the values of \( \hat{\lambda}_t \) are very small, between 0 and 0.070; the mean is 0.027 and the standard deviation 0.014. This indicates that \( \lambda_t \) is practically zero and consequently the observability restriction is not binding. In contrast, in the unstable period, the values of \( \hat{\lambda}_t \) oscillate between 0 and 0.684 with mean 0.213 and standard

\(^4\)The estimates in Tables 1.2 and 1.3 need to be multiplied by \(-1\), see expression (1.3.10).
Figure 1.6: Estimated Inverse Mill’s Ratio: Stable Period (2004/1/1-2007/1/1)

Figure 1.7: Estimated Inverse Mill’s Ratio: Unstable Period (2007/1/1-2011/4/29)
deviation 0.181; these values imply that the relevant portion of the function $\hat{\lambda}_t$ is not entirely linear in the regressors, see Figure 2. In the unstable period, there are a few regions where $\hat{\lambda}_t$ is very close to zero; this happens mainly in the highly volatile period of the end of 2008, when the range of the interval is very large, so that the observability restriction is much less binding than in the rest of the sample.

With the estimated inverse of the Mill's ratio $\hat{\lambda}_{t-1}$, we are able to implement the second step of the estimation. We calculate the second-step estimator by running the feasible regressions (1.3.14) and (1.3.15), and the modified second-step minimum distance estimator by solving the problem in (1.3.24).

We also implement a stationary block bootstrap procedure to obtain the standard errors of the modified second-step estimator. The analytical expression of the standard errors will be difficult to obtain as we carry three sources of uncertainty, i.e. the estimates $\hat{\Lambda}$ and $\hat{\sigma}_m$ in the first step, the estimates $(\tilde{C}_l, \tilde{C}_u)$ in the modified second step, and the idiosyncratic uncertainty of the errors in the IAR system. Thus, we resort to the bootstrap as this is a common practice to overcome the difficulties of the estimation of asymptotic variances in various contexts, see Efron (1979), Buchinsky (1995), and Ledoit, Santa-Clara, and Wolf (2003), among others. For the consistency of the bootstrap estimator for dependent data see Goncalves and White (2005). In the stationary block bootstrap, the block size follows a Geometric distribution with mean equal to $b$. To choose the optimal block size $b$, we follow the method proposed by Politis and White (2004) and Patton, Politis and White (2009). The optimal value of $b$ minimizes the $MSE(\hat{\sigma}_b^2)$ with $\sigma^2_\infty = \sum_{s=-\infty}^{\infty} R(s)$, where $R(s)$ is the auto-covariance function. This procedure considers only the bootstrapping for a scalar time series, however with interval time series we need to jointly bootstrap a $2 \times 1$ vector time series $\{(y_{lt}, y_{ut})\}_{l=0}^{T}$. We proceed by
selecting separately the optimal block sizes $b_l$ and $b_u$ for the lower bound $\{y_{lt}\}_{t=0}^T$ and
the upper bound $\{y_{ut}\}_{t=0}^T$ series respectively. Then, we use the average $(b_l + b_u)/2$ as
the unified block size length to bootstrap the vector sequence $\{(y_{lt}, y_{ut})\}_{t=0}^T$. We report
the optimal block sizes for both periods in Table 1.14.

Table 1.14: Block Sizes for Stationary Block Bootstrapping

<table>
<thead>
<tr>
<th>Block Size</th>
<th>Stable Period</th>
<th>Unstable Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_l$</td>
<td>1.8055</td>
<td>53.2275</td>
</tr>
<tr>
<td>$b_u$</td>
<td>2.5007</td>
<td>53.4241</td>
</tr>
<tr>
<td>$b = (b_l + b_u)/2$</td>
<td>2.1531</td>
<td>53.3258</td>
</tr>
</tbody>
</table>

The optimal block size for the stable period is around 2 and for the unstable period
53. This large difference in the block size can be interpreted as the existence of a much
larger persistence in the IAR system of the unstable period than in the stable period.

We report the second-step estimates for the stable period in Tables 1.15 and 1.16.
Table 1.15: Lower Bound Regression Results of Three Models (Stable Period)

<table>
<thead>
<tr>
<th>regressors</th>
<th>OLS</th>
<th>Two-step</th>
<th>Modified Two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>std. err.</td>
<td>coefficient</td>
</tr>
<tr>
<td>const</td>
<td>-0.3062 (0.0509)</td>
<td>-0.0729 (0.1400)</td>
<td>-0.2947 (0.0499)</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>0.0586 (0.0451)</td>
<td>0.1400 (0.0568)</td>
<td>0.0626 (0.0461)</td>
</tr>
<tr>
<td>$r_{t-2}$</td>
<td>-0.0283 (0.0470)</td>
<td>-0.0295 (0.0413)</td>
<td>-0.0283 (0.0444)</td>
</tr>
<tr>
<td>$r_{t-2}$</td>
<td>0.1100 (0.0453)</td>
<td>0.2408 (0.0920)</td>
<td>0.1165 (0.0462)</td>
</tr>
<tr>
<td>$\hat{\lambda}_{t-1}$</td>
<td>-0.0840 (0.0463)</td>
<td>-0.1410 (0.0606)</td>
<td>-0.0868 (0.0472)</td>
</tr>
</tbody>
</table>

Degree of Freedom 749 748 749
Standard Error of Regression 0.4143 0.4139 0.4143
Adjusted $R^2$ 0.5208 0.5217 0.5147
$F$-statistic 164.9 138.1 160.9
Table 1.16: Upper Bound Regression Results of Three Models (Stable Period)

<table>
<thead>
<tr>
<th>regressors</th>
<th>OLS</th>
<th>Two-step</th>
<th>Modified Two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>std. err.</td>
<td>coefficient</td>
</tr>
<tr>
<td>const</td>
<td>0.3674</td>
<td>(0.0485)</td>
<td>0.8233</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>-0.0938</td>
<td>(0.0430)</td>
<td>0.0652</td>
</tr>
<tr>
<td>$r_{t-1}^*$</td>
<td>-0.0209</td>
<td>(0.0448)</td>
<td>-0.0233</td>
</tr>
<tr>
<td>$r_{t-2}$</td>
<td>-0.1377</td>
<td>(0.0432)</td>
<td>0.1179</td>
</tr>
<tr>
<td>$r_{t-2}^*$</td>
<td>0.0163</td>
<td>(0.0441)</td>
<td>-0.0950</td>
</tr>
<tr>
<td>$\lambda_{t-1}$</td>
<td>0.1708</td>
<td>(0.0432)</td>
<td>0.1730</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.1555</td>
<td>(0.0393)</td>
<td>0.5848</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.0670</td>
<td>(0.0218)</td>
<td>0.1708</td>
</tr>
</tbody>
</table>

Degree of Freedom: 749
Standard Error of Regression: 0.3954
Adjusted $R^2$: 0.5846
$F$-statistic: 213.2
The estimates of $\hat{\lambda}_t$ in the first step already suggest that the observability restriction is not binding, thus OLS should suffice. However, we also report the estimates of the two-step and modified two-step estimation procedures to underline the presence of multicollinearity caused by $\hat{\lambda}_t$ being almost zero. Observe that the OLS estimates and the modified minimum-distance estimates are almost identical, and that there is not loss of efficiency by implementing the modified estimator. This is what we expect when the restriction is not binding. On the contrary, the two-step estimator is much less reliable, the estimates are different from the OLS estimates, even changing signs, and overall their standard errors are large as a consequence of the induced multicollinearity.

We report the second-step estimates for the unstable period in Tables 1.17 and 1.18.
Table 1.17: Lower Bound Regression Results of Three Models (Unstable Period)

<table>
<thead>
<tr>
<th>regressors</th>
<th>OLS</th>
<th>Two-step</th>
<th>Modified Two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>std. err.</td>
<td>coefficient</td>
</tr>
<tr>
<td>const</td>
<td>−0.1281</td>
<td>(0.0649)</td>
<td>−0.0951</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>0.0434</td>
<td>(0.0425)</td>
<td>0.0456</td>
</tr>
<tr>
<td>$r_{t-2}$</td>
<td>0.0420</td>
<td>(0.0492)</td>
<td>0.0435</td>
</tr>
<tr>
<td>$r_{t-3}$</td>
<td>0.1025</td>
<td>(0.0443)</td>
<td>0.1055</td>
</tr>
<tr>
<td>$r_{t-4}$</td>
<td>−0.1075</td>
<td>(0.0486)</td>
<td>−0.1080</td>
</tr>
<tr>
<td>$r_{t-5}$</td>
<td>0.1346</td>
<td>(0.0458)</td>
<td>0.1371</td>
</tr>
<tr>
<td>$r_{t-6}$</td>
<td>0.0094</td>
<td>(0.0480)</td>
<td>0.0102</td>
</tr>
<tr>
<td>$r_{t-7}$</td>
<td>0.1751</td>
<td>(0.0462)</td>
<td>0.1784</td>
</tr>
<tr>
<td>$r_{t-8}$</td>
<td>−0.1581</td>
<td>(0.0461)</td>
<td>−0.1595</td>
</tr>
<tr>
<td>$r_{t-9}$</td>
<td>0.1202</td>
<td>(0.0469)</td>
<td>0.1230</td>
</tr>
<tr>
<td>$r_{t-10}$</td>
<td>−0.1100</td>
<td>(0.0436)</td>
<td>−0.1117</td>
</tr>
<tr>
<td>$\hat{\lambda}_{t-1}$</td>
<td>−0.0874</td>
<td>(0.2263)</td>
<td>−0.6775</td>
</tr>
</tbody>
</table>

Degree of Freedom 993 992 993
Standard Error of Regression 1.142 1.142 1.144
Adjusted $R^2$ 0.4953 0.4948 0.4646
$F$-statistic 90.57 82.96 80.2
Table 1.18: Upper Bound Regression Results of Three Models (Unstable Period)

<table>
<thead>
<tr>
<th>regressors</th>
<th>OLS</th>
<th>Two-step</th>
<th>Modified Two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>std. err.</td>
<td>coefficient</td>
</tr>
<tr>
<td>const</td>
<td>0.1210</td>
<td>(0.0562)</td>
<td>-0.0469</td>
</tr>
<tr>
<td>$r_{it-1}$</td>
<td>-0.2563</td>
<td>(0.0369)</td>
<td>-0.2673</td>
</tr>
<tr>
<td>$r_{ut-1}$</td>
<td>0.0622</td>
<td>(0.0426)</td>
<td>0.0548</td>
</tr>
<tr>
<td>$r_{it-2}$</td>
<td>-0.1839</td>
<td>(0.0384)</td>
<td>-0.1994</td>
</tr>
<tr>
<td>$r_{ut-2}$</td>
<td>0.0191</td>
<td>(0.0422)</td>
<td>0.0214</td>
</tr>
<tr>
<td>$r_{it-3}$</td>
<td>-0.0866</td>
<td>(0.0397)</td>
<td>-0.0992</td>
</tr>
<tr>
<td>$r_{ut-3}$</td>
<td>0.0626</td>
<td>(0.0416)</td>
<td>0.0583</td>
</tr>
<tr>
<td>$r_{it-4}$</td>
<td>-0.0377</td>
<td>(0.0400)</td>
<td>-0.0548</td>
</tr>
<tr>
<td>$r_{ut-4}$</td>
<td>-0.0393</td>
<td>(0.0400)</td>
<td>-0.0321</td>
</tr>
<tr>
<td>$r_{it-5}$</td>
<td>-0.0823</td>
<td>(0.0407)</td>
<td>-0.0968</td>
</tr>
<tr>
<td>$r_{ut-5}$</td>
<td>0.0339</td>
<td>(0.0378)</td>
<td>0.0425</td>
</tr>
<tr>
<td>$\hat{\lambda}_{t-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_t^2$</td>
<td>1.2909</td>
<td></td>
<td>1.2080</td>
</tr>
<tr>
<td>$\sigma_u^2$</td>
<td>0.9699</td>
<td></td>
<td>0.9960</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.6780</td>
<td></td>
<td>0.5752</td>
</tr>
<tr>
<td>Degree of Freedom</td>
<td>993</td>
<td></td>
<td>992</td>
</tr>
<tr>
<td>Standard Error of Regression</td>
<td>0.9888</td>
<td></td>
<td>0.9888</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.5294</td>
<td></td>
<td>0.5304</td>
</tr>
<tr>
<td>$F$-statistic</td>
<td>103.7</td>
<td></td>
<td>95.49</td>
</tr>
</tbody>
</table>
In this period, we already know that $\hat{\lambda}_t$ is large and different from zero, so that the observability restriction is binding in a substantial part of the sample. As expected, the modified two-step estimates are different from the OLS estimates, more so in the regression for the lower bound. Note that the estimates associated with $\hat{\lambda}_t$, though with the right signs, are barely significant in the two-step estimation because of some mild multicollinearity, which is corrected in the modified two-step estimation.

The severity of observability restriction is better illustrated in Figures 1.8 and 1.9.
Figure 1.8: Observability Restriction: Stable Period (2004/1/1-2007/1/1)

Figure 1.9: Observability Restriction: Unstable Period (2007/1/1-2011/4/29)
The ellipses are contours of the bivariate normal probability density of the errors with different confidence levels (from 50% to 99%). The contours are drawn according to the estimates produced by the modified two-step estimation procedure. The 45-degree lines indicate the role of the observability restrictions for each time \( t \) (see Figure 1), so that the area of the density below the line is truncated. Observe that in the stable period, Figure 1.8, the contours are smaller than those in the unstable period, Figure 1.9 because of smaller variances. In the stable period, the lines corresponding to the observability restriction are clustered outside the 99% contour level, so that the truncation is minimal; however for the unstable period, the truncation of the bivariate density is large, mainly in the direction of the south-east quadrant, indicating the severity of the observability restriction.

### 1.6 Comparison with Existing Approaches

In this section, we compare our two-step (TS) and modified two-step (MTS) estimators with the estimation techniques proposed in the current literature. We implement the approaches suggested by Lima Neto and De Carvalho (2008, 2010), henceforth LNC, and we also estimate a location-scale model from which we construct interval estimates.

Given an interval-valued time series \( \{Y_t\} = \{[Y_{lt}, Y_{ut}]\} \), we obtain the time series of the centers, i.e., \( y_{ct} = (y_{lt} + y_{ut})/2 \), and the time series of the radius, i.e., \( y_{rt} = (y_{ut} - y_{lt})/2 \). LNC estimate the following system of equations

\[
\begin{align*}
    y_{ct} &= \beta_0^c + \beta_1^c y_{ct-1} + \cdots + \beta_p^c y_{ct-p} + \epsilon_t^c \quad (1.6.48) \\
    y_{rt} &= \beta_0^r + \beta_1^r y_{rt-1} + \cdots + \beta_p^r y_{rt-p} + \epsilon_t^r. \quad (1.6.49)
\end{align*}
\]
Their center/range method (CRM) consists of estimating each equation by least squares; and their constrained center/range method (CCRM) consists of imposing the restriction $\beta_j^r \geq 0, j = 0, \ldots, p$ on the equation of the radius to ensure that $\hat{y}_{rt} \geq 0$ and, therefore, $\hat{y}_{lt} \leq \hat{y}_{ut}$. Then, the equation of the center is estimated by least squares and the constrained equation of the radius is estimated by adapting Lawson and Hanson’s algorithm.

Before we proceed with the comparison among methodologies, it is very important to underline the implications of the LNC system of center/radius equations for the system of lower/upper bound equations, which is estimated in this paper. By defining the matrix $M$

$$M = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

it is easy to transform the center/radius vector to the lower/upper bound vector, i.e.,

$$\begin{bmatrix} y_{ct} \\ y_{rt} \end{bmatrix} = M \begin{bmatrix} y_{lt} \\ y_{ut} \end{bmatrix}.$$

Hence, the system of equations (4.4.18) and (4.4.19) is transformed into a system of lower/upper bound equations as follows,

$$\begin{bmatrix} y_{lt} \\ y_{ut} \end{bmatrix} = M^{-1} \begin{bmatrix} \beta_0^c \\ \beta_0^r \end{bmatrix} + \sum_{l=1}^{p} M^{-1} \begin{bmatrix} \beta_l^c & 0 \\ 0 & \beta_l^r \end{bmatrix} M \begin{bmatrix} y_{lt-1} \\ y_{ut-1} \end{bmatrix} + M^{-1} \begin{bmatrix} \epsilon_t^c \\ \epsilon_t^r \end{bmatrix},$$

$$= \begin{bmatrix} \beta_0^c - \beta_0^r \\ \beta_0^c + \beta_0^r \end{bmatrix} + \sum_{l=1}^{p} \begin{bmatrix} (\beta_l^c + \beta_l^r)/2 & (\beta_l^c - \beta_l^r)/2 \\ (\beta_l^c - \beta_l^r)/2 & (\beta_l^c + \beta_l^r)/2 \end{bmatrix} \begin{bmatrix} y_{lt-1} \\ y_{ut-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^c - \epsilon_t^r \\ \epsilon_t^c + \epsilon_t^r \end{bmatrix},$$

which is an extremely restricted system because, for each of the $p$ matrices of coefficients,
the diagonal elements must be identical, both equal to \((\beta^c_l + \beta^r_l)/2\), as well as the off-diagonal elements, both equal to \((\beta^c_l - \beta^r_l)/2\). Only when these restrictions are satisfied, which is highly unlikely, the LNC approach and our approach will deliver the same results.

The second set of comparisons is with a location-scale model applied to the time series of centers. We estimate the following GARCH(1,1) model

\[
y_{ct} = \mu_c + \varepsilon_t \\
= \mu_c + \sigma_t \zeta_t , \\
\sigma^2_t = \omega + \alpha \varepsilon^2_{t-1} + \beta \sigma^2_{t-1}
\]

where the i.i.d. standardized error term \(\zeta_t\) follows a standard normal or Student-\(t\) density with \(\nu\) degrees of freedom. Based on this model, we construct \((1-\alpha)\)-probability intervals, which will depend on the distributional assumptions on \(\zeta_t\), i.e.,

\[
[\hat{y}_{lt}, \hat{y}_{ut}]_{\alpha} = \left[ \bar{y}_{ct} - z_{\frac{\nu}{2}} \hat{\sigma}_t, \bar{y}_{ct} + z_{\frac{\nu}{2}} \hat{\sigma}_t \right] \\
[\hat{y}_{lt}, \hat{y}_{ut}]_{\alpha} = \left[ \bar{y}_{ct} - \nu \nu - 2 \bar{y}_{ct} + \nu \nu - 2 \right] \\
\]

Since the original data \([y_{lt}, y_{ut}]\) are the observed extreme values of the process at time \(t\), we will stretch the estimated interval \([\hat{y}_{lt}, \hat{y}_{ut}]_{\alpha}\) to cover as much as 99% or 99.5% probability, so that \(\hat{y}_{lt}\) and \(\hat{y}_{ut}\) are far away into the tails of the distribution.

In Table 1.19, we summarize the eight methodologies that we implement in the estimation of an IAR(1) model. We simulate data from four DGPs, which are characterized

---

5We are grateful to a referee who suggested the 5-parameter location-scale model as a classical
Table 1.19: Estimation Methodologies

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCRM</td>
<td>Constrained Center and Range Method</td>
</tr>
<tr>
<td>CRM</td>
<td>Center and Range Method</td>
</tr>
<tr>
<td>TS</td>
<td>Two-Step Estimation</td>
</tr>
<tr>
<td>MTS</td>
<td>Modified Two-Step Estimation</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>99% C.I. from GARCH(1,1) with conditional normality</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>99.5% C.I. from GARCH(1,1) with conditional normality</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>99% C.I. from GARCH(1,1) with conditional Student-t</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>99.5% C.I. from GARCH(1,1) with conditional Student-t</td>
</tr>
</tbody>
</table>

by whether the observability restriction is binding or not, whether there is high or low persistence in the dynamics of the conditional mean, and whether the errors of the model are drawn from a bivariate normal density or from a bivariate Student-t density with five degrees of freedom. In Table 1.20 we show the four DGPs.

Table 1.20: Data Generating Processes

<table>
<thead>
<tr>
<th>β₀₀</th>
<th>β₀₁</th>
<th>β₁₁</th>
<th>β₁₂</th>
<th>β₂₁</th>
<th>β₂₂</th>
<th>σ₁²</th>
<th>σ₂²</th>
<th>ρ</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1</td>
<td>0</td>
<td>0</td>
<td>-0.8</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.8</td>
<td>3</td>
<td>1</td>
<td>-0.8</td>
</tr>
<tr>
<td>DGP2</td>
<td>0</td>
<td>0</td>
<td>-0.1</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.1</td>
<td>3</td>
<td>1</td>
<td>-0.8</td>
</tr>
<tr>
<td>DGP3</td>
<td>-2</td>
<td>2</td>
<td>-0.8</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.8</td>
<td>3</td>
<td>1</td>
<td>-0.8</td>
</tr>
<tr>
<td>DGP4</td>
<td>-2</td>
<td>2</td>
<td>-0.1</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.1</td>
<td>3</td>
<td>1</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

* B: binding observability restriction; NB: non-binding; H: high persistence; L: low persistence; N: normal errors; T5: Student-t errors

1.6.1 In-Sample Evaluation Criteria: Loss Functions

For every DGP in Table 1.20 we generate 1000 samples, and we evaluate the performance of each estimation methodology in Table 1.19 according to four criteria: (i) Root Mean Squared Error (RMSE) for each of the upper and lower bounds, (ii) Coverage (CR) and Efficiency Rates (ER) of the estimated intervals (Rodrigues and Salish, 2011), (iii) Multivariate Loss Functions (MLF) for the vector of lower and upper bounds (Komunjer benchmark)
and Owyang, 2011), and (iv) Mean Distance Error (MDE) between the fitted and actual intervals (Arroyo et al., 2010).

For a sample of size \( T \), let us call \( \hat{y}_t = [\hat{y}_{lt}, \hat{y}_{ut}] \) the fitted values of the corresponding interval \( y_t = [y_{lt}, y_{ut}] \) obtained by each methodology. These are the definitions of the four criteria:

(i) RMSE:

\[
RMSE_{lt} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{lt} - y_{lt})^2} \\
RMSE_{ut} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{ut} - y_{ut})^2}.
\]

(ii) CR and ER:

\[
CR = \frac{1}{T} \sum_{t=1}^{T} \frac{w(y_t \cap \hat{y}_t)}{w(y_t)}, \\
ER = \frac{1}{T} \sum_{t=1}^{T} \frac{w(y_t \cap \hat{y}_t)}{w(\hat{y}_t)}.
\]

where \( y_t \cap \hat{y}_t \) is the intersection of the actual and fitted intervals, and \( w(\cdot) \) represents the range or width of the interval.

The coverage rate is the average proportion of the actual interval that is covered by the fitted interval, and the efficiency rate is the average proportion of the fitted interval that is covered by the actual interval. Both rates are between zero and one and a large rate
means a better fit. However, given an actual interval, a wide fitted interval will imply a large coverage rate but a low efficiency rate, on the contrary, a tight fitted interval will imply a low coverage rate but a high efficiency rate. Therefore, we take into account the potential trade-off between the two rates by calculating an average of the two rates, i.e., $(CR + ER)/2$.

(iii) MLF: We implement the following multivariate loss function

$$L_p(\tau, e) \equiv \left( \| e \|_p + \tau \| e \|_p^{-1} \right) / \| e \|_p^{-1}$$

where $\| \cdot \|_p$ is the $l_p$-norm, $\tau$ is two-dimensional parameter vector bounded by the unit ball $B_q$ in $\mathbb{R}^2$ with $l_q$-norm (where $p$ and $q$ satisfy $1/p + 1/q = 1$), and $e = (e_l, e_u)$ is the bivariate residual interval $(\hat{y}_{lt} - y_{lt}, \hat{y}_{ut} - y_{ut})$. We consider two norms, $p = 1$ and $p = 2$ and their corresponding $\tau$ parameter vectors within the unit balls $B_\infty$ and $B_2$ respectively,

$$MLF_1 = \int_{\tau \in B_\infty} (|e_l| + |e_u| + \tau_1 e_l + \tau_2 e_u) d\tau$$

$$MLF_2 = \int_{\tau \in B_2} \left[ e_l^2 + e_u^2 + (\tau_1 e_l + \tau_2 e_u)(e_l^2 + e_u^2)^{1/2} \right] d\tau.$$

(iv) MDE: Let $D^q(\hat{y}_t, y_t)$ be a distance measure of order $q$ between the fitted and the actual intervals, the mean distance error is defined as

$$MDE^q(\{\hat{y}_t\}, \{y_t\}) = \left( \frac{\sum_{t=1}^T D^q(\hat{y}_t, y_t)}{T} \right)^{1/q}$$
We consider two cases, $q = 1$ and $q = 2$, with a distance measure such as

$$D(\hat{y}_t, y_t) = \frac{1}{\sqrt{2}} \sqrt{(\hat{y}_{ut} - y_{ut})^2 + (\hat{y}_{ut} - y_{ut})^2}.$$  

In Tables 1.21, 1.22, 1.23, and 1.24, we report the values of the four aforementioned evaluation criteria for the four DGPs considered in Table 1.20 estimated according to the methodologies described in Table 1.19.
Table 1.21: Methodology Evaluation for DGP1 (HIGH persistence and BINDING observability restriction)

<table>
<thead>
<tr>
<th>Methodology</th>
<th>RMSE</th>
<th>CR &amp; ER</th>
<th>MLF</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Upper</td>
<td>CR</td>
<td>ER</td>
<td>$\frac{CR + ER}{2}$</td>
</tr>
<tr>
<td>DGP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multivariate Normal Distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>1.5851</td>
<td>1.2286</td>
<td>0.7099</td>
<td>0.6086</td>
</tr>
<tr>
<td>CRM</td>
<td>1.5201</td>
<td>1.2973</td>
<td>0.7066</td>
<td>0.6073</td>
</tr>
<tr>
<td>TS</td>
<td><strong>1.2735</strong></td>
<td><strong>0.7689</strong></td>
<td>0.8244</td>
<td>0.7023</td>
</tr>
<tr>
<td>MTS</td>
<td><strong>1.2738</strong></td>
<td><strong>0.7691</strong></td>
<td>0.8244</td>
<td>0.7022</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>2.9030</td>
<td>2.6388</td>
<td>0.9877</td>
<td>0.3604</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>3.1914</td>
<td>2.9510</td>
<td>0.9928</td>
<td>0.3343</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>2.2336</td>
<td>1.8782</td>
<td>0.9543</td>
<td>0.4440</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>2.5063</td>
<td>2.1911</td>
<td>0.9734</td>
<td>0.4064</td>
</tr>
<tr>
<td><strong>Multivariate Student’s t Distribution ($\nu = 5$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>2.1135</td>
<td>1.6705</td>
<td>0.7064</td>
<td>0.5856</td>
</tr>
<tr>
<td>CRM</td>
<td>2.0382</td>
<td>1.7506</td>
<td>0.7055</td>
<td>0.5864</td>
</tr>
<tr>
<td>TS</td>
<td><strong>1.6877</strong></td>
<td><strong>1.0099</strong></td>
<td>0.8323</td>
<td>0.6894</td>
</tr>
<tr>
<td>MTS</td>
<td><strong>1.6890</strong></td>
<td><strong>1.0107</strong></td>
<td>0.8327</td>
<td>0.6894</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>3.8119</td>
<td>3.4153</td>
<td>0.9819</td>
<td>0.3373</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>4.1732</td>
<td>3.8073</td>
<td>0.9874</td>
<td>0.3131</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>3.1679</td>
<td>2.6299</td>
<td>0.9598</td>
<td>0.4013</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>3.6105</td>
<td>3.1287</td>
<td>0.9761</td>
<td>0.3609</td>
</tr>
</tbody>
</table>
Table 1.22: Methodology Evaluation for DGP2 (LOW persistence and BINDING observability restriction)

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>CR &amp; ER</th>
<th>MLF</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>CR</td>
<td>ER</td>
</tr>
<tr>
<td>DGP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate Normal Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>1.0963</td>
<td>0.6885</td>
<td>0.7992</td>
<td>0.6583</td>
</tr>
<tr>
<td>CRM</td>
<td>1.0954</td>
<td>0.6896</td>
<td>0.7989</td>
<td>0.6582</td>
</tr>
<tr>
<td>TS</td>
<td>1.0903</td>
<td>0.6809</td>
<td>0.8025</td>
<td>0.6607</td>
</tr>
<tr>
<td>MTS</td>
<td>1.0906</td>
<td>0.6811</td>
<td>0.8026</td>
<td>0.6606</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>1.1266</td>
<td>0.7353</td>
<td>0.8719</td>
<td>0.6061</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>1.1580</td>
<td>0.7825</td>
<td>0.8959</td>
<td>0.5826</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>1.1538</td>
<td>0.7756</td>
<td>0.8926</td>
<td>0.5859</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>1.2220</td>
<td>0.8736</td>
<td>0.9222</td>
<td>0.5497</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>1.6448</td>
<td>1.0786</td>
<td>0.8976</td>
<td>0.5454</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>1.6967</td>
<td>1.1556</td>
<td>0.9154</td>
<td>0.5205</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>1.7500</td>
<td>1.2253</td>
<td>0.9261</td>
<td>0.5029</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>1.9388</td>
<td>1.4802</td>
<td>0.9511</td>
<td>0.4515</td>
</tr>
</tbody>
</table>

(ν = 5)
Table 1.23: Methodology Evaluation for DGP3 (HIGH persistence and NON-BINDING observability restriction)

<table>
<thead>
<tr>
<th>Methodology</th>
<th>RMSE</th>
<th>CR &amp; ER</th>
<th>MLF</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate Normal Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>2.4973</td>
<td>2.0577</td>
<td>0.8655</td>
<td>0.8500</td>
</tr>
<tr>
<td>CRM</td>
<td>2.1549</td>
<td>2.2905</td>
<td>0.8650</td>
<td>0.8493</td>
</tr>
<tr>
<td>TS</td>
<td>1.7255</td>
<td>0.9964</td>
<td>0.9181</td>
<td>0.9063</td>
</tr>
<tr>
<td>MTS</td>
<td>1.7259</td>
<td>0.9965</td>
<td>0.9181</td>
<td>0.9063</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>8.3435</td>
<td>8.0010</td>
<td>0.9999</td>
<td>0.7039</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>9.4680</td>
<td>9.1662</td>
<td>1.0000</td>
<td>0.6872</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>3.1196</td>
<td>2.0920</td>
<td>0.7845</td>
<td>0.8239</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>3.0973</td>
<td>2.0581</td>
<td>0.8430</td>
<td>0.8370</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate Student’s t Distribution (ν = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>3.0472</td>
<td>2.5013</td>
<td>0.8463</td>
<td>0.8253</td>
</tr>
<tr>
<td>CRM</td>
<td>2.6563</td>
<td>2.7776</td>
<td>0.8447</td>
<td>0.8237</td>
</tr>
<tr>
<td>TS</td>
<td>2.1303</td>
<td>1.2356</td>
<td>0.9110</td>
<td>0.8945</td>
</tr>
<tr>
<td>MTS</td>
<td>2.1308</td>
<td>1.2359</td>
<td>0.9110</td>
<td>0.8945</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>8.9350</td>
<td>8.4656</td>
<td>0.9993</td>
<td>0.6992</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>10.0825</td>
<td>9.6664</td>
<td>0.9997</td>
<td>0.6831</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>3.7183</td>
<td>2.3954</td>
<td>0.8416</td>
<td>0.8174</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>3.8918</td>
<td>2.6454</td>
<td>0.8914</td>
<td>0.8223</td>
</tr>
<tr>
<td>Method</td>
<td>RMSE</td>
<td>CR &amp; ER</td>
<td>MLF</td>
<td>MDE</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>CR</td>
<td>ER</td>
</tr>
<tr>
<td>CCRM</td>
<td>1.5523</td>
<td>0.9149</td>
<td>0.8544</td>
<td>0.7714</td>
</tr>
<tr>
<td>CRM</td>
<td>1.5508</td>
<td>0.9168</td>
<td>0.8543</td>
<td>0.7713</td>
</tr>
<tr>
<td>TS</td>
<td>1.5448</td>
<td>0.9044</td>
<td>0.8559</td>
<td>0.7728</td>
</tr>
<tr>
<td>MTS</td>
<td>1.5452</td>
<td>0.9047</td>
<td>0.8559</td>
<td>0.7728</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>1.6669</td>
<td>1.0944</td>
<td>0.7208</td>
<td>0.8421</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>1.6217</td>
<td>1.0242</td>
<td>0.7591</td>
<td>0.8264</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>1.6594</td>
<td>1.0816</td>
<td>0.7276</td>
<td>0.8394</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>1.6000</td>
<td>0.9879</td>
<td>0.7817</td>
<td>0.8155</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>1.9226</td>
<td>1.2020</td>
<td>0.7640</td>
<td>0.8142</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>1.8892</td>
<td>1.1479</td>
<td>0.8003</td>
<td>0.7951</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>1.8989</td>
<td>1.0810</td>
<td>0.7951</td>
<td>0.7978</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>1.8741</td>
<td>1.1119</td>
<td>0.8508</td>
<td>0.7591</td>
</tr>
</tbody>
</table>
The numbers in boldface correspond to the minimum losses when we consider the functions RMSE, MLF, and MDE, and to the maximum rates when we consider the weighted CR/ER rates. In each table, we provide two scenarios: in the upper panel, the DGP is simulated with multivariate normal errors so that our methods TS and MTS perform under the correct distributional assumption, and in the lower panel, the DGP is simulated with multivariate Student-t errors to assess the performance of TS and MTS under misspecification in the distributional assumptions. These are our findings:

1. Across the four tables, TS and MTS exhibit superior performance over the other six methodologies considered.

2. Across methods, TS and MTS are superior to CCRM and CRM, and these are far better than the GARCH models. The classical methodology embedded in normal or fat-tail location-scale models is by far the worst performer across all evaluation functions and it is very inefficient on delivering an acceptable fitted interval as the efficiency rates (ER) shows.

3. In the misspecified scenarios with Student-t errors (lower panels), the losses across all methods are larger than those under correct error specification, which is expected, but nevertheless TS and MTS provide the smallest losses.

4. Across DGPs, DGP1 and DGP3, which have high persistence in the conditional mean, have the smallest losses, and in particular, TS and MTS deliver unmatched performance even in the cases of misspecified distributional assumptions.

5. DGP2 and DGP4 have low persistence in the conditional mean. Observe that in these two specifications, the coefficients are all very close to zero, thus, in these cases the constraints imposed by CCRM and CRM are not so restrictive and, as a consequence, the
performance of CCRM and CRM is close to that of TS and MTS, but the performance of the location-scale models is still far behind the other methods.

6. Only for DGP4 with low persistence in mean and non-binding observability restriction, the performance of all methods is roughly equivalent, which is expected as all constraints are relaxed.

In summary, when the researcher faces an interval-valued data set, a priori, she does not know the persistence of the data and whether the observability restriction is or is not binding, thus, it is advisable to start the estimation of the model by implementing TS or/and MTS. If there is high persistence in the conditional mean, even if the observability restriction is non-binding, it pays off to implement TS and MTS as the losses are substantially smaller than those from the competing methodologies. In addition, the implementation of a location-scale model also entails the choice of distributional assumptions, which is subject to misspecification issues.

1.6.2 In-Sample Evaluation Criteria: Mean Estimates, Bias, and MSE

In the previous section, we have evaluated the methodologies according to several goodness-of-fit criteria, and we have eliminated the location-scale models as a competing method. In this section, we compare the mean estimates of the parameters in the conditional mean delivered by our proposed TS and MTS in relation to those provided by CCRM and CRM. As before, we consider the four DGPs with correctly specified multivariate normal errors, and with Student-\(t\) errors to assess the effect of misspecification in the distribution.

In Table 1.25 (DGP1 and DGP2) and Table 1.26 (DGP3 and DGP4), we present the
simulation results for the case of multivariate normal errors.

Table 1.25: Simulation Results of DGP1 and GDP2 with Multivariate Normal Errors

<table>
<thead>
<tr>
<th>Mean</th>
<th>DGP1 (high persistence and binding observability restriction)</th>
<th>DGP2 (low persistence and binding observability restriction)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>$b_{12}$</td>
</tr>
<tr>
<td>true value</td>
<td>-0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.0986</td>
<td>-0.0986</td>
</tr>
<tr>
<td>CRM</td>
<td>-0.1553</td>
<td>-0.0419</td>
</tr>
<tr>
<td>TS</td>
<td><strong>-0.7930</strong></td>
<td><strong>0.1081</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>-0.7970</strong></td>
<td><strong>0.1046</strong></td>
</tr>
<tr>
<td>Bias$^2$</td>
<td>CCRM</td>
<td>0.4920</td>
</tr>
<tr>
<td>CRM</td>
<td>0.4156</td>
<td>0.0201</td>
</tr>
<tr>
<td>TS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td>MSE</td>
<td>CCRM</td>
<td>0.4922</td>
</tr>
<tr>
<td>CRM</td>
<td>0.4164</td>
<td>0.0202</td>
</tr>
<tr>
<td>TS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
</tbody>
</table>

The numbers in boldface are the best estimates, the lowest bias and the lowest mean-square error. In DGP1 and DGP2, Table 1.25, the observability restriction is binding. When the process has high persistence (DGP1), CCRM and CRM perform very badly.
Table 1.26: Simulation Results of DGP3 and DGP4 with Multivariate Normal Errors

<table>
<thead>
<tr>
<th></th>
<th>DGP3 (high persistence and non-binding observability restriction)</th>
<th>DGP4 (low persistence and non-binding observability restriction)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>$b_{12}$</td>
</tr>
<tr>
<td>true value</td>
<td>-0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.2168</td>
<td>-0.2168</td>
</tr>
<tr>
<td>CRM</td>
<td>-0.3703</td>
<td>-0.0634</td>
</tr>
<tr>
<td>TS</td>
<td>-0.8002</td>
<td>0.1014</td>
</tr>
<tr>
<td>MTS</td>
<td>-0.7998</td>
<td>0.1018</td>
</tr>
<tr>
<td>Bias$^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.3401</td>
<td>0.1004</td>
</tr>
<tr>
<td>CRM</td>
<td>0.1847</td>
<td>0.0267</td>
</tr>
<tr>
<td>TS</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>MTS</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.3404</td>
<td>0.1007</td>
</tr>
<tr>
<td>CRM</td>
<td>0.1861</td>
<td>0.0267</td>
</tr>
<tr>
<td>TS</td>
<td>0.0002</td>
<td>0.0005</td>
</tr>
<tr>
<td>MTS</td>
<td>0.0002</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>-0.0935</td>
<td>-0.0976</td>
</tr>
<tr>
<td>CRM</td>
<td>-0.1016</td>
<td>-0.0895</td>
</tr>
<tr>
<td>TS</td>
<td>-0.1671</td>
<td>-0.0905</td>
</tr>
<tr>
<td>MTS</td>
<td>-0.1000</td>
<td>-0.0527</td>
</tr>
<tr>
<td>Bias$^2$</td>
<td>0.0000</td>
<td>0.0218</td>
</tr>
<tr>
<td>CRM</td>
<td>0.0000</td>
<td>0.0194</td>
</tr>
<tr>
<td>TS</td>
<td>0.0045</td>
<td>0.0197</td>
</tr>
<tr>
<td>MTS</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.0002</td>
<td>0.0219</td>
</tr>
<tr>
<td>CRM</td>
<td>0.0004</td>
<td>0.0196</td>
</tr>
<tr>
<td>TS</td>
<td>4.722</td>
<td>10.59</td>
</tr>
<tr>
<td>MTS</td>
<td>0.0017</td>
<td>0.0044</td>
</tr>
</tbody>
</table>
The mean estimates have a large bias and frequently the wrong sign. On the contrary, TS and MTS deliver unbiased estimates with the lowest mean-square error. When the process has low persistence (DGP2), the best estimation method is MTS, which delivers unbiased estimates. Observe that TS suffers from the multicollinearity problem explained in the previous sections, and thus it is not recommended if our interest is to understand the dynamics in the conditional mean. CCRM and CRM estimates are not recommended either because of their large bias. In DGP3 and DGP4, Table 1.26, the observability restriction is non-binding but the results are not any different from those in Table 1.25. When the process has high persistence, either TS or MTS deliver unbiased estimates with the lowest mean-square error, and CCRM and CRM generate highly biased estimates. When the process has low persistence, MTS is the best performer because it not only takes care of the multicollinearity problem but also delivers unbiased estimates.

In Table 1.27 (DGP1 and DGP2) and Table 1.28 (DGP3 and DGP4), we present the simulation results for the case of multivariate Student-$t$ errors (5 degrees of freedom).

In DGP1 and DGP2, Table 1.27, the observability restriction is binding. When the process has high persistence (DGP1), the best performer is TS followed by MTS as they provide estimates with the lowest biases and capture the right dynamics. On the other hand, CCRM and CRM do not capture the persistence in the conditional mean and their estimates are highly biased. A common problem to these four methods is that the estimates of the constants are very biased. However, in TS and MTS, these biases are somehow compensated by the estimates of the coefficients corresponding to the regressor $\lambda_{t-1}$ so that the overall estimation generates good fitted intervals with
Table 1.27: Simulation Results of DGP1 and DGP2 with Multivariate Student-t Errors

<table>
<thead>
<tr>
<th></th>
<th>$b_{11}$</th>
<th>$b_{12}$</th>
<th>$b_{0L}$</th>
<th>$b_{21}$</th>
<th>$b_{22}$</th>
<th>$b_{0U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP1 (high persistence and binding observability restriction)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>true value</td>
<td>-0.8</td>
<td>0.1</td>
<td>0</td>
<td>-0.1</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.0976</td>
<td>-0.0976</td>
<td><strong>-0.1508</strong></td>
<td><strong>-0.0976</strong></td>
<td>-0.0976</td>
<td>3.2462</td>
</tr>
<tr>
<td>CRM</td>
<td>-0.1461</td>
<td>-0.0490</td>
<td>-0.3158</td>
<td>-0.0490</td>
<td>-0.1461</td>
<td>3.4111</td>
</tr>
<tr>
<td>TS</td>
<td><strong>-0.8889</strong></td>
<td><strong>0.0016</strong></td>
<td>1.1406</td>
<td><strong>-0.0522</strong></td>
<td><strong>0.8529</strong></td>
<td><strong>-0.6148</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>-0.9284</strong></td>
<td>-0.0329</td>
<td>1.5307</td>
<td>-0.0295</td>
<td><strong>0.8729</strong></td>
<td>-0.8406</td>
</tr>
<tr>
<td><strong>Bias$^2$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.4934</td>
<td>0.0390</td>
<td><strong>0.0228</strong></td>
<td><strong>0.0000</strong></td>
<td>0.8056</td>
<td>10.5376</td>
</tr>
<tr>
<td>CRM</td>
<td>0.4276</td>
<td>0.0222</td>
<td>0.0997</td>
<td>0.0026</td>
<td>0.8951</td>
<td>11.6359</td>
</tr>
<tr>
<td>TS</td>
<td><strong>0.0079</strong></td>
<td><strong>0.0097</strong></td>
<td>1.3010</td>
<td><strong>0.0023</strong></td>
<td><strong>0.0028</strong></td>
<td><strong>0.3780</strong></td>
</tr>
<tr>
<td>MTS</td>
<td>0.0165</td>
<td>0.0177</td>
<td>2.3431</td>
<td>0.0050</td>
<td>0.0053</td>
<td>0.7067</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.4937</td>
<td>0.0393</td>
<td><strong>0.0244</strong></td>
<td><strong>0.0002</strong></td>
<td>0.8058</td>
<td>10.5479</td>
</tr>
<tr>
<td>CRM</td>
<td>0.4285</td>
<td>0.0223</td>
<td>0.1008</td>
<td>0.0027</td>
<td>0.8960</td>
<td>11.6540</td>
</tr>
<tr>
<td>TS</td>
<td><strong>0.0134</strong></td>
<td><strong>0.0156</strong></td>
<td>1.7679</td>
<td><strong>0.0042</strong></td>
<td><strong>0.0047</strong></td>
<td><strong>0.5458</strong></td>
</tr>
<tr>
<td>MTS</td>
<td>0.0200</td>
<td>0.0220</td>
<td>3.5410</td>
<td>0.0061</td>
<td>0.0066</td>
<td>1.1562</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$b_{11}$</th>
<th>$b_{12}$</th>
<th>$b_{0L}$</th>
<th>$b_{21}$</th>
<th>$b_{22}$</th>
<th>$b_{0U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP2 (low persistence and binding observability restriction)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>true value</td>
<td>-0.1</td>
<td>0.05</td>
<td>0</td>
<td>-0.05</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.0613</td>
<td>-0.0685</td>
<td>-1.4316</td>
<td>-0.0685</td>
<td>-0.0613</td>
<td>1.0178</td>
</tr>
<tr>
<td>CRM</td>
<td><strong>-0.0671</strong></td>
<td><strong>-0.0627</strong></td>
<td>-1.4459</td>
<td><strong>-0.0627</strong></td>
<td>-0.0671</td>
<td>1.0322</td>
</tr>
<tr>
<td>TS</td>
<td>-0.1958</td>
<td>-0.2252</td>
<td>188.95</td>
<td>0.1240</td>
<td>0.3712</td>
<td>-4.9869</td>
</tr>
<tr>
<td>MTS</td>
<td>-0.2389</td>
<td>-0.0594</td>
<td>17.5335</td>
<td>0.0293</td>
<td><strong>0.1663</strong></td>
<td>-9.8989</td>
</tr>
<tr>
<td><strong>Bias$^2$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.0015</td>
<td>0.0140</td>
<td>2.0494</td>
<td>0.0003</td>
<td>0.0260</td>
<td>1.0360</td>
</tr>
<tr>
<td>CRM</td>
<td><strong>0.0011</strong></td>
<td><strong>0.0127</strong></td>
<td>2.0906</td>
<td><strong>0.0002</strong></td>
<td>0.0279</td>
<td>1.0653</td>
</tr>
<tr>
<td>TS</td>
<td>0.0092</td>
<td>0.0757</td>
<td>3.57e+04</td>
<td>0.0303</td>
<td>0.0735</td>
<td>24.8691</td>
</tr>
<tr>
<td>MTS</td>
<td>0.0193</td>
<td><strong>0.0120</strong></td>
<td>307.43</td>
<td>0.0063</td>
<td><strong>0.0044</strong></td>
<td>97.9887</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.0017</td>
<td>0.0142</td>
<td>2.0508</td>
<td>0.0005</td>
<td><strong>0.0263</strong></td>
<td>1.0368</td>
</tr>
<tr>
<td>CRM</td>
<td><strong>0.0015</strong></td>
<td><strong>0.0129</strong></td>
<td>2.0925</td>
<td><strong>0.0003</strong></td>
<td>0.0283</td>
<td>1.0667</td>
</tr>
<tr>
<td>TS</td>
<td>70.43</td>
<td>197.8</td>
<td>8.0e+07</td>
<td>39.10</td>
<td>88.69</td>
<td>6.3e+07</td>
</tr>
<tr>
<td>MTS</td>
<td>0.2874</td>
<td>0.7468</td>
<td>753.22</td>
<td>0.1046</td>
<td>0.2813</td>
<td>256.81</td>
</tr>
</tbody>
</table>
Table 1.28: Simulation Results of DGP3 and DGP4 with Multivariate Student-t Errors

<table>
<thead>
<tr>
<th>DGP3 (high persistence and non-binding observability restriction)</th>
<th>( b_{11} )</th>
<th>( b_{12} )</th>
<th>( b_{0L} )</th>
<th>( b_{21} )</th>
<th>( b_{22} )</th>
<th>( b_{0U} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>true value</td>
<td>-0.8</td>
<td>0.1</td>
<td>-2</td>
<td>-0.1</td>
<td>0.8</td>
<td>2</td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.2080</td>
<td>-0.2080</td>
<td>1.492</td>
<td>-0.2080</td>
<td>-0.2080</td>
<td>12.502</td>
</tr>
<tr>
<td>CRM</td>
<td>-0.3537</td>
<td>-0.0624</td>
<td>-0.1120</td>
<td>-0.0624</td>
<td>-0.3537</td>
<td>14.1055</td>
</tr>
<tr>
<td>TS</td>
<td><strong>-0.7947</strong></td>
<td><strong>0.1072</strong></td>
<td><strong>-2.1132</strong></td>
<td><strong>-0.1030</strong></td>
<td><strong>0.7957</strong></td>
<td><strong>2.0657</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>-0.7963</strong></td>
<td><strong>0.1059</strong></td>
<td><strong>-2.0963</strong></td>
<td><strong>-0.1021</strong></td>
<td><strong>0.7965</strong></td>
<td><strong>2.0559</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>CCRM</td>
<td>0.3504</td>
<td>0.0949</td>
<td>12.192</td>
<td>0.0117</td>
<td>1.0161</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td>0.1992</td>
<td>0.0264</td>
<td>3.5645</td>
<td>0.0014</td>
<td>1.3310</td>
</tr>
<tr>
<td></td>
<td>TS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0001</strong></td>
<td><strong>0.0128</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td></td>
<td>MTS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0093</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td>Bias²</td>
<td>CCRM</td>
<td>0.3508</td>
<td>0.0952</td>
<td>12.225</td>
<td>0.0120</td>
<td>1.0160</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td>0.2007</td>
<td>0.0264</td>
<td>3.5710</td>
<td>0.0015</td>
<td>1.3325</td>
</tr>
<tr>
<td></td>
<td>TS</td>
<td><strong>0.0003</strong></td>
<td><strong>0.0006</strong></td>
<td><strong>0.0827</strong></td>
<td><strong>0.0001</strong></td>
<td><strong>0.0002</strong></td>
</tr>
<tr>
<td></td>
<td>MTS</td>
<td><strong>0.0002</strong></td>
<td><strong>0.0006</strong></td>
<td><strong>0.0710</strong></td>
<td><strong>0.0001</strong></td>
<td><strong>0.0002</strong></td>
</tr>
<tr>
<td>MSE</td>
<td>CCRM</td>
<td>0.0001</td>
<td>0.0205</td>
<td><strong>0.0001</strong></td>
<td>0.0019</td>
<td>0.0356</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0184</strong></td>
<td>0.0019</td>
<td>0.0013</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>TS</td>
<td>0.0027</td>
<td>0.0121</td>
<td>429.39</td>
<td><strong>0.0001</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td></td>
<td>MTS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0128</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DGP4 (low persistence and non-binding observability restriction)</th>
<th>( b_{11} )</th>
<th>( b_{12} )</th>
<th>( b_{0L} )</th>
<th>( b_{21} )</th>
<th>( b_{22} )</th>
<th>( b_{0U} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>true value</td>
<td>-0.1</td>
<td>0.05</td>
<td>-2</td>
<td>-0.05</td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td>CCRM</td>
<td>-0.0886</td>
<td>-0.0931</td>
<td><strong>-2.0091</strong></td>
<td>-0.0931</td>
<td>-0.0886</td>
<td>2.6154</td>
</tr>
<tr>
<td>CRM</td>
<td><strong>-0.0962</strong></td>
<td>-0.0855</td>
<td><strong>-2.0441</strong></td>
<td>-0.0855</td>
<td>-0.0962</td>
<td>2.6504</td>
</tr>
<tr>
<td>TS</td>
<td>-0.1519</td>
<td>-0.0599</td>
<td>-22.722</td>
<td><strong>-0.0417</strong></td>
<td><strong>1.063</strong></td>
<td><strong>2.8518</strong></td>
</tr>
<tr>
<td>MTS</td>
<td><strong>-0.1031</strong></td>
<td><strong>0.0475</strong></td>
<td>-1.8869</td>
<td><strong>-0.0487</strong></td>
<td><strong>1.001</strong></td>
<td><strong>1.9402</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>CCRM</td>
<td>0.0001</td>
<td>0.0205</td>
<td><strong>0.0001</strong></td>
<td>0.0019</td>
<td>0.0356</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0184</strong></td>
<td>0.0019</td>
<td>0.0013</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>TS</td>
<td>0.0027</td>
<td>0.0121</td>
<td>429.39</td>
<td><strong>0.0001</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td></td>
<td>MTS</td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0128</strong></td>
<td><strong>0.0000</strong></td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td>Bias²</td>
<td>CCRM</td>
<td><strong>0.0003</strong></td>
<td><strong>0.0206</strong></td>
<td><strong>0.0020</strong></td>
<td>0.0020</td>
<td>0.0358</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td><strong>0.0004</strong></td>
<td><strong>0.0185</strong></td>
<td>0.0058</td>
<td>0.0014</td>
<td>0.0388</td>
</tr>
<tr>
<td></td>
<td>TS</td>
<td>4.7210</td>
<td>7.3450</td>
<td>1.5e+06</td>
<td>0.9674</td>
<td>2.0540</td>
</tr>
<tr>
<td></td>
<td>MTS</td>
<td>0.0021</td>
<td><strong>0.0061</strong></td>
<td>0.0488</td>
<td><strong>0.0007</strong></td>
<td><strong>0.0019</strong></td>
</tr>
</tbody>
</table>
substantially lower losses than those generated by CCRM and CRM as we have seen in Table 1.21 (lower panel). Thus, the misspecification of the multivariate density does not seem to affect greatly the performance of TS and MTS. When the process has low persistence (DGP2), no method seems to deliver overall unbiased estimates, and the problem of the estimation of the constant is severe. Note that the design of DGP2 represents the worst scenario because, by construction, the intervals are very tight, the specification of the conditional means deliver very small values around zero, so that the regressor $\lambda_{t-1}$ carries all the weight to estimate fitted intervals with the right order. Yet TS delivers the smallest losses as we have seen in Table 1.22 (lower panel). In DGP3 and DGP4, Table 1.28 the observability restriction is non-binding. When the process has high persistence, TS and MTS are superior performers, they deliver unbiased estimates with the lowest mean-square error. CCRM and CRM produce highly biased estimates. When the process has low persistence, MTS is the best performer overall.

In summary, evaluating the estimation performance of the four methods, we reach similar conclusions as those when we evaluate their goodness of fit. Even under misspecification of the multivariate density of the errors, if there is high persistence in the conditional mean, whether the observability restriction is binding or not, TS and MTS are superior estimation techniques. If the persistence is low and the observability restriction is non-binding, we recommend MTS, even with a misspecified density. Only when the persistence is low and the observability restriction is binding, the misspecification of the density may play a role on estimating the right dynamics but yet TS and MTS are not dominated by the competing methods and they offer the advantage of preserving the natural order of an interval.
1.6.3 In-Sample Evaluation Criteria: S&P 500 Daily Low/High Interval Returns

In Table 1.29 we compare the performance of the different estimation techniques in the modelling of the interval time series of the daily low/high returns to the SP500 Index.
Table 1.29: Methodology Evaluation for S&P500 Daily Low/High Interval Returns

<table>
<thead>
<tr>
<th>Period</th>
<th>RMSE</th>
<th>CR &amp; ER</th>
<th>MLF</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>p = 1</td>
<td>p = 2</td>
</tr>
<tr>
<td><strong>Unstable Period</strong> (2007/1/1-2011/4/29); BINDING observability restriction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>1.1541</td>
<td>0.9990</td>
<td>0.6811</td>
<td>0.5939</td>
</tr>
<tr>
<td>CRM</td>
<td>1.1541</td>
<td>0.9990</td>
<td>0.6811</td>
<td>0.5939</td>
</tr>
<tr>
<td>TS</td>
<td>1.1356</td>
<td>0.9828</td>
<td>0.6810</td>
<td>0.5974</td>
</tr>
<tr>
<td>MTS</td>
<td>1.1379</td>
<td>0.9831</td>
<td>0.6824</td>
<td>0.5952</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>1.8625</td>
<td>1.8329</td>
<td>0.9557</td>
<td>0.3630</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>2.0469</td>
<td>2.0246</td>
<td>0.9671</td>
<td>0.3399</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>2.2210</td>
<td>2.2064</td>
<td>0.9735</td>
<td>0.3272</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>2.6042</td>
<td>2.5989</td>
<td>0.9857</td>
<td>0.2921</td>
</tr>
<tr>
<td><strong>Stable Period</strong> (2004/1/1-2007/1/1); NON-BINDING observability restriction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCRM</td>
<td>0.4146</td>
<td>0.3958</td>
<td>0.7177</td>
<td>0.6427</td>
</tr>
<tr>
<td>CRM</td>
<td>0.4146</td>
<td>0.3958</td>
<td>0.7177</td>
<td>0.6427</td>
</tr>
<tr>
<td>TS</td>
<td>0.4123</td>
<td>0.3914</td>
<td>0.7184</td>
<td>0.6466</td>
</tr>
<tr>
<td>MTS</td>
<td>0.4129</td>
<td>0.3942</td>
<td>0.7173</td>
<td>0.6443</td>
</tr>
<tr>
<td>GARCH-N (99%)</td>
<td>0.6368</td>
<td>0.6240</td>
<td>0.9473</td>
<td>0.4410</td>
</tr>
<tr>
<td>GARCH-N (99.5%)</td>
<td>0.7025</td>
<td>0.6908</td>
<td>0.9621</td>
<td>0.4143</td>
</tr>
<tr>
<td>GARCH-T (99%)</td>
<td>0.7415</td>
<td>0.7302</td>
<td>0.9686</td>
<td>0.4002</td>
</tr>
<tr>
<td>GARCH-T (99.5%)</td>
<td>0.8599</td>
<td>0.8499</td>
<td>0.9818</td>
<td>0.3621</td>
</tr>
</tbody>
</table>
As in the simulation sections, we consider the following RMSE, CR & ER, MLF, and MDE functions. The upper panel shows the results for the unstable period (2007-2011) when the observability restriction is binding, and the lower panel for the stable period (2004-2007) when the observability restriction is non-binding. Overall and across panels, the estimation of a location-scale model, either with normal or Student-t errors, is not satisfactory, as the RMSE, MLF, and MDE losses are the largest among all methods. The location-scale model seems to provide slightly better CR & ER rates. From the estimation of the model in Section 5 of the manuscript, we know that the persistence in the conditional mean in both periods is low. Thus, in the unstable period, we are dealing with a scenario similar to the simulated DGP2, and in the stable period, similar to the simulated DGP4. Furthermore, we observe similar results to those provided by the simulations in Tables 1.22 and 1.24. In the unstable period, when the observability restriction is binding, TS and MTS provide the smaller losses; and in the stable period, when the observability restriction is non-binding, the losses of TS and MTS are equivalent to those of CCRM and CRM, as the restrictions become lax.

1.7 Conclusion

In many disciplines, massive data sets are released in an aggregate format such as intervals or histograms. Most of the current literature on the analysis of interval-valued data has focused on fitting classical regression models to the lower and upper bounds of the intervals, or to the center and radius. However, the bounds of the interval have a natural order that has not been taken into account in the estimation of the regression model, and as a result, it is very possible that for some observations the fitted lower bound could be larger than the fitted upper bound. In this paper, we have constrained
the regression model in such a way that a reversal of the order of the bounds will never happen. The constraint is probabilistic in nature as the errors of the model must come from a truncated bivariate probability density function to guarantee the natural order of the interval. The truncation has several consequences for the estimation of the model. Even when the regression model is linear, a maximum likelihood estimator will be nonlinear and difficult to compute. If we were to apply OLS estimation, the estimator would not be consistent because the truncation makes the error term correlated with the regressors. To solve both predicaments, we have proposed a two-step estimation procedure that is very easy to implement and that delivers consistent estimators of the parameters of the model. The two-step procedure consists of a maximum likelihood estimator in the first step and either least-squares estimation or minimum distance estimation in the second step. The minimum distance estimator is a neat solution when there is substantial multicollinearity as it is able to identify all parameters of the model regardless of how large or small the truncation of the density is.

We have analyzed the daily interval of low/high returns to the SP500 index and showed two instances, one in which the truncation is minimal and another in which the truncation is severe, to highlight the value of implementing the proposed two-step estimator.

We have shown that our estimators are superior over the existing approaches. We have examined several goodness-of-fit measures across methodologies and concluded that even when the observability restriction is non-binding, it pays off to implement our estimators because their losses are smaller than those from competing methodologies. We have also examined the biases and mean-squared errors of our estimators with those of other methods and our conclusions remain unchanged. Even under misspecification of the multivariate density of the errors, when there are relevant dynamics in the conditional
mean of the model, our estimators are superior.


Chapter 2

Modified Multivariate Adaptive Regression Splines

2.1 Introduction

Spline approaches are among the most frequently used nonparametric methods that can be applied to least squares regression. Such nonparametric estimation is robust to the misspecification of functional form, which is a constant concern for parametric estimation. Popular methods involved with splines include regression splines, smooth splines, and penalized splines etc. The popularity of spline methods is due to several advantages: they are global in implementation, easy in imposing constraints on resulting estimate, straightforward least squares solutions, and fast to compute. However, several disadvantages limit their usability in econometric modeling. For example, as the number of regressors increases (linearly), the number of total coefficients of multivariate (tensor-product) spline bases grows exponentially. With such curse of dimensionality, spline
methods may not be robust and deliver fragile estimation results, limiting the use to the analysis in which sample size is large even with a few regressors. To alleviate the dimensionality problem, additivity and/or partial linearity are imposed on the functional form. However, imposing such restrictions undermines the flexibility of nonparametric model. In this paper, borrowing the idea of adaptive splines selection in Friedman (1991), we propose a modified multivariate adaptive regression splines (MMARS) method that is more flexible to model multivariate nonlinear relationship. The adaptive procedure constructs and selects relevant univariate and multivariate spline bases. This procedure is hierarchical in the sense that the higher dimensional splines are constructed and selected based on the existing lower dimensional splines in the regression. By restricting the maximal dimension of splines in the procedure, we can control the complexity of the regression spline estimates. If no interactions among the splines from different regressors are allowed (i.e., the maximal dimension of spline is one), the model becomes additive as a special case.

Our proposed modified multivariate adaptive regression splines (MMARS) is different from Friedman (1991)’s original MARS version in knots specification and type of splines. In MARS, the knots of splines consist of all the data points, and therefore, when the sample size is large, MARS can be very slow due to the large number of candidate univariate basis functions. In our MMARS, knots are simply equally spaced, and the number of candidate univariate basis functions is much smaller than that of MARS especially when the sample size is large. In addition, a univariate spline basis in MARS is a pair of piecewise linear function associate with each knot, while ours is B-splines, which is more tractable in computational implementation and theoretical deductions.

The investigation of theoretical properties of methods based on polynomial splines (B-
splines as a special case) has been an active area of research for three decades, and both global and local asymptotic properties have been thoroughly studied in existing literature. Specifically, for multivariate tensor-product polynomial splines, global rates of convergence of spline estimates were examined in Stone (1994). The local asymptotic properties are first studied in Zhou, Shen and Wolfe (1998) for univariate spline regression. Alternatively, Huang (2003) proposed the geometric approach to study the local asymptotics of polynomial spline regression. The general result applies to a variety of types of splines, including univariate splines, tensor product splines, and bivariate or multivariate splines on triangulations. In this paper, we follow the geometric approach to derive the local asymptotics of our adaptive regression spline estimator. The adaptive selection procedure is hierarchical, and is a greedy algorithm. Therefore, it may stop at local optima and fail locating the global optimum. We establish the asymptotic normality of our adaptive estimator, and obtain the optimal convergence rate that it can possibly achieve. The optimal convergence rate depends on the order ratio of the number of selected spline basis functions to the total potential ones. In the Monte Carlo simulation, we compare our adaptive estimator with classical regression splines in various settings of data generating processes. The results show that our estimator outperforms others by producing smaller average mean squared errors when multivariate covariates and/or data clustering appears.

The paper is organized as follows. Section 2 gives a brief introduction to splines. Section 3 introduces the general setting of the model and the adaptive selection procedure. The Asymptotic Properties of our proposed adaptive regression splines are investigated in Section 3. Some simulation studies and an empirical application are given in Section 4 and 5 respectively. Finally, Section 6 concludes.
2.2 Model

2.2.1 General Setting

We consider a nonparametric regression model of the general form

\[ y = m(X) + \varepsilon \]

where \( X \) is a \( 1 \times d \) vector \((x_1, \cdots, x_k, \cdots, x_d)\) with \( d \) regressors, and \( m(X) = E(y|X) \) is the unspecified conditional expectation function, and \( \varepsilon \) is the error term. Our primary interest is in estimating \( m(X) \). We consider least squares estimates over polynomial spline spaces and refer to the corresponding estimation procedures as polynomial spline regression. Broadly speaking, a polynomial spline is referred to any possibly smooth, piecewise polynomial function. In this paper, we focus on one widely used type of spline functions known as B-spline. Theoretically, a linear combination of B-splines can be used to uniquely represent any spline function given the same degree, smoothness, and domain partition. Besides, B-splines are also fast to compute and evaluate in computer.

The polynomial spline space consists of all the spline functions generated by the linear combination of multivariate tensor product spline bases associated with the \( d \) regressors, which is also called as estimation space. The Multivariate tensor product splines are built from univariate ones by multiplication. For each regressor \( X_k, k = 1, \cdots, d \), we build the univariate candidate spline bases \( \{h_{jk}^{(1)}(x_k)\}_{j=1}^{J_k} \), and the total number of univariate spline bases for all \( d \) regressors is \( \sum_{k=1}^{d} J_k \). As a convention, we use \( i, k \) and \( j \) to index the observation, regressor, and spline basis respectively, and \( J_k \) to denote the total number of univariate spline bases for \( k \)-th regressor. Then, we build
bivariate tensor product bases \( \{ h_{j_1 k_1, j_2 k_2}^{(2)}(x_{k_1}, x_{k_2}) \} \) based on the univariate spline bases by multiplications,

\[
h_{j_1 k_1, j_2 k_2}^{(2)}(x_{k_1}, x_{k_2}) = h_{j_1 k_1}^{(1)}(x_{k_1}) \times h_{j_2 k_2}^{(1)}(x_{k_2})
\]

with \( 1 \leq j_1 \leq J_{k_1}, \ 1 \leq j_2 \leq J_{k_2}, \) and \( 1 \leq k_1 \neq k_2 \leq d \). Therefore, the total number of bivariate spline bases is \( \sum_{1 \leq k_1 \neq k_2 \leq d} J_{k_1} J_{k_2} \). Recursively, we can build \( m \)-variate spline bases \( \{ h_{j_1 k_1, \ldots, j_m k_m}^{(m)}(x_{k_1}, \ldots, x_{k_m}) \} \) as

\[
h_{j_1 k_1, \ldots, j_m k_m}^{(m)}(x_{k_1}, \ldots, x_{k_m}) = \prod_{l=1}^{m} h_{j_l k_l}^{(1)}(x_{k_l}).
\]

with \( 1 \leq j_l \leq J_{k_l} \) with \( 1 \leq l \leq m \) and \( m \leq d \). Total number of \( m \)-variate spline bases is \( \binom{d}{m} \prod_{l=1}^{m} J_{k_l} \). Finally, the \( d \)-variate tensor product spline bases are \( \{ h_{j_1, \ldots, j_d}^{(d)}(x_1, \ldots, x_d) \} \) in which

\[
h_{j_1, \ldots, j_d}^{(d)}(x_1, \ldots, x_d) = \prod_{k=1}^{d} h_{j_k k}^{(1)}(x_k)
\]

with \( 1 \leq j_1 \leq J_1, \ldots, 1 \leq j_d \leq J_d \). The total number of \( d \)-variate spline bases is \( \prod_{l=1}^{d} J_{k_l} \). If we assume that the number of univariate splines are the same for all regressors, i.e., \( J_k = J_1 \) for all \( k = 1, \ldots, d \), the total number of splines is

\[
(J_1 + 1)^d - 1 = J_1^d + \binom{d}{d-1} J_1^{d-1} + \cdots + \binom{d}{k} J_1^k + \cdots + d J_1.
\]

However, since in such construction of tensor product splines, lower dimensional splines can always be expressed as a linear combination of higher dimensional ones, the total number of linearly independent bases in the polynomial spines space is \( J_1^d \), denoted as
The curse of dimensionality is the fact the dimension of polynomial splines space $J_n = J_1^d$ tends to increase exponentially with respect to the increase of the number of regressors $d$, and it is often impractical to include all the spline bases in the spline regression even when $d$ is moderate. Borrowing the idea of the adaptive selection procedure in Friedman (1991), our paper tries to overcome such problem by choosing those relevant and useful splines from the bases by generalized cross-validation (GCV). The adaptive splines selection reduces the number of splines involved in the polynomial spline regression while retaining the flexibility of the model.

### 2.2.2 Adaptive Selection Procedure

The adaptive spline selection procedure includes two steps: a forward step and a backward step. The forward step is a “bottom up” model building procedure in which relevant splines are added into the model adaptively according to the following steps.

1. At the beginning of the $M$-th stage ($M \geq 0$), suppose the model has the form,

   $$ m(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X_m) $$

   where $h_m(X_m)$ are the spline functions with $d_m$ regressors in the polynomial splines space constructed adaptively, and $X_m$ is $1 \times d_m$ vector $[X_{k_1}, \cdots, X_{k_{d_m}}]$. When $M = 0$, the model only contains the constant function $h_0(X) = 1$. Therefore, we build the model from the simplest constant mean model.

2. During the $M$-th stage, we construct the multiplication of $h_m(X_m)$ in the current
model with all the univariate spline bases \( \{ h^{(1)}_{jk}(x_k) \}_{j=1}^J \) in all possible combinations excluding the univariate splines that are associated with the regressors \( X_m \) in \( h_m(X_m) \) (i.e., \( 1 \leq m \leq M \), \( 1 \leq j \leq J \), and \( 1 \leq k \leq d \) while \( k \notin \{ k_1, \ldots, k_{d_m} \} \)) as the potential spline functions to be added in the regression model.

3. At the end of the \( M \)-th stage, we **select** the spline function \( h_l(X_l) \cdot h^{(1)}_{jk}(x_k) \) constructed in the former step that produces the largest improvement of model fitting based on generalized cross-validation. The selected spline function (denoted as \( h_{M+1}(X_{M+1}) \)) is added to the polynomial splines regression model

\[
m(X) = \beta_0 + \sum_{m=1}^{M+1} \beta_m h_m(X_m),
\]

and the procedure goes to the beginning of the \((M+1)\)-th stage.

Such forward step continues until there is no more improvement in GCV or the number of splines in the model reaches a maximal value.

At the end of the forward step, we may have a large model that overfits the data. Therefore, a backward deletion procedure is applied to prune off some irrelevant regressors. Those spline functions whose removal do not worsen model fitting based on GCV is deleted from the model. After this step, we reach the final output of least squares estimation based on the adaptive selection procedure which selects \( J^*_n \) non-constant spline functions from the spline bases

\[
y = \hat{\beta}_0 + \sum_{m=1}^{J^*_n} \hat{\beta}_m h_m(X_m) + u,
\]

where the adaptive splines estimator is \( \hat{\beta}_0 + \sum_{m=1}^{J^*_n} \hat{\beta}_m h_m(X_m) \).
The adaptive selection procedure in this paper is very similar to MARS in Friedman (1991). However, the differences are mainly in several aspects. First, in MARS, the knots of splines consist of all the data points, while ours is equally spaced knots. Second, the spline basis in MARS is a pair of piecewise linear functions for each knots, and ours is cubic splines, which is more tractable in implementations and theoretical deductions. Therefore, given the specification of spline basis, MARS has $2dn$ univariate basis functions, ours has only $dJ$ univariate bases. When the sample size $n$ is large, the MARS is very slow due to the large size of univariate bases, from which the spline functions in the regression are constructed and selected adaptively.

The splines regression with adaptive selection procedure has both advantages and disadvantages. The advantages are two folded. First, the upper limit on the number of interaction of regressors in the construction of spline functions can be set. By this way, we can control the complexity of the splines regression model. For example, if the maximal number of interaction is one, the adaptive selection procedure only select spline functions from the univariate bases, and does not consider any multivariate tensor product splines. Second, such adaptive selection procedure aviods the search over an exponentially growing space of alternatives, since the total number of linearly dependent spline bases is $J^d_1$ and the total number of sub-models which include non-empty subsets of total spline bases is $2^J_1 - 1$. It is impractical to search over such a huge amount of models. Therefore, by reducing the number of splines in the regression, the adaptive selection help the splines regression model to alleviate the curse of dimensionality, since only those relevant splines that improve model fitting in terms of GCV are included in the model.
On the other hand, the adaptive selection procedure also suffers from some disadvantages. This modeling strategy is hierarchical in the sense that higher order interaction splines must be built upon lower order ones, which may not be true in reality. Therefore, as this procedure is a greedy algorithm and not an exhaustive search, the procedure may only stop at some local optima and fail locating the global optimum.

2.3 Asymptotic Properties

Let $G = G_n$ denote the linear space spanned by all the spline bases with finite dimension $J_n$. The adaptive estimate $\hat{m}(X; G^*)$ of $m(X)$ is defined as the element $g \in G^* \subseteq G$ such that minimizes certain information criterion (generalized cross-validation). Set $\tilde{m}(X; G^*) = E(\hat{m}|X; G^*)$, where $G^*$ only includes the spline bases after adaptive selection procedure. Then, we have the following decomposition

$$\hat{m}(X; G^*) - m(X) = [\hat{m}(X; G^*) - \tilde{m}(X; G^*)] + [\tilde{m}(X; G^*) - m(X)]$$  \hspace{1cm} (2.3.1)

where the first bracket on the right hand side of the equation is the variance term and the second bracket is the bias term. We first prove asymptotic normality of the variance term, and then analyze the bias term. Finally, we obtain the optimal convergence rate the the adaptive estimator can possibly achieve.

**Theorem 2.3.1.** (Asymptotic Normality of Variance Term)

Assume that (i) the polynomial spline spaces $G$ and $G^*$ are theoretical identifiable in the sense that $g \in G$ ($g^* \in G^*$) and $\|g\| = 0$ ($\|g^*\| = 0$) together imply that $g = 0$ ($g^* = 0$); (ii) the density function $p_x$ of $X$ is bounded away from zero and infinity; (iii) the partition $\Delta$ and $\Delta^*$ associated with polynomial spline spaces $G$ and $G^*$...
satisfies the regularity condition (Condition A.2 in Huang (2003)) on their supports respectively; (iv) \( \lim_{n \to \infty} J_n \log n/n = 0 \); (v) homoskedastic error \( \sigma^2(x) = \sigma^2 \) and \( \lim_{\lambda \to \infty} E(\varepsilon^2\mathbf{1}[|\varepsilon| \lambda]|X = x) = 0 \). Then, given the polynomial splines space \( G^* \subseteq G \) after adaptive selection, for \( x \in \mathcal{X} \),

\[
\frac{\hat{m}(x;G^*) - \tilde{m}(x;G^*)}{\sqrt{\text{var}(\hat{m}(x,G^*)|X)}} \xrightarrow{d} N(0,1)
\]

where

\[
\text{var}(\hat{m}(x;G^*)|X) = \frac{1}{n} \mathbf{B}'(x)[E_n(\mathbf{B}(X)\mathbf{B}'(X))]^{-1}\mathbf{B}(x)\sigma^2.
\]

Consequently, we have

\[
\sup |\hat{m}(x;G) - \tilde{m}(x;G^*)| = O_p \left( \sqrt{\frac{J_n}{n}} \right) \tag{2.3.2}
\]

In Assumption (iii), the regularity conditions of partition \( \Delta \) is as follows,

**Condition 1.** *(Condition A.2 Huang (2003))*

(i) (No mixture) Given any distinct \( \delta, \delta' \in \Delta \), the closures of \( \delta \) and \( \delta' \) are disjoint or intersect in a common vertex, edge, face and so on;

(ii) (No degeneration) The ratio of the sizes of inscribed and circumscribed balls of each \( \delta \in \Delta \) is bounded away from zero; (iii) (Quasi-uniform)

\[
\frac{\max\{\text{diam}(\delta) : \delta \in \Delta\}}{\min\{\text{diam}(\delta) : \delta \in \Delta\}} \leq C < \infty.
\]

The proof of this theorem is similar to Theorem 3.1 in Huang (2003), in which geometric
approach of Hilbert Space is used to obtain a quite general understanding of the local asymptotics of polynomial spline regression. Assumption (i) means that the estimation space $G$ is a Hilbert Space equipped with the theoretical inner product and $G^* \subseteq G$ is its subspace. Assumption (ii) - (iv) gives sufficient conditions for the theoretical norm to be close to the empirical norm uniformly over the estimation spaces in the sense of $\sup_{g \in G} \|g\|_n/\|g\| - 1 = o_p(1)$ and $\sup_{g^* \in G^*} \|g^*\|_n/\|g^*\| - 1 = o_p(1)$, so that the estimation space $G$ is a Hilbert Space equipped with the empirical inner product and $G^* \subseteq G$ is its subspace. Therefore, Assumption (i) - (iv) yield sufficient conditions for the least squared estimate to be well defined, and the homoskedasticity in Assumption (v) is not critical and can be generalized into heteroskedastic case.

The bias term $\tilde{m}(x; G^*) - m(x)$ in (2.3.1) can be further decomposed as

$$\tilde{m}(x; G^*) - m(x) = [\tilde{m}(x; G^*) - \tilde{m}(x; G)] + [\tilde{m}(x; G) - m(x)].$$

The second term in (2.3.3) is the bias induced by approximating unknown conditional mean function with the function in polynomial spline space $G$. The asymptotic behavior of the second term is well studied in the approximation theory, see Schumaker (1981), Stone (1982), DeVore and Lorentz (1993). A function $f$ on $\mathcal{X}$ is said the satisfy a Hölder condition with exponent $\beta$ if there is a positive number $\gamma$ such that $|f(x_2) - f(x_1)| \leq \gamma |f(x_2) - f(x_1)|^\beta$ for $x_1, x_2 \in \mathcal{X}$; here $|x|$ is the Euclidean norm of $x = (x_1, \ldots, x_d) \in \mathcal{X}$.

Given a $d$-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers, set $[\alpha] = \alpha_1 + \cdots + \alpha_d$ and let $D^\alpha$ denote the differential operator defined by

$$D^\alpha = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$
Let $k$ be a nonnegative integer and set $p = k + \beta$. A function on $\mathcal{X}$ is said to be $p$-smooth if it is $k$ times continuously differentiable on $\mathcal{X}$ and $D^\alpha$ satisfies a Hölder condition with exponent $\beta$ for all $\alpha$ with $[\alpha] = k$. Now suppose the unknown conditional mean function $m(X)$ is $p$-smooth, we have

$$\sup_x |\tilde{m}(x; G) - m(x)| = O_p(J_n^{-p/d}) \tag{2.3.4}$$

where $\tilde{m}(X; G)$ is the function in tensor product $B$-spline space $G$ with $J_n$ number of spline bases. See the proof in Chapter 12 of Schumaker (1981).

The first term on the right hand side of the equation (2.3.3) is the bias induced by projecting the conditional mean function on $G^*$ instead of $G$, and $G^*$ is a subspace of $G$. Suppose $\{\phi_j(x)\}_{j=1}^{J_n}$ is the orthonormal basis of $G$. The subspace $G^*$ is spanned by a subset of $J^*_n$ basis $\{\phi_{j^*}(x)\}$, and $G' = G \setminus G^*$ is spanned by the rest of $(J_n - J^*_n)$ basis $\{\phi_{j'}(x)\}$. The empirical inner product is $\langle m, \phi_j \rangle_n = \sum_{i=1}^n m(x)\phi_j(X_i)/n$. We assume that by generalized cross validation, the empirical inner product of the selected spline bases $\phi_{j^*}$ in $G$ and conditional mean function $m(x)$ is $O_p(1)$, i.e., $\langle m, \phi_{j^*} \rangle_n = \sum_{i=1}^n m(x)\phi_{j^*}(X_i)/n = O_p(1)$; and the empirical inner product of unselected spline bases $\phi_{j'}$ and $m(x)$ is $O_p(n^{-\ell})$, i.e., $\langle m, \phi_{j'} \rangle_n = \sum_{i=1}^n m(x)\phi_{j'}(X_i)/n = O_p(n^{-\ell})$, where $\ell > 0$. Then, based on the inequality (2.3.5),

$$|\tilde{m}(x; G^*) - \tilde{m}(x; G')| = |\tilde{m}(x; G')|$$

$$= \left| \sum_{j'} \langle m, \phi_{j'} \rangle_n \phi_{j'}(x) \right|$$

$$\leq \sum_{j'} |\langle m, \phi_{j'} \rangle_n| \phi_{j'}(x)$$

$$\leq C \cdot (J_n - J^*_n) \sup_{j' \in G'} |\langle m, \phi_{j'} \rangle_n| \tag{2.3.5}$$
we have that

\[
\sup |\tilde{m}(x; G^*) - \tilde{m}(x; G)| = O_p \left( \frac{J_n - J_n^*}{n^{\ell}} \right).
\]

(2.3.6)

Based on (2.3.2), (2.3.4) and (2.3.6), we derive the convergence rate of \( \hat{m}(X; G^*) \) to \( m(X) \) as

\[
\sup |\hat{m}(x; G^*) - m(x)| \\
\leq \sup |\hat{m}(x; G^*) - \tilde{m}(x; G^*)| + \sup |\tilde{m}(x; G^*) - \tilde{m}(x; G)| + \sup |\tilde{m}(x; G) - m(x)| \\
= O_p \left( \sqrt{\frac{J_n^*}{n}} \right) + O_p \left( \frac{J_n - J_n^*}{n^{\ell}} \right) + O_p(J_n^{-p/d}).
\]

Therefore we can obtain the optimal convergence rate by balancing the three terms, and have the following theorem.

**Theorem 2.3.2. (Optimal convergence rate of adaptive estimator)**

Suppose the adaptive estimator obtains the optimal convergence rate by generalized cross validation and assume that, (i) the numbers of total and selected spline bases are \( J_n \) and \( J_n^* \) respectively, satisfying \( J_n^*/J_n = O_p(n^{-r}) \); and that (ii) by generalized cross validation, the empirical inner product of the selected spline bases \( \phi_j \) in \( G \) and conditional mean function \( m(x) \) is \( O_p(1) \); and the empirical inner product of unselected spline bases \( \phi_j' \) and \( m(x) \) is \( O_p(n^{-\ell}) \), \( \ell > 0 \);

then,

(a) if \( \frac{1 + r}{\ell} \leq \frac{2p + d}{p + d} \) and \( r < \frac{d}{2p} \), the optimal convergence rate is obtained by balancing the variance and the second bias term

\[
\sup |\hat{m}(x; G^*) - m(x)|^2 = O_p \left( n^{-(1+r)2p/(2p+d)} \right), \quad J_n = O_p \left( n^{(1+r)d/(2p+d)} \right);
\]
(b) if \( \frac{1 + r}{\ell} > \frac{2p + d}{p + d} \) and \( \ell < \frac{p + d}{2p} \), the optimal convergence rate is obtained by balancing the first and second bias term

\[
\sup |\hat{m}(x; G^*) - m(x)|^2 = O_p \left( n^{-2p\ell/(p+d)} \right), \quad J_n = O_p \left( n^{\ell d/(p+d)} \right).
\]

Note that in (a) the optimal convergence rate is an increasing function of \( r \) which controls the significance of dimension reduction in the adaptive spline estimation. It implies that the more significantly the adaptive splines estimator reduces the dimension of the model, the more gain in convergence rate the adaptive splines estimator achieves. In (b), the optimal convergence rate is also an increasing function of \( \ell \) which captures how fast the irrelevant spline bases vanish in the model. It also indirectly controls the significance of dimension reduction in the adaptive spline estimation. Therefore, it also suggests that the gain in convergence rate is larger when the dimension reduction is more significant. Such implication is also suggested by the results of Monte Carlo simulation.

### 2.4 Monte Carlo

In the section, we compare our adaptive spline estimator with classical regression splines in both univariate and multivariate cases when the model includes up to 3 independent variables.

#### 2.4.1 Data Generating Processes

In the data generating processes, we consider up to three independent random variables \( X_1, X_2, \) and \( X_3 \) that come from either uniform \([0, 1]\) distribution or mixed normal
distribution, i.e.,

\begin{align*}
\text{DIST-u:} & \quad X_1, X_2, X_3 \sim U[0, 1], \\
\text{DIST-mm:} & \quad X_1, X_2, X_3 \sim \begin{cases} 
N(0.25, (0.25/4)^2) & \text{with probability 0.5,} \\
N(0.75, (0.25/4)^2) & \text{with probability 0.5.}
\end{cases}
\end{align*}

The random variables $X_1$ and $X_2$ with mixed normal distribution is used to capture the presence of clustering data in reality. The dependent variable $Y$ is generated from the following equations that include one, two or three independent variables,

\begin{align*}
\text{EQ1:} & \quad Y = 10 \sin(X_1^2) + \varepsilon, \\
\text{EQ2:} & \quad Y = 10 \sin(X_1^2 + X_2^2) + \varepsilon, \\
\text{EQ3:} & \quad Y = 10 \sin(X_1^2 + X_2^2 + X_3^2) + \varepsilon,
\end{align*}

where the error term $\varepsilon$ follows normal distribution $N(0,1)$. Therefore, in total we have 6 DGPs. They are, (i) univariate cases: EQ1/uniform, EQ1/mixed normal; (ii) bivariate cases: EQ2/uniform, EQ2/mixed normal; and (iii) trivariate cases: EQ3/mixed normal, EQ3/uniform. For each DGP, the sample size is 500, and we replicate each DGP with 1000 times. We perform our adaptive spline estimation, and classical regression splines estimation with generalized cross validation for the purpose of comparison.

### 2.4.2 Evaluation

To compare the performance of our adaptive spline estimator with that of classical regression spline estimator, we use Average Mean Squared Error (AMSE) as the criterion. For each DGP, we perform our adaptive spline estimation, and classical regression splines
estimation with generalized cross validation for the purpose of comparison.

<table>
<thead>
<tr>
<th>DGP</th>
<th>Classical Splines</th>
<th>Adaptive Splines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( J_n )</td>
<td>AMSE</td>
</tr>
<tr>
<td><strong>univariate cases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ1/uniform</td>
<td>4.753</td>
<td><strong>0.0143</strong></td>
</tr>
<tr>
<td>EQ1/mixed normal</td>
<td>4.726</td>
<td><strong>0.0142</strong></td>
</tr>
<tr>
<td><strong>bivariate cases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ2/uniform</td>
<td>16.644</td>
<td><strong>0.0332</strong></td>
</tr>
<tr>
<td>EQ2/mixed normal</td>
<td>16.484</td>
<td><strong>0.0332</strong></td>
</tr>
<tr>
<td><strong>trivariate cases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ3/uniform</td>
<td>64.000</td>
<td><strong>0.1268</strong></td>
</tr>
<tr>
<td>EQ3/mixed normal</td>
<td>64.037</td>
<td><strong>0.1282</strong></td>
</tr>
</tbody>
</table>

Sample Size = 500, Simulation Times = 1000

Table 2.1: Model Comparison

Table 2.1 compares the estimation results of both spline methods in several aspects for six DGP settings. It includes the average total number of splines \( J_n \) for both splines estimations, the average selected number of splines \( J_n^* \) for adaptive splines estimation, and the average mean squared errors (AMSE) for both methods. The histograms of 1000 MSE values for each DGP from both splines estimations are drawn in Figure 2.1 and 2.2 respectively.

Here are the findings based on the comparison.

1. The curse of dimensionality that plagues in the classical regression splines is largely ameliorated in the adaptive splines estimation. As the sample size is fixed to be 500, when the number of regressors increase from one to three, the average effective splines in classical approach exponentially increases from 4 to 64, while that only increases linearly to 5 to 8 in adaptive approach which sharply reduces the dimensions of the models.

2. The AMSEs show the benefits from the reduction of dimensionality of the models.
The AMSE of classical regression splines estimation increases from approximately 0.01 to 0.13 for both uniform and mixed normal distribution of regressors; and the AMSE of adaptive splines estimation, however, only increase from approximately 0.01 to 0.03 and 0.06 for uniform and mixed normality distribution of regressors respectively. This finding is consistent with our theoretical conclusion the optimal convergence rate that the adaptive splines estimation converges quicker as the dimension reduction is more significant (i.e., as $r$ and $\ell$, controlling the significance of dimension reduction, become larger).

3. The horizontal comparison of the AMSEs in the table further shows that the more dimensions the adaptive splines reduces, the more gain in AMSEs the adaptive splines achieve. When, in the two univariate cases, the average number of selected splines $\overline{J}_n^r$ does not differ too much from that of total splines $\overline{J}_n$, there is no significant gains of AMSE in the adaptive approach. In the bivariate cases, the adaptive approach has more dramatical dimension reduction on average in the DGP with uniform distribution than with mixed normal distribution, and there is significant AMSE gain in the former DGP and not in the latter. Similar results also appear in the trivariate cases.

4. Overall, the simulation study delivers an important message that with moderate sample size, classical splines approach is reliable only one or two regressors are involved, and that when more regressors are included, the estimation results from classical approach becomes fragile and unreliable. Dimension reduction of spline regression model based on adaptive selection help to robustify the estimation results.
Figure 2.1: Histograms of MSE for Splines with Uniformity in DGPs
Figure 2.2: Histograms of MSE for Splines with Mixed Normality in DGPs
2.5 Empirical Application

2.5.1 Data

We consider a multivariate economic application, which investigates the effect of public capital stock on state productivity growth in the United States. The data set is a panel of 48 observations from 1970 to 1986 (17 years) with 816 observations in total. The variables includes gross state product ($gsp$), private capital stock ($pc$), employment level ($emp$), unemployment rate ($unemp$), public capital stock ($pcap$) and its three components including highway and streets ($hwy$), water and sewer facilities ($water$), and other public buildings and structures ($util$). This data is used in Baltagi and Pinnoi (1995) and available in the “plm” package of R software.

2.5.2 Results

In the multivariate case, we apply our proposed adaptive splines estimation to the analysis of the effect of public capital stock on gross state product for the 48 continental states in the United States. The nonparametric regression model of interest is

\[ gsp_{it} = m(X_{it}) + \varepsilon_{it} \]

\[ X_{it} = (pcap_{it}, pc_{it}, emp_{it}), \]

in which the panel data are pooled together and treat the data set as cross-sectional. The adaptive splines estimation is then applied to the nonparametric model. Using linear combination of spline functions to approximate the unknown conditional mean.
function $m(X_{it})$, we have

$$gsp_{it} = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X_{mit}) + u_{it}$$

$$B_{M+1}(X_{it}) = \begin{bmatrix} 1 & h_1(X_{1it}) & \cdots & h_M(X_{Mit}) \end{bmatrix}_{1 \times (M+1)}$$

$$\beta = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_M \end{bmatrix}'_{(M+1) \times 1}$$

where $h_m(\cdot)$ and $X_{mit}$ denote the constructed and selected $m$-th spline function and its associated variables. Based on the estimation results, we can use generalized cross validation to adaptively construct and select the spline functions that should be included in $B_{M+1}(X_{it})$. Table 2.2 reports the constructed and selected spline functions and their estimates of coefficients. In total, we select 15 spline functions, and the variables in each of the 15 spline functions are shown in the second column of the table (variables), and the third column (knots) reports the univariate spline bases from which the spline functions are constructed. The variables ($pcap$) and numbers in bold represents the selection of univariate spline bases of public capital stock. From the result of splines selection, the public capital stock variable enters into the model with through the interactions with other two variables (employment level and private capital stock).

Based on the splines estimation results, we can obtain the marginal effect of public capital stock on the gross state product by taking the derivative of the estimates $\hat{m}(X_{it}; G^*)$ with respect to $pcap_{it}$, i.e.,

$$\hat{\gamma}_{it} \equiv \frac{\partial \hat{m}(X_{it}; G^*)}{\partial pcap_{it}} = \sum_{m=1}^{J^*} \left( \frac{\partial h_m(X_{it})}{\partial pcap_{it}} \right) \hat{\beta}_m, \quad i = 1, \cdots, N, \ t = 1, \cdots, T.$$
Dependent Variable: log Gross State Product  
Sample Size: 816 (48 states, from 1970 to 1986)

<table>
<thead>
<tr>
<th>spline functions</th>
<th>variables</th>
<th>knots</th>
<th>estimates</th>
<th>std. error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp01</td>
<td>intercept</td>
<td></td>
<td>8.3843</td>
<td>0.0609</td>
<td>137.6158</td>
</tr>
<tr>
<td>sp02</td>
<td>emp</td>
<td>(1)</td>
<td>-0.5199</td>
<td>0.1346</td>
<td>-3.8630</td>
</tr>
<tr>
<td>sp03</td>
<td>emp</td>
<td>(4)</td>
<td>4.7099</td>
<td>0.0564</td>
<td>83.5791</td>
</tr>
<tr>
<td>sp04</td>
<td>emp</td>
<td>(2)</td>
<td>3.0629</td>
<td>0.0889</td>
<td>34.4529</td>
</tr>
<tr>
<td>sp05</td>
<td>emp</td>
<td>(3)</td>
<td>3.3200</td>
<td>0.1077</td>
<td>30.8362</td>
</tr>
<tr>
<td>sp06</td>
<td>emp, pc</td>
<td>(1,3)</td>
<td>14.2914</td>
<td>0.5166</td>
<td>27.6654</td>
</tr>
<tr>
<td>sp07</td>
<td>pc</td>
<td>(1)</td>
<td>1.0749</td>
<td>0.1148</td>
<td>9.3614</td>
</tr>
<tr>
<td>sp08</td>
<td>emp, pcap</td>
<td>(3,1)</td>
<td>-2.8210</td>
<td>0.2149</td>
<td>-13.1264</td>
</tr>
<tr>
<td>sp09</td>
<td>emp, pc, pcap</td>
<td>(1,3,2)</td>
<td>-16.7689</td>
<td>1.2926</td>
<td>-12.9734</td>
</tr>
<tr>
<td>sp10</td>
<td>emp, pc</td>
<td>(2,2)</td>
<td>-4.5308</td>
<td>0.3027</td>
<td>-14.9693</td>
</tr>
<tr>
<td>sp11</td>
<td>emp, pc, pcap</td>
<td>(2,2,2)</td>
<td>5.1153</td>
<td>0.5251</td>
<td>9.7417</td>
</tr>
<tr>
<td>sp12</td>
<td>emp, pcap</td>
<td>(4,1)</td>
<td>35.6787</td>
<td>4.3315</td>
<td>8.2371</td>
</tr>
<tr>
<td>sp13</td>
<td>emp, pc, pcap</td>
<td>(2,4,4)</td>
<td>12.0902</td>
<td>1.1190</td>
<td>10.8041</td>
</tr>
<tr>
<td>sp14</td>
<td>emp, pcap</td>
<td>(4,2)</td>
<td>-2.6585</td>
<td>0.5786</td>
<td>-4.5949</td>
</tr>
<tr>
<td>sp15</td>
<td>emp, pc</td>
<td>(1,1)</td>
<td>-0.6969</td>
<td>0.2448</td>
<td>-2.8469</td>
</tr>
</tbody>
</table>

Total Sum of Squares: 849.81, Residual Sum of Squares: 3.569  
R-Squared: 0.9958, Adj. R-Squared: 0.9775  
F-statistic: 13565.8 on 14 and 801 DF, p-value: <2.22E-16

Table 2.2: Estimation of Adaptive Regression Splines: Gross State Product
Then, we calculate the average marginal effect of public capital stock on the gross state product for each state and each year respectively, i.e.,

\[
\bar{\gamma}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{it}, \quad \text{for each state} \quad i = 1, \cdots, N
\]

\[
\bar{\gamma}_t = \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_{it}, \quad \text{for each year} \quad t = 1, \cdots, T
\]

\[
\bar{\gamma} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\gamma}_{it} \approx 0.10.
\]

In Figure 2.3, we plot the series of \(\bar{\gamma}_i\) and find that the average marginal effects of public capital stocks varies a lot across the 48 states. The three largest average marginal effects are in Texas, Louisiana, and South Carolina, and are around 0.4; the states with significant negative marginal effects of public capital stock include Wyoming, New Mexico, New Jersey, Montana, Florida and North Dakota; and other states have average marginal effects between 0.3 and \(-0.1\). Figure 2.4 plots the average marginal effects of public capital stock for each year. We find that on average in the United State the marginal effects are between 0.115 and 0.085 with a general decreasing trend. In addition, the three short period spikes correspond to the three business cycles dated by NBER during the time span of the data (November 1973 – March 1975, January 1980 – July 1980, and July 1981 – November 1982). Such correspondence between increase of marginal effect and business cycle is also shown for California (see Figure 2.5), and hints that the increasing public spending is an effective measure for the policymaker to overcome or alleviate economic downturn.
Figure 2.3: Average Marginal Effects of Public Capital Stocks on GSP for Each States

Figure 2.4: Average Marginal Effects of Public Capital Stocks on GSP for Each Year (the United States)
Figure 2.5: Marginal Effects of Public Capital Stocks on GSP for Each Year (California)

2.6 Conclusion

In this paper, we propose an adaptive spline estimator based on Friedman (1991)’s multivariate adaptive regression splines (MARS). The model takes the form of an expansion in the cross product spline basis functions, where the numbers of basis functions, the degree of product degree and knot locations are automatically selected adaptively by the data using generalized cross validation. Our proposed model is different from MARS in the fact that our model uses equally spaced knots and cubic B-spline basis functions while MARS uses all the data points and their associated piecewise linear function pairs as its potential knots and spline basis functions. With such modifications, our estimator is more tractable not only in computational implementations but in theoretical deductions as well. The adaptive selection procedure is hierarchical, and is a greedy algorithm. Therefore, it may stop at local optima and fail locating the global optimum.

We establish the asymptotic normality of our adaptive estimator, and obtain the optimal convergence rate that it can possibly achieve. It turns out that the optimal convergence
rate depends on the order ratio of the number of selected spline basis functions to the total potential ones.

In the Monte Carlo simulation, we compare our adaptive estimator with classical regression splines in the univariate, bivariate and trivariate data generating processes with uniformly and mixed normally distributed regressors. The results show that our estimator has more significant improvement upon classical regression splines by producing smaller average mean squared errors when multivariate covariates are in the DGP. We also apply our adaptive spline estimator to the study of the effect of public capital stock on the gross state product using the pooled panel data set in Baltagi and Pinnoi (1995).
Bibliography


Chapter 3

Maximum Likelihood Estimation of Interval-valued Returns and Trading Intensity

3.1 Introduction

With the seminar paper by Billard and Diday (2003, 2006), a variety of regression models are proposed to fit the interval data, see Arroyo, González-Rivera and Maté (2011) for a review. The simplest approach, proposed by Billard and Diday (2000), is to fit a regression model to the centers of the intervals of the dependent variable and of the regressors. Further approaches consider two separate regressions, one for the lower bound and another for the upper bound of intervals, neither with no constraints in the regression coefficients (Brito 2007). In a similar line, Lima Neto and de Carvalho (2008) propose running two different regressions, one for the center and another for the range of
the intervals, with no constraints. None of these approaches guarantees that the fitted values from the regressions will satisfy the natural order of an interval, i.e., $\hat{y}_l \leq \hat{y}_u$, for all observations in the sample. Later on, Lima Neto and de Carvalho (2010) modify their previous model by imposing non-negative constraints on the regression coefficients of the model for the range and solve a quadratic programming problem to find the least squares solution. Such ad hoc restrictions, requiring that the regressors in range regression must be positive, limit the usefulness of the model. Recently, González-Rivera and Lin (2013) propose a constrained regression model that generalizes the previous regression models to lower/upper bounds and center/radius of intervals, and that impose an observability restriction that guarantee the natural order of the fitted intervals.

All the existing models mainly focus on regressions of minimum and maximum (or center and radius) of intervals on their lag terms with no or restrictive assumption on the distribution of error terms, and such a specification ignores the extreme nature of the lower and upper bounds of intervals. Alternatively, we adopt an alternative modelling approach that exploits the extreme property. In the model, we assume that there is an underlying stochastic process $y_t$ for the interval-valued time series, and that the lower and upper bounds of the intervals ($y_{lt}$ and $y_{ut}$) are the realized extreme observations (minima and maxima) based on the $n_t$ random draws from the conditional densities of underlying strictly stationary stochastic process $y_t$. In addition, the number of random draws follows certain discrete stochastic process. In the model specification, the conditional densities of the two underlying stochastic processes $y_t$ and $n_t$ are updated each period, following certain dynamic rules. In this paper, we consider the case that the underlying stochastic processes are normal and negative binomial. After the dynamic rules of conditional densities are specified, we use maximum likelihood method

115
to estimate the model. As an example, we apply our model to the USDA national daily boxed beef cuts negotiated sales data. We find that our model produces estimates that successfully capture the fluctuations in the low/high prices and number of trades.

The paper is organized as follows. In section 2, we propose the general framework of the model. In Section 3, we introduce a special case of the model, and establish forecast based on the estimation results. Section 4 is the empirical application. Finally, Section 5 concludes.

### 3.2 General Framework

In the data generating process for interval-valued data, we assume that there is an underlying stochastic process for the interval-valued time series, and the high/low values of intervals are the realized highest and lowest order statistics based on the random draws from the conditional densities of the underlying stochastic process. Put it more formally, let \( \{y_t : t = 1, \cdots, T\} \) be a underlying stochastic process. The underlying or latent random variable \( y_t \) at time \( t \) has a conditional density as \( f(y_t | \mathcal{F}_t; \theta_1) \); and, at each time \( t \), from the conditional density of \( y_t \) there are \( n_t \) independent draws which follows certain discrete distribution \( h(n_t | \mathcal{F}_t; \theta_2) \), where \( \mathcal{F}_t \) is the information set available at time \( t \) and \( \theta_1 \) and \( \theta_2 \) are the parameter vectors. We use \( y_{lt} \) and \( y_{ut} \) to denote the largest and smallest value of the random sample \( S_t \equiv \{y_{it} : i = 1, 2, \cdots, n_t\} \):

\[
\begin{align*}
    y_{lt} &\equiv \min_i S_t = \min_{1 \leq i \leq n_t} \{y_{it}\}, \\
    y_{ut} &\equiv \max_i S_t = \max_{1 \leq i \leq n_t} \{y_{it}\}.
\end{align*}
\]
Then, \( \{(y_{lt}, y_{ut}, n_t) : t = 1, \cdots, T\} \) forms the observed interval time series and number of random draws. The intuition is straightforward. For instance, in the tick-by-tick high frequency financial trading data, for every period (within minutes or seconds), we have observations for low return, high return and number of trades. In this setting, we assume that the conditional distribution of underlying return \( f(y_t|\mathcal{F}_t, \theta_1) \) is updated every period according to certain dynamic rule, and that the number of trades in the one minute interval represents the number of random draws \( n_t \) from that conditional distribution of return. Then, at time \( t \), the low and high returns \( (y_{lt}, y_{ut}) \) is the lowest and highest ranked order statistics of the random sample \( S_t \) formed by the \( n_t \) draws.

Since \( (y_{lt}, y_{ut}) \) are the lowest and highest order statistics of random sample \( S_t \), the joint conditional probability density of \( (y_{lt}, y_{ut}) \) given \( n_t \) and information set \( \mathcal{F}_t \) is,

\[
g(y_{lt}, y_{ut}|n_t, \mathcal{F}_t; \theta_1) = n_t(n_t - 1) [F(y_{ut}|\mathcal{F}_t; \theta_1) - F(y_{lt}|\mathcal{F}_t; \theta_1)]^{n_t-2} \\
\times f(y_{lt}|\mathcal{F}_t; \theta_1) f(y_{ut}|\mathcal{F}_t; \theta_1)
\]

where \( F(\cdot|\mathcal{F}_t; \theta_1) \) is the cumulative distribution function corresponding to the conditional density \( f(y_t|\mathcal{F}_t; \theta_1) \). Given that, the joint probability density of \( (y_{lt}, y_{ut}, n_t) \) conditional on information set \( \mathcal{F}_t \) is,

\[
p(y_{lt}, y_{ut}, n_t|\mathcal{F}_t; \theta) = g(y_{lt}, y_{ut}|n_t, \mathcal{F}_t; \theta_1) h(n_t|\mathcal{F}_t; \theta_2)
\]

where \( \theta = \theta_1 \cup \theta_2 \) is the joint set of parameter vectors. Then, we can write the log
likelihood function of \( \{(y_{lt}, y_{ut}, n_t) : t = 1, 2, \cdots, T\} \) is,

\[
L(\theta | y_u, y_l, n) = \sum_{t=1}^{T} \log p(y_{lt}, y_{ut}, n_t | \mathcal{F}_t; \theta)
= \sum_{t=1}^{T} \log g(y_{lt}, y_{ut} | n_t, \mathcal{F}_t; \theta_1) + \sum_{t=1}^{T} \log h(n_t | \mathcal{F}_t; \theta_2)
\equiv L_1(\theta_1 | y_u, y_l, n) + L_2(\theta_2 | n). \tag{3.2.1}
\]

If the intersection of \( \theta_1 \) and \( \theta_2 \) is non-empty, i.e., some parameters appear in both conditional densities \( f(y_t | \mathcal{F}_t, \theta_1) \) and \( h(n_t | \mathcal{F}_t, \theta_2) \) at the same time, the ML estimator \( \hat{\theta}_{ML} \) must be jointly estimated by maximizing (3.2.1):

\[
\hat{\theta}_{ML} = \arg \max_{\theta} L(\theta | y_u, y_l, n).
\]

On the other hand, if the intersection of \( \theta_1 \) and \( \theta_2 \) is empty, i.e., the parameter vector \( \theta \) can be partitioned into \( \theta_1 \) and \( \theta_2 \) that belong to the two conditional densities, the ML estimator \( \hat{\theta}_{ML} \) that maximizes the joint log-likelihood function \( L(\theta | y_u, y_l, n) \) is equivalent to the ML estimators \( \hat{\theta}_{1,ML} \) and \( \hat{\theta}_{2,ML} \) that maximize \( L_1(\theta_1 | y_u, y_l, n) \) and \( L_2(\theta_2 | n) \) in (3.2.1) separately. Therefore, the estimation procedure is simplified as follows,

\[
\hat{\theta}_{1,ML} = \arg \max_{\theta_1} \sum_{t=1}^{T} \log g(y_{lt}, y_{ut} | n_t, \mathcal{F}_t; \theta_1)
\hat{\theta}_{2,ML} = \arg \max_{\theta_2} \sum_{t=1}^{T} \log h(n_t | \mathcal{F}_t; \theta_2).
\]

In this paper, we only entertain this simplified case, and give a specific example in the next section.
3.3 A Special Case

3.3.1 Estimation

To implement the maximum likelihood estimation, we further assume that the conditional densities of underlying stochastic process $y_t$ and number of random draws $n_t$ are normal and negative binomial respectively, i.e.,

$$f(y_t | F_t; \theta_1) = \frac{1}{\sqrt{2\pi \sigma^2_t}} \exp \left\{ -\frac{(y_t - \mu_t)^2}{2\sigma^2_t} \right\},$$

$$h(n_t | F_t; \theta_2) = \frac{\Gamma(n_t + d - 2)}{(n_t - 2)!\Gamma(d)} \left( \frac{d}{\lambda_t + d} \right)^d \left( \frac{\lambda_t}{\lambda_t + d} \right)^{n_t - 2},$$

where $\mu_t$ and $\sigma^2_t$ are the conditional mean and variance of the normal density; $\lambda_t$ and $d$ are the intensity and dispersion parameter of negative binomial density. To close the model, we specify the dynamics of the conditional mean, variance and intensity of the underlying random processes $y_t$ and $n_t$, which are described in (3.3.2), (3.3.3), and (3.3.4) respectively,

$$\mu_t(\alpha) = f_1(F_t; \alpha) \hspace{1cm} (3.3.2)$$

$$\sigma^2_t(\beta) = f_2(F_t; \beta) \hspace{1cm} (3.3.3)$$

$$\lambda_t(\gamma) = f_3(F_t; \gamma) \hspace{1cm} (3.3.4)$$

where $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$ define the dynamics; and $\alpha$, $\beta$, and $\gamma$ are the parameters.

Given all the assumptions above, the joint density of $(y_{lt}, y_{ut}, n_t)$ can be written explicitly.

---

1 We use negative binomial distribution as a robust alternative to the Poisson distribution, since the additional dispersion parameter $d$ in negative binomial distribution captures the deviation from the Poisson, which characterizes the phenomenon of over-dispersion. When $d$ goes to infinity, the negative binomial converges to Poisson. For simplicity here, we only use Poisson distribution.
as,

\[
p(y_t, y_{ut}, n_t | \mathcal{F}_t; \theta) = g(y_t, y_{ut} | n_t, \mathcal{F}_t; \theta_1) h(n_t | \mathcal{F}_t; \theta_2)
\]

\[
= n_t(n_t - 1) \left[ \Phi \left( \frac{y_{ut} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) - \Phi \left( \frac{y_{lt} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \right]^{n_t-2}
\]

\[
\times \frac{1}{\sigma_t(\beta)^d} \phi \left( \frac{y_{lt} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \frac{1}{\sigma_t(\beta)^d} \phi \left( \frac{y_{ut} - \mu_t(\alpha)}{\sigma_t(\beta)} \right)
\]

\[
\times \frac{\Gamma(n_t + d - 2)}{(n_t - 2)! \Gamma(d)} \frac{d}{\lambda_t + d} \left( \frac{\lambda_t}{\lambda_t + d} \right)^{n_t-2}
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are the standard normal probability density and cumulative distribution functions; and that \( \theta_1 = (\alpha, \beta) \) and \( \theta_2 = (\gamma, d) \) are the parameter vectors with empty intersection. Now we consider simple autoregressive specifications of the conditional mean, variance and intensity as follows,

\[
\mu_t = \alpha_0 + \alpha_{1t} y_{lt-1} + \alpha_{ht} y_{ht-1} + \alpha_{n1} \log n_{t-1}, \quad (3.3.5)
\]

\[
\log \sigma_t^2 = \beta_0 + \beta_{r1} \log(y_{ht-1} - y_{lt-1})^2 + \beta_{n1} \log n_{t-1}, \quad (3.3.6)
\]

\[
\log \lambda_t = \gamma_0 + \gamma_{r1} \log(y_{ht-1} - y_{lt-1})^2 + \gamma_{r1} \log n_{t-1}. \quad (3.3.7)
\]

With equations (3.3.5) – (3.3.7), now the model is fully specified, and the log-likelihood function is given in (3.2.1), where

\[
L_1(\theta_1 | y_u, y_l, n) = \sum_{t=1}^{T} \log g(y_t, y_{ut} | n_t, \mathcal{F}_t; \theta_1)
\]

\[
= \sum_{t=1}^{T} \log n_t + \sum_{t=1}^{T} \log(n_t - 1) - 2 \sum_{t=1}^{T} \log \sigma_t(\beta)
\]

\[
+ \sum_{t=1}^{T} (n_t - 2) \log \left[ \Phi \left( \frac{y_{ut} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) - \Phi \left( \frac{y_{lt} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \right]
\]

\[
+ \sum_{t=1}^{T} \log \phi \left( \frac{y_{ut} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) + \sum_{t=1}^{T} \log \phi \left( \frac{y_{lt} - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \quad (3.3.8)
\]
and

\[
L_2(\theta_2|\mathbf{n}) \equiv \sum_{t=1}^{T} \log h(n_t|\mathcal{F}_t; \theta_2)
\]
\[
= \sum_{t=1}^{T} \log \Gamma(n_t + d - 2) - \sum_{t=1}^{T} \log \Gamma(d) - \sum_{t=1}^{T} \log(n_t - 2)!
\]
\[
+ d \sum_{t=1}^{T} \log \left( \frac{d}{\lambda_t + d} \right) + \sum_{t=1}^{T} (n_t - 2) \log \left( \frac{\lambda_t}{\lambda_t + d} \right).
\] (3.3.9)

The ML estimate \( \hat{\theta}_{ML} \) is obtained by maximizing the log likelihood function in (3.2.1).

Alternatively, since the intersection of parameter vector \( \theta_1 \) and \( \theta_2 \) is empty, the ML estimates \( \hat{\theta}_{1,ML} \) and \( \hat{\theta}_{2,ML} \) can be also obtained equivalently by maximizing separately the two parts of log-likelihood functions \( L_1(\theta_1|\mathbf{y}_u, \mathbf{y}_l, \mathbf{n}) \) and \( L_2(\theta_2|\mathbf{n}) \) in (3.3.8) and (3.3.9).

The asymptotic properties of the ML estimator have been extensively studied under general conditions in statistical literature. To prove the consistency and asymptotic normality, we only need to verify that those certain regularity conditions hold, see Hayashi (2000) and Greene (2012) for detail. In this paper, the verification is left for future work. Numerically, we use Monte Carlo simulations to show that the our ML estimator has low mean squared errors and normal-like empirical distributions.

### 3.3.2 Forecasting

Given the model specification, the conditional mean of lower bound of interval is,

\[
E(y_{lt}|\mathcal{F}_t, \theta) = E[E(y_{lt}|\mathcal{F}_t, n_t; \theta_1)|\mathcal{F}_t; \theta_2]
\]
\[
= E \left[ n_t \int_{-\infty}^{+\infty} s \left( 1 - \Phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \right)^{n_t-1} \frac{1}{\sigma_t(\beta)} \phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) ds | \mathcal{F}_t; \theta_2 \right]
\]
\[
= \sum_{n_t=2}^{\infty} n_t \int_{-\infty}^{+\infty} s \left( 1 - \Phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \right)^{n_t-1} \frac{1}{\sigma_t(\beta)} \phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) ds \cdot h(n_t|F_t; \theta_2(3.3.10))
\]

where the law of iterated expectation and the density of minimum order statistic \(y_{lt}\) are used in the first and second equation respectively. Similarly, the conditional mean of upper bound of interval is,

\[
E(y_{ut}|F_t, \theta) = \sum_{n_t=2}^{\infty} n_t \int_{-\infty}^{+\infty} s \left( \Phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) \right)^{n_t-1} \frac{1}{\sigma_t(\beta)} \phi \left( \frac{s - \mu_t(\alpha)}{\sigma_t(\beta)} \right) ds \cdot h(n_t|F_t; \theta_2(3.3.11))
\]

Therefore, after the estimation procedure, we can plug the ML estimates \(\hat{\theta}_{ML}\) into the \((3.3.10)\) and \((3.3.11)\) to obtain the fitted values of lower and upper bounds, denoted as \(\hat{y}_{lt}\) and \(\hat{y}_{ut}\). Using these equations, we can also do out-of-sample forecasting by updating the information set \(F_t\): First, divide the whole sample (sample size \(T\)) into estimation period (sample size \(T - h\)) and evaluation period (sample size \(h\)). Then, use the data in estimation period to estimate the parameters. Finally, use the ML estimates \(\hat{\theta}_{ML}\) and equations \((3.3.10)\) and \((3.3.11)\) with updated information set \(F_{T-h+1}\) to obtain one-step-ahead forecast interval \(\hat{y}_{T-h+1} = \{\hat{y}_{l,T-h+1}, \hat{y}_{u,T-h+1}\}\). Multi-step-ahead forecast can be obtained by using equations \((3.3.10)\) and \((3.3.11)\) recursively.

### 3.4 Empirical Application

In this section, we apply our model to the daily livestock sales data which are documented by the Agriculture Marketing Service (AMS) in the United States Department of Agriculture (USDA). AMS provides current, unbiased price and sales information to assist in the orderly marketing and distribution of farm commodities. It daily market
news reports include information on prices, volume, quality, condition, and other mar-
ket data on farm products in specific markets and marketing areas. Reports cover both
domestic and international markets.

The specific data set used in the paper are the national daily boxed beef cuts negotiated
sales prices. The historical data are archived, and can be downloaded from the following
website\footnote{The original USDA url link is very long, so we shorten the original url by Google url shortener.} \url{http://goo.gl/76WYQ}. The name of the daily beef report is “Boxed Beef
Cutout & Cuts-Negotiated Sales PM CSV”, coded as “LM_XB403”. In the daily report,
sales prices of different parts (choice cuts) of beef are provided, and we use the price of
“Rib, Ribeye, lip-on, bn-in”, which is item “109E” in the report. The variables include
number of trades, total pounds, low price, high price, and weighted average price, where
prices reflect U.S. dollars per 100 pounds. The data ranges form January 3rd 2012 to
June 29th 2012 with 127 observations in total. Figure 3.1a and Figure 3.1b plot the
time series of daily prices and number of trades respectively.

In the graph, the prices have clear upward trends, which suggest that the all the price
series are non-stationary. To convert the original high and low prices to be stationary,
we calculate the high and low returns with respect to the weighted average price of the
previous day, that is,

$$r_{ht} = \frac{P_{\text{high,}t} - P_{\text{avg,}t-1}}{P_{\text{avg,}t-1}} \times 100\%$$

$$r_{lt} = \frac{P_{\text{low,}t} - P_{\text{avg,}t-1}}{P_{\text{avg,}t-1}} \times 100\%.$$

Figure 3.1c plots the low and high returns, for which stationary assumption is plausible.

In addition, the number of daily trades shows periodical ups and downs in the graph,
Figure 3.1: Time Series Plots of Daily Prices/Returns and Number of Trades
and implies the existence of seasonality. In Figure 3.2a and Figure 3.2b we draw the ACF and PACF for the number of trades respectively. Clearly, PACF has a clear cut-off at the tenth-period lag, and ACF decays to zero gradually. The ACF and PACF suggest the existence of autoregressive and (weekly) seasonal dynamics in the intensity of trading times.

Table 3.1 reports descriptive statistics for the three series: low return \( r_{lt} \), high return \( r_{ht} \) and number of trades \( n_t \). The low and high returns range from \(-13.21\%\) to \(1.75\%\) and \(0.6\%\) to \(15.5\%\) respectively. In addition, both series have similar variance and their kurtosis are close to 3, implying that the returns do not have heavy tail and that normal distribution is a plausible assumption for the returns. In addition, for the number of trades, we observe that the variance 159.81 is much larger than the mean 32.5, it suggests that, when modeling the number of trades, we should entertain negative binomial distribution which is capable of capturing over-dispersion.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>low return ( r_{lt} )</th>
<th>high return ( r_{ht} )</th>
<th># of trades ( n_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>(-13.2100)</td>
<td>0.6002</td>
<td>9</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>(-4.2610)</td>
<td>3.6550</td>
<td>23</td>
</tr>
<tr>
<td>Median</td>
<td>(-2.2930)</td>
<td>5.5600</td>
<td>31</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>(-0.9475)</td>
<td>8.0170</td>
<td>40</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.7500</td>
<td>15.5100</td>
<td>73</td>
</tr>
<tr>
<td>Mean</td>
<td>(-2.8820)</td>
<td>5.9640</td>
<td>32.5</td>
</tr>
<tr>
<td>Variance</td>
<td>7.4485</td>
<td>8.1451</td>
<td>159.81</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.3978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>(-1.1280)</td>
<td>0.5632</td>
<td>0.5385</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.2116</td>
<td>3.0494</td>
<td>2.9563</td>
</tr>
</tbody>
</table>

Table 3.1: Descriptive Statistics

In the model specification, we maintain the assumptions that the conditional densities of underlying daily return \( r_t \) and number of trades \( n_t \) follow normal and negative bino-
mial distributions respectively. In addition, the dynamics of the conditional mean and variance of \( r_t \) and intensity of \( n_t \) are as follows,

\[
\mu_t = \alpha_0 + \alpha_{l1} r_{l,t-1} + \alpha_{h1} r_{h,t-1} + \alpha_{n1} \log n_{t-1} \quad (3.4.12)
\]

\[
\log \sigma_t^2 = \beta_0 + \beta_{r1} \log(r_{h,t-1} - r_{l,t-1})^2 + \beta_{n1} \log n_{t-1} \quad (3.4.13)
\]

\[
\log \lambda_t = \gamma_0 + \gamma_{r1} \log(r_{h,t-1} - r_{l,t-1})^2 + \gamma_{n1} \log n_{t-1} + \gamma_{n4} \log n_{t-4} + \gamma_{n5} \log n_{t-5} \quad (3.4.14)
\]

In the dynamic specification of intensity \( \log \lambda_t \), we include 1-period, 4-period, and 5-period lags for logarithm of number of trades to capture the autoregressive dynamics and stochastic (weekly) seasonality shown in the ACF and PACF graphs.

As the parameters can be partitioned into two parts, \( \theta_1 = (\alpha, \beta) \) in the joint conditional density of low and high returns and \( \theta_2 = (\gamma, d) \) in the conditional density of number of trades \( n_t \), we obtain the ML estimates \( \hat{\theta}_{1,\text{ML}} \) and \( \hat{\theta}_{1,\text{ML}} \) by maximizing the two parts of the log-likelihood function separately. The upper panel of Table 3.2 reports the estimation results for the dynamics of conditional mean and variance of the underlying return \( r_t \). The standard errors of the estimates are obtained from the sandwich form (BHHH) variance-covariance matrix of the ML estimator. Given 5% significance level, only \( \hat{\alpha}_{l1} \) and \( \hat{\beta}_{r1} \) are significant for conditional mean and variance respectively. Therefore, it shows that the lagged value of low returns \( r_{l,t-1} \) and range of returns \( \log(r_{h,t-1} - r_{l,t-1}) \) have significant effects on the determination of the conditional densities of underlying returns \( r_t \). In addition, both \( \hat{\alpha}_{n1} \) and \( \hat{\beta}_{n1} \) are insignificant. Thus, the lagged values of number of trades do not affect the conditional mean and variance of \( r_t \).

\^Although the 10-period lag is the fourth significant spike in the PACF of \( n_t \) (Figure 3.2b), its coefficient is not statistical significant when included. Therefore, we do not put it in the (3.4.14).
Figure 3.2: ACF and PACF for Number of Trades
The lower panel of Table 3.2 reports the estimation results for the dynamics of trading intensity. In the dynamics, the estimates $\hat{\gamma}_0$, $\hat{\gamma}_{n4}$, and $\hat{\gamma}_{n5}$ are significant at 5% level, while $\hat{\gamma}_{r1}$ and $\hat{\gamma}_{n1}$ are not. It indicates that the trading intensity $\lambda_t$ is dominated by long term dynamics and/or weekly seasonality in the number of trades (i.e., $\log n_{t-4}$ and $\log n_{t-5}$ terms in (3.4.14)) instead of short term dynamics $\log n_{t-1}$ and lagged range of returns $\log (r_{h,t-1} - r_{l,t-1})$. Notice that the conditional densities of underlying $r_t$ and $n_t$ are affected by different sets of variables. Therefore, it suggests that the two conditional densities may be updated irrelevantly.

With the ML estimates, we calculate estimated conditional mean $\hat{\mu}_t$, conditional variance $\hat{\sigma}^2_t$ and $\hat{\lambda}_t$ based on (3.4.12) – (3.4.14). In addition, we also calculate the estimated low and high returns ($\hat{r}_l$ and $\hat{r}_h$) based on the expression of conditional mean of lower and

<table>
<thead>
<tr>
<th>Conditional Density of $(r_{lt}, r_{ht})$</th>
<th>estimate</th>
<th>s.e.</th>
<th>t statistic</th>
<th>p-value$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>-1.3392</td>
<td>0.9085</td>
<td>-1.4741</td>
<td>0.1404</td>
</tr>
<tr>
<td>$\alpha_{l1}$</td>
<td>-0.3224</td>
<td>0.0818</td>
<td>-3.9419</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\alpha_{h1}$</td>
<td>0.0972</td>
<td>0.0850</td>
<td>1.1437</td>
<td>0.2528</td>
</tr>
<tr>
<td>$\alpha_{n1}$</td>
<td>0.4160</td>
<td>0.3105</td>
<td>1.3397</td>
<td>0.1804</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.2359</td>
<td>0.8144</td>
<td>0.2896</td>
<td>0.7720</td>
</tr>
<tr>
<td>$\beta_{r1}$</td>
<td>0.3445</td>
<td>0.0443</td>
<td>7.7815</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\beta_{n1}$</td>
<td>-0.0336</td>
<td>0.0558</td>
<td>-0.6029</td>
<td>0.5466</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional Density of $(n_t)$</th>
<th>estimate</th>
<th>s.e.</th>
<th>t statistic</th>
<th>p-value$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>3.5375</td>
<td>0.5640</td>
<td>6.2722</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\gamma_{r0}$</td>
<td>0.0014</td>
<td>0.0242</td>
<td>0.0570</td>
<td>0.9546</td>
</tr>
<tr>
<td>$\gamma_{n1}$</td>
<td>-0.1414</td>
<td>0.1469</td>
<td>-0.9627</td>
<td>0.3356</td>
</tr>
<tr>
<td>$\gamma_{n4}$</td>
<td>-0.1627</td>
<td>0.0795</td>
<td>-2.0449</td>
<td>0.0408</td>
</tr>
<tr>
<td>$\gamma_{n5}$</td>
<td>0.2641</td>
<td>0.0448</td>
<td>5.9005</td>
<td>0.0000</td>
</tr>
<tr>
<td>$d$</td>
<td>9.6702</td>
<td>1.8469</td>
<td>5.2360</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$^1 p$-values are based on standard normal.
upper bounds of interval: (3.3.10) and (3.3.11). Figure 3.3a, 3.3b and 3.3c plot those estimated series respectively.

In Figure 3.3a, estimated low, high and mean returns are plotted with actual low and high returns. The estimated mean returns \( \hat{\mu}_t \) lie between the actual low and high returns. Estimated low and high returns also generally capture the ups and downs in the actual returns. Figure 3.3b plots the estimated variance. As the true variance is latent and therefore unobservable, we can only plot the estimated values, which fluctuate over time and hence indicate conditional heteroskedasticity. In Figure 3.3c, we plot the actual number of trades \( n_t \) and estimated intensity \( \hat{\lambda}_t \). Although these two series are not directly comparable, however, the estimated conditional intensity still generally pin down the ups and downs in the number of trades since the estimated conditional intensity should be very close to the conditional expectation of the number of trades.

### 3.5 Conclusion

All the existing models mainly focus on the regressions of minimum and maximum (or center and radius) of intervals on its lag terms with no or restrictive assumption on the distribution of error terms, and such specification ignores the extreme nature of the lower and upper bounds of intervals. Alternatively, we adopt a modelling approach that exploits such extreme property. In the model, we assume that there is an underlying stochastic process \( y_t \) for the interval-valued time series, and that the lower and upper bounds of the intervals \( (y_{lt} \text{ and } y_{ut}) \) are the realized extreme observations (minima and maxima) based on the \( n_t \) random draws from the conditional densities of underlying strictly stationary stochastic process \( y_t \). In addition, the number of ran-
(a) Estimated Daily Returns: $\mu_t, r_{lt}, r_{ht}$

(b) Estimated Variance: $\sigma_t^2$

(c) Estimated Intensity: $\lambda_t$

Figure 3.3: Time Series Plots of Daily Prices/Returns and Number of Trades
dom draws follows certain discrete stochastic process. In the model specification, the conditional densities of the two underlying stochastic processes $y_t$ and $n_t$ are updated each period, following certain dynamic rules. In this paper, we consider the case that the underlying stochastic processes are normal and negative binomial. After the dynamic rules of conditional densities are specified, we use maximum likelihood method to estimate the model. As an example, we apply our model to the USDA national daily boxed beef cuts negotiated sales data. We find that our model produces estimates that successfully capture the fluctuations in the low/high prices and number of trades. Specifically, the lagged value of low returns and range of returns have significant effects on the determination of the conditional densities of underlying returns $r_t$, while both $\hat{\alpha}_n$ and $\hat{\beta}_n$ are insignificant, i.e., the lagged values of number of trades do not affect the conditional mean and variance of underlying returns $r_t$. In addition, the trading intensity $\lambda_t$ is dominated by long term dynamics and/or weekly seasonality in the number of trades instead of the short term dynamics or lagged range of returns. The conditional densities of underlying returns $r_t$ and number of trades $n_t$ are affected by different sets of variables. Therefore, it suggests that the two conditional densities may be updated irrelevantly for this specific daily boxed beef cuts trading data set.
Bibliography


Chapter 4

Semiparametric Estimation of Interval-valued Time Series Using Extreme Value Approach

4.1 Introduction

With the advancement of sophisticated information system that reduces the cost of generating, retrieving and storing large data, more and more massive data sets have been available to both researchers and practitioners in many disciplines. Economics and business are also in such trend. For instance, financial data is now available at very high frequencies for almost every asset that is traded in a public market providing data sets with millions of observations. Marketing data sets offer high granularity about consumers and products characteristics. Environmental stations produce data sets that contain high and low frequency records of temperatures, atmospheric conditions, pol-
lutants, etc. across many regions. Statistical institutes, like the Census Bureau, collect socioeconomic information about all individuals in a nation. These data sets containing massive information tend to be released in an aggregated format, either because of confidentiality or because the interest of study is not the individual unit but a collective of units. In these circumstances, the researcher does not face classical cross-sectional or time series data sets where each observations are single values in the real line, but data aggregated in some fashion. For example, the observation of interval-valued data $[y_l, y_u]$ offers information on the lower and upper bound of the variable of interest; and information about income or networth very often comes in interval format. So do low and high prices of an asset in a given day, daily temperature intervals, and low/high prices of certain commodities sold in different places, etc.

The interval-valued data are also as a type of symbolic data. With the symbolic approach pioneered by Billard and Diday (2003, 2006), there are a variety of proposals to fit a regression model to interval data. Arroyo, González-Rivera and Maté (2011) provides a review for them. The simplest approach proposed by Billard and Diday (2000) is to fit a regression model to the centers of the intervals of the dependent variable and of the regressors. Further approaches consider two separate regressions, one for the lower bound and another for the upper bound of intervals, neither with any constraints in the regression coefficients (Brito 2007). In a similar line, Lima Neto and de Carvalho (2008) propose running two different regressions, one for the center and another for the range of the intervals, with no constraints. None of these approaches guarantees that the fitted values from the regressions will satisfy the natural order of an interval, i.e., $\hat{y}_l \leq \hat{y}_u$, for all observations in the sample. Later on, Lima Neto and de Carvalho (2010) modify their previous model by imposing non-negative constraints on the regression coefficients of the
model for the range and solve a quadratic programming problem to find the least squares solution. Such ad hoc restrictions, requiring that the regressors in range regression must be positive, limit the usefulness of the model. Recently, González-Rivera and Lin (2012) propose a constrained regression model that generalizes the previous regression models to lower/upper bounds and center/radius of intervals, and that impose a observability restriction that guarantee the natural order of the fitted intervals. All the existing models mainly focus on the regressions of minimum and maximum (or center and radius) of intervals on its lag terms with no or restrictive assumption on the distribution of error terms, and such specification ignores the extreme nature of the lower and upper bounds of intervals.

In this paper, we propose a semiparametric model that exploits the extreme property of lower and upper bounds of intervals by assuming that there is an underlying stochastic process for the interval-valued time series, and that the lower and upper bounds of the intervals are the realized extreme observations (minima and maxima) based on the random draws from the conditional densities of underlying stochastic process which is strictly stationary. The specification of this data generating process decomposes the analysis of interval-valued time series into two parts. The first part is the classical point-valued time series analysis for the underlying stochastic process, for which the time series of centers is used as a proxy. The second part is to model the conditional mean of the lower/upper bounds nonparametrically, since the conditional mean of extreme value is often highly nonlinear and intractable even for commonly used distributions such as normal and students’ t, etc. Following this methodology, we combines the expertise of the two fields to analyze the interval-valued time series data, and propose a two step procedure for the estimation of interval-valued time series: 1) construct and estimate
a parametric model for the time series of centers, and obtain the estimated conditional moments of it; 2) and then perform nonparametric regression of lower/upper bounds with the estimated conditional moments as the regressors. Based on the literature on nonparametric regression with generated regressors, the effect of parameter uncertainty in the second step is asymptotically negligible given some regularity conditions, and therefore, our two-step estimator has typical nonparametric convergence rate. As an illustration of the methods that we propose, we apply our two step estimator to the data of five-minute low/high stock returns of Wells Fargo Corporation. In the first step, we obtain conditional mean and standard deviation from the AR(1)-GARCH(1,1) model for the center values of the low/high returns, and in the second step, we perform nonparametric regression (local constant and local linear smoothing) by regressing low/high returns on the number of trades and estimated conditional moments generated in the first step. Then we compare our two-step estimator with its several competing approaches, including (constrained) center-radius model, VAR model and location-scale model with confidence intervals. The performance of each estimation methodology is evaluated according to four criteria: root mean squared errors, coverage and efficiency rates, multivariate loss functions, and mean distance error. The comparison shows that our proposed estimators are superior to the competing approaches, and that both estimated moments and number of trades are indispensable in capturing the spikes and dynamics of low/high returns correctly.

The paper is organized as follows. In section 2, we provide the general framework the basic assumptions. In section 3, we present the two-step estimation procedure and establish the asymptotic properties of the nonparametric regression in the second step using generated regressors. In section 4, we model the five-minute interval of low/high
stock returns of Wells Fargo, and provide an extensive comparison of the proposed estimation methods with the competing ones. Finally, section 5 concludes.

4.2 Basic Assumptions and Methodology

In this section, we describe the data generating process of the interval-valued time series, and then we build our model heuristically. In the procedure of model building, we impose several assumptions which are moderate enough to admit a lot of potential processes that are frequently used in financial data modeling.

**Assumption 6 (Data Generating Process).** Let \( \{y_t : t = 1, \ldots, T\} \) be a underlying stationary or nonstationary time series. The random variable \( y_t \) at time \( t \) has a conditional density as \( f_t(y_t|\mathcal{F}_{t-1}; \theta_t) \), where \( \mathcal{F}_{t-1} \) is the information set available at time \( t \) and \( \theta_t \) is the parameter vector characterizing the shape, location and scale of the distribution. Also, at each time \( t \), from the conditional density of \( y_t \) there are \( n_t \) independent draws \( 1 \), i.e., \( S_t \equiv \{y_{it} : i = 1, \ldots, n_t\} \) is an asymptotic random sample from the conditional density of \( y_t \) in each time period.

The assumption of independent draws in each time period can be relaxed to asymptotic independence in the sense of satisfying Condition D\((u_{nt})\) and Condition D'\((u_{nt})\).

**Condition 2** (D\((u_{nt})\)). For any integers \( p, q \) and \( n_t \)

\[
1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_q \leq n_t
\]

such that \( j_1 - i_p \geq l \), random sequence \( \{Y_u\} \) and the threshold sequence \( \{u_{nt}\} \) we have

\[
\left| P\left( \max_{i \in A_1 \cup A_2} Y_{it} \leq u_{nt} \right) - P\left( \max_{i \in A_1} Y_{it} \leq u_{nt} \right) P\left( \max_{i \in A_2} Y_{it} \leq u_{nt} \right) \right| \leq \alpha_{n_t, l_t},
\]

where \( A_1 = \{i_1, \cdots, i_p\}, A_2 = \{j_1, \cdots, j_q\} \) and \( \alpha_{n_t, l_t} \to 0 \) as \( n_t \to \infty \) for some sequence \( l_t = l_{nt} = o(n_t) \).

**Condition 3** (D'\((u_{nt})\)). For random sequence \( \{Y_u\} \) and the threshold sequence \( \{u_{nt}\} \), we have

\[
\lim_{k \to \infty} \limsup_{n \to \infty} \sum_{j=2}^{[n_t/k]} P(Y_{it} > u_{nt}, Y_{jt} > u_{nt}) = 0.
\]

Then the stationary sequence \( \{Y_u\} \) have very much the same extremal behavior as iid sequence, and therefore the dependence in the original sequence can be ignored. The two conditions are discussed in detail in Leadbetter et al. (1983). The verification of the two conditions are tedious in general. For Gaussian \( \{Y_u\} \) the it is very nice, and reduces to showing \( \text{cov}(Y_{i1}, Y_{n1}) \to o(1/\ln n) \), known as Berman's condition.
density \( f_t(y_t|\mathcal{F}_{t-1}; \theta_t) \) with sample size \( n_t \). We use \( y_{lt} \) and \( y_{ut} \) to denote the largest and smallest value of the random sample \( S_t \):

\[
y_{lt} \equiv \min_i S_t = \min_{1 \leq i \leq n_t} \{ y_{it} \},
\]

\[
y_{ut} \equiv \max_i S_t = \max_{1 \leq i \leq n_t} \{ y_{it} \}.
\]

Then, \( \{(y_{lt}, y_{ut}) : t = 1, \cdots, T\} \) forms the observed interval time series of lower and upper bounds.

The intuition behind Assumption 6 is straightforward. For instance, in the tick-by-tick high frequency financial trading data, for every five minutes (or even higher frequencies), we have observations for low return, high return and number of trades. In this setting, we assume that the conditional distribution of return \( f_t(y_t|\mathcal{F}_{t-1}; \theta_t) \) is updated every five minutes according to certain stochastic mechanism, and that the number of trades in the five minutes interval represents the number of random draws \( n_t \) from that conditional distribution of return. Then, the low and high returns \( (y_{lt} \text{ and } y_{ut}) \) is the lowest and highest ranked order statistics of the sample \( S_t \) formed by the \( n_t \) random draws.

The specification of this data generating process decomposes the analysis of interval-valued time series into two parts: (i) classical point-valued time series analysis for the underlying time series \( \{y_t\} \), and (ii) extreme value theory for the extremal observations \( \{(y_{lt}, y_{ut})\} \). Both fields have been well developed in econometrics and statistics. Suppose that the underlying conditional distribution \( f(y_t|\mathcal{F}_{t-1}; \theta) \) of the time series of interest is known, we can use classical extreme value theory to model the upper bound \( y_{ut} \) (maximum value) and lower bound \( y_{lt} \) (minimum value) associated with each underlying conditional density. The asymptotic theories for maxima (and minima) is very
different from those for sums, where we have central limit theorems with normal limiting distribution in various settings by properly normalizing and centering the sum with mean and variance. In contrast, in the limit theorems for maxima, the normalizing and centering terms depend on the tail of underlying distribution and can be difficult to derive. In addition, the shape of limiting distribution of maxima (also known as extreme value distribution) is not unique and depends on the underlying distribution. The central result in the classical extreme value theory is the Fisher-Tippett theorem which specifies the form of the limit distribution for centered and normalized maxima.

**Theorem 4.2.1** (Fisher-Tippett Theorem, limit laws for maxima). Let \( \{Y_i : i = 1, \cdots, n\} \) be a sequence of iid random variables, and \( Y_u \) be the maximal value among the sequence. If there exist norming constants \( c_n > 0, d_n \in \mathbb{R} \) and some non-degenerate distribution function \( H \) such that

\[
c_n^{-1}(Y_u - d_n) \xrightarrow{d} H
\]

then \( H \) belongs to the type of one of the following three standard extreme value distribution functions (\( \alpha > 0 \)):

- **Fréchet:** \( \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \)
- **Weibull:** \( \Psi_\alpha(x) = \begin{cases} \exp\{-(x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \)
- **Gumbel:** \( \Lambda(x) = \exp\{-e^{x}\}, \ x \in \mathbb{R}. \)
The three standard cases can be nested into a one-parameter generalized extreme value distribution (GEV) $H_\xi$, which is defined as,

$$H_\xi(x) = \begin{cases} 
\exp\{- (1 + \xi x)^{-1/\xi}\} & \text{if } \xi \neq 0, \\
\exp\{- \exp\{-x\}\} & \text{if } \xi = 0,
\end{cases}$$

where $1 + \xi x > 0$ and $\xi$ is a shape parameter. Then, (i) $\xi = \alpha^{-1} > 0$ corresponds to the Fréchet distribution $\Phi_\alpha$, and includes the Pareto, inverse gamma, Student $t$, loggamma and the $F$ distributions; (ii) $\xi = 0$ corresponds to the Gumbel distribution $\Lambda$, and includes the exponential, normal, lognormal, hyperbolic and generalized hyperbolic distributions; and (iii) $\xi = \alpha^{-1} < 0$ corresponds to the Weibull distribution $\Psi_\alpha$, and includes the uniform, Beta distributions and other distributions with power law behavior at endpoint. Theorem 4.2.1 states that non-degenerate limit distribution, if exists, must be a generalized extreme distribution with certain shape parameter $\xi$. However, it is possible that non-degenerate limit distribution does not exist for certain underlying distribution functions. We rule out such possibility by impose Assumption 2 on the underlying conditional density of $y_t$.

**Assumption 7.** Let $F_t(y_t|F_{t-1}; \theta)$ be the underlying distribution function of $y_t$ with right endpoint $y_{RF_t} \leq +\infty$ and left endpoint $y_{LF_t} \geq -\infty$, and $F_t(y_t|F_{t-1}; \theta)$, for all $t = 1, \cdots, T$, satisfies

$$\lim_{y_t \to y_{RF_t}} \frac{F_t(y_t|F_{t-1}; \theta)}{F_t(y_{t-1}; \theta)} = 1$$

(4.2.1)

$$\lim_{y_t \to y_{LF_t}} \frac{F_t(y_t|F_{t-1}; \theta)}{F_t(y_t+|F_{t-1}; \theta)} = 1$$

(4.2.2)

These distributions have heavier tails, and for $m \geq 1/\xi$, their $m$-th moment does not exist.
and

\[ F_t(y_{RFt} - \mathcal{F}_{t-1}; \theta) = 1 \]
\[ F_t(y_{LFt} + \mathcal{F}_{t-1}; \theta) = 1 \]

where \( \bar{F}(\cdot) = 1 - F(\cdot) \).

By imposing restriction on the tail of distribution function \( F \), this assumption ensures that the non-degenerate limit distribution for maximum \( Y_u \) and minimum \( Y_l \) exists, which is given in Leadbetter, Lindgren and Rootzén (1983). Intuitively, if the jump heights of the distribution function \( F \) decay sufficiently slow on the tail, then a non-degenerate limit distribution for extremal value does not exist. If the right endpoint is finite, all the distribution function without a jump at its right endpoint satisfies this condition. If the right endpoint is infinite, all the continuous distribution functions with only finite discontinuous points also satisfy this condition. Therefore, although delicate, this assumption is mild and admits a large amount of distribution functions.

We say that the random variable \( Y \) belongs to the maximum domain of attraction of the extreme value distribution \( H_\xi \) if the standardized maxima \( Y_u \) converges in probability to the extreme value distribution \( H_\xi \); and given different conditions on the tail behavior of distribution function \( F \) by restricting the decaying rate of the tail, the random variable \( Y \) falls into different maximum domains of attraction of extreme value distribution. The construction of normalizing and centering term \( c_n \) and \( d_n \) in Theorem 4.2.1 depends on two things: (i) the maximum domain of attraction the random variable \( Y \) and (ii) the functional form of distribution function \( F(y|\mathcal{F}_{t-1}, \theta_t) \). In this paper, we need the

3Examples include poisson, geometric and negative binomial distributions. Note that for integer-valued random variable and the right endpoint is infinite, then 4.2.1 translates into \( \lim_{n \to \infty} \bar{F}(n)/\bar{F}(n-1) = 1 \). 4.2.2 has similar interpretation.
following assumption that the underlying conditional distributions $F_t(y_t|F_{t-1}, \theta_t)$ is time invariant.

**Assumption 8.** All the underlying distribution functions $\{F_t(y_t|F_{t-1}, \theta_t) : t = 1, \ldots, T\}$ are time invariant, i.e.,

$$F_t(y_t|F_{t-1}, \theta_t) = F(y_t|F_{t-1}, \theta_t), \forall t = 1, 2, \ldots, T,$$

and the shape parameters $\xi_l$ and $\xi_u$ of extreme value distribution function associated with $Y_t$ for maximum and minimum is between $-1/2$ and $1/2$, i.e., $-1/2 < \xi_l < 1/2$ and $-1/2 < \xi_u < 1/2$.

This assumption stabilizes the extremal behavior of the underlying distribution functions of series $\{y_t\}$. Therefore, the construction of normalizing and centering term $c_{n_t}$ and $d_{n_t}$ is time invariant, based on which, we can build models to investigate the extremal behavior.

**Lemma 1.** Given Assumption 6 – 8 for all $t = 1, \ldots, T$, there exists normalizing and centering constants $c_{n_t}$ and $d_{n_t}$, such that the standardized maxima and minima i.e., $c_{un_t}^{-1}(Y_{ut} - d_{un_t})$ and $c_{ln_t}^{-1}(Y_{lt} - d_{ln_t})$ converge in distribution to the same extreme value distribution with shape parameters $\xi_u$ and $\xi_l$ respectively, i.e.,

$$c_{un_t}^{-1}(Y_{ut} - d_{un_t}) \xrightarrow{d} H_{\xi_u}, \forall t = 1, \ldots, T$$

$$c_{ln_t}^{-1}(Y_{lt} - d_{ln_t}) \xrightarrow{d} H_{\xi_l}, \forall t = 1, \ldots, T,$$

and the normalizing and centering terms are time invariant functions of $n_t$ and $\theta_t$, i.e.,

$$c_{un_t} = c_u(n_t, \theta_t) \text{ and } c_{ln_t} = c_l(n_t, \theta_t)$$
\[ d_{un_t} = d_u(n_t, \theta_t) \text{ and } d_{ln_t} = d_l(n_t, \theta_t) \]

Since the model we intend to build is based on expectation, the convergence in distribution is too weak to use directly. Therefore, in order to take the advantage of this limit law, we need to impose a restriction on the moment of \( Y_u \) and \( Y_l \) to obtain convergence in stronger mode.

**Assumption 9.** For all \( t = 1, \cdots, T \), there exists \( \delta > 0 \), such that

\[
\sup_{n_t} E \left( \left| c_{un_t}^{-1} (Y_{ut} - d_{un_t}) \right|^{2+\delta} \right) = M_u < \infty,
\]

\[
\sup_{n_t} E \left( \left| c_{ln_t}^{-1} (Y_{lt} - d_{ln_t}) \right|^{2+\delta} \right) = M_l < \infty.
\]

With this additional assumption, we have the following theorem,

**Theorem 4.2.2.** Let \( Y_{\xi_u} \) and \( Y_{\xi_l} \) be the random variable with distribution functions \( H_{\xi_u} \) and \( H_{\xi_l} \) respectively. If Assumption 6 – 9 hold, then for all \( t = 1, \cdots, T \), and for any \( r \in (0, 2] \),

\[
\lim_{n \to \infty} E \left( \left| c_{un_t}^{-1} (Y_{ut} - d_{un_t}) \right|^r \right) = E(|Y_{\xi_u}|^r).
\]

\[
\lim_{n \to \infty} E \left( \left| c_{ln_t}^{-1} (Y_{lt} - d_{ln_t}) \right|^r \right) = E(|Y_{\xi_l}|^r).
\]

Specifically, for expectation, we have

\[
E \left( c_{un_t}^{-1} (Y_{ut} - d_{un_t}) \right) = E(Y_{\xi_u}) + o(1) = \frac{\Gamma(1 - \xi_u) - 1}{\xi_u} + o(1), \quad (4.2.3)
\]
where $\Gamma(\cdot)$ is the Gamma function. Based on (4.2.3), we have,

$$E(Y_{ut}) = d_u(n_t, \theta_t) + c_u(n_t, \theta_t) \frac{\Gamma(1 - \xi_u) - 1}{\xi_u} + o(c_u(n_t, \theta_t)), \quad (4.2.4)$$

and similarly, the expectation of minima $Y_{lt}$ is

$$E(Y_{lt}) = d_l(n_t, \theta_t) + c_l(n_t, \theta_t) \frac{\Gamma(1 - \xi_l) - 1}{\xi_l} + o(c_l(n_t, \theta_t)). \quad (4.2.5)$$

Normalizing and centering functions can be highly nonlinear, and their analytical expressions, if exist, are impractical to be used, even when the underlying distribution $F(y_t|\mathcal{F}_{t-1}; \theta_t)$ is commonly used normal and Stduent’s $t$-distribution. Therefore, in this paper we simply use unknown function $m_u(n_t, \theta_t)$ to denote the conditional mean of maxima $Y_{ut}$, and rely on nonparametrics in the estimation,

$$E(Y_{ut}) = m_u(n_t, \theta_t) \quad (4.2.6)$$

and similarly for minima $Y_{lt}$,

$$E(Y_{lt}) = m_l(n_t, \theta_t) \quad (4.2.7)$$

---

*4 If $y_t$ is normally distributed as $N(\mu_t, \sigma_t^2)$, we have

$$c_u(n_t, \mu_t, \sigma_t) = \frac{1}{\sigma_t \sqrt{2 \ln n_t}},$$

$$d_u(n_t, \mu_t, \sigma_t) = \mu_t + \sigma_t \sqrt{2 \ln n_t} - \sigma_t \frac{\ln(4\pi) + \ln \ln n_t}{2(2 \ln n_t)^{1/2}}.$$

If $y_t$ has $t$-distribution with mean $\mu_t$ and degrees of freedom $\nu_t$, we have $d_u(n_t, \mu_t, \sigma_t) = 0$ and $c_u(n_t, \mu_t, \nu_t)$ solves the following reduced form model,

$$\frac{1}{n} = \frac{1}{2} - (c - \mu_t) \Gamma \left( \frac{\nu_t + 1}{2} \right) \cdot \frac{2F_1 \left( \frac{1}{2}, \frac{\nu_t + 1}{2}; \frac{3}{2}; -\frac{(c - \mu_t)^2}{\nu_t} \right)}{\sqrt{\pi \nu_t} \Gamma \left( \frac{3}{2} \right)},$$

where $2F_1$ is the hypergeometric function.
where $\varepsilon_t$ is the error term with finite second moment which is guaranteed by Assumption 9. Such nonparametric specification avoids the misspecification of underlying distribution and the difficulty of calculating the associated normalizing and centering functions $c(n_t, \theta_t)$ and $d(n_t, \theta_t)$.

In (4.2.6), if the parameter $\theta_t$, charactering the location, scale and shape of the distribution of underlying process $\{y_t\}$, were known, we can directly perform nonparametric methods to estimate the model. However, usually the underlying process $\{y_t\}$ parameter $\theta_t$ is unobservable, therefore, we use the middle points of the interval $\{y_{ct}\}$, defined as,

$$y_{ct} = \frac{y_{lt} + y_{ut}}{2}$$

as the proxy for the underlying process $\{y_t\}$, and specify the dynamics for $\theta_t$ and estimate them as a compromise so that the infeasible model (4.2.6) becomes feasible. To keep the model parsimonious and less onerous in computation, we only include first two moments of the underlying distribution in parameter $\theta_t$, i.e.,

$$\theta_t = (\mu_t, \sigma_t),$$

which generally captures the location and scale of the underlying distribution. In addition, we further modify the model (4.2.6) and (4.2.7) by standardizing the dependent variable $Y_{lt}$ and $Y_{ut}$ with $\mu_t$ and $\sigma_t$.

$$E\left(\frac{Y_{lt} - \mu_t}{\sigma_t}\right) = m_l(\mu_t, \sigma_t, n_t)$$

$$E\left(\frac{Y_{ut} - \mu_t}{\sigma_t}\right) = m_u(\mu_t, \sigma_t, n_t)$$
so that we have,

\[ Y_{lt} = \mu_t + m_l(\mu_t, \sigma_t, n_t)\sigma_t + \varepsilon_{lt}, \]  
(4.2.8)  

\[ Y_{ul} = \mu_t + m_u(\mu_t, \sigma_t, n_t)\sigma_t + \varepsilon_{ul}. \]  
(4.2.9)  

The standardization of the dependent variable does not change the maximum domain of attraction of extreme value distribution which the random variable \( Y_{ul} \) belongs to. In addition, the model after the transformation establishes intuitive relationship between extremal values (upper or lower bounds) and the underlying process \( \{y_t\} \): the upper and lower bounds can be deemed as shifting up and down from mean \( \mu_t \) with \( m_u(\mu_t, \sigma_t, n_t) \) and \( m_l(\mu_t, \sigma_t, n_t) \) times of standard deviation \( \sigma_t \) respectively. So now we can impose various dynamic structures to \( (\mu_t, \sigma_t) \): from simple models, such as ARMA and fractional ARIMA models, to complicated ones, such as ARCH, GARCH, regime-switching and nonlinear transition models.

In this paper, we consider the AR\((s)\)-GARCH\((p,q)\) specification for the underlying process to complete the whole model described as follows,  

**The First Part:**

\[ y_{ct} = \mu_{t|t-1} + u_t \]  
(4.2.10)  

\[ = \mu_{t|t-1} + \sigma_{t|t-1}z_t, \quad z_t \sim W.N.(0, 1) \]  

\[ \mu_{t|t-1} = \phi_0 + \hat{\phi}_1y_{c,t-1} + \cdots + \hat{\phi}_sy_{c,t-s} \]  

\[ \sigma_{t|t-1}^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \]  

\[ \eta \equiv (\phi_0, \cdots, \hat{\phi}_s, \omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q), \]  

147
The Second Part:

\[
y_t = \mu_{t|t-1}(\eta) + m_t(\mu_{t|t-1}(\eta), \sigma_{t|t-1}(\eta), n_t)\sigma_{t|t-1}(\eta) + \varepsilon_t \\
y_{ut} = \mu_{t|t-1}(\eta) + m_u(\mu_{t|t-1}(\eta), \sigma_{t|t-1}(\eta), n_t)\sigma_{t|t-1}(\eta) + \varepsilon_{ut}.
\]

(4.2.11) (4.2.12)

The AR-GARCH specification is commonly used to characterize the stylized features of financial data such heavy tails and volatility clustering. Note that after the specification is transformed from structural form in (4.2.4) and (4.2.5) to reduced form (4.2.10), (4.2.11) and (4.2.12), the model becomes a semiparametric time series model. To make the econometric method concrete, we impose the following conditions to restrict the time dependence of the process.

**Assumption 10** (Weakly Dependence Conditions).

(i) (\(\beta\)-mixing condition). The triple \((y_{lt}, y_{ut}, n_t)\) is a \(\beta\)-mixing process.

(ii) (M.D.S.). The error terms \(\varepsilon_{lt}\) and \(\varepsilon_{ut}\) are martingale difference sequences.

Although it is plausible to combine the two steps and estimate the model, however, such combination complicates the estimation procedure, causing the model less applicable. Therefore, we propose a two step estimation procedure: (i) estimate the underlying AR-GARCH model in the first step, then (ii) plug the fitted values obtained in the first step into the nonparametric models in the second step. Such two step procedure for this model has several advantages: first, we can choose appropriate models in the first step according to the feature of the data, such as homo/hetero-scelasticity, ARCH, and GARCH effects, etc.; second, the estimation strategies in both steps can be conveniently implemented in many, if not all, mainstream statistical softwares.
4.3 Estimation Procedure

Before proceeding to the estimation and asymptotic properties of the two step estimator, we first clarify some abbreviations in notations for convenience of illustration. Let \( \hat{\eta} \) denote the estimates of the true parameter \( \eta_0 \) in AR\((s)\)-GARCH\((p,q)\) model (4.2.10) in the first step, \( m_{lt}(\eta_0) \) and \( m_{ut}(\eta_0) \) denote the true and unknown conditional mean functions \( m_t(\mu_{t|t-1}(\eta_0), \sigma_{t|t-1}(\eta_0), n_t) \) and \( m_u(\mu_{t|t-1}(\eta_0), \sigma_{t|t-1}(\eta_0), n_t) \) in the nonparametric estimations (4.2.11) and (4.2.12) of the second step. In addition, we use the abbreviations \( \hat{\mu}_t \) and \( \hat{\sigma}_t \) to denote \( \mu_{t|t-1}(\hat{\eta}) \) and \( \sigma_{t|t-1}(\hat{\eta}) \). Thus, the infeasible regressions (4.2.11) and (4.2.12) in the second step are written as

\[
y_{jt} = \mu_t + m_{jt}(\eta_t) \sigma_t + \varepsilon_{jt}, \tag{4.3.13}
\]

where we use \( j = u, l \) as the abbreviations for upper and lower bounds.

In our two step estimation, we first estimate the AR\((s)\)-GARCH\((p,q)\) model (4.2.10) and obtain the estimates \( \hat{\eta} \), the estimated conditional mean \( \hat{\mu}_t \) and conditional variance \( \hat{\sigma}_t^2 \), then we plug those fitted values into the nonparametric models in (4.3.13) and implement the feasible nonparametric estimations,

\[
y_{jt} = \hat{\mu}_t + m_{jt}(\hat{\eta}) \hat{\sigma}_t + e_{jt}, \tag{4.3.14}
\]

where the new error terms \( e_{jt} \) become

\[
e_{jt} = \mu_t - \hat{\mu}_t + m_{jt}(\eta) \sigma_t - m_{jt}(\hat{\eta}) \hat{\sigma}_t + \varepsilon_{jt}, \tag{4.3.15}
\]

where the extra terms are introduced because of parameter uncertainty in \( \hat{\eta} \). Note that
in the feasible nonparametric regressions the new error terms may be correlated with regressors and exhibit serial correlation, violating the required assumption for consistency of the nonparametric estimator. However, since \( \hat{\eta} \) is strong consistent with parametric \( \sqrt{T} \) convergence rate, \( e_{jt} \) are also strong consistent estimators for \( \varepsilon_{jt} \). Therefore, when sample size \( T \) is large (which is true for high frequency financial data), the serial correlation and endogeneity of the error term \( e_{jt} \) in feasible nonparametric regression vanish. Therefore, in this paper we regard the error term \( e_{jt} \) as they were purged of all correlation and avoid the modelling complications incurred by such correlation in error terms. In addition, we show that, under some regularity conditions, our two step estimator has a sort of “oracle” property: the feasible estimator is asymptotically equivalent to the infeasible estimator, since the nonparametric estimator in the second step has a convergence rate that is slower than the parametric rate for \( \hat{\eta} \) in the first step.

Note that the regressors \( \hat{\mu}_t \) and \( \hat{\sigma}_t \) in the second step are generated via the estimate \( \hat{\eta} \), therefore the estimate \( \hat{\eta} \) in the first step has two effects on the nonparametric fitted values in the second step: the nonparametric projection (i.e., estimated conditional mean function) is obtained by regressing the dependent variable on the regressors generated by the estimates \( \hat{\eta} \) instead of true \( \eta \), which is called projection effect; and we plug the regressors generated by the estimates \( \hat{\eta} \) (instead of true \( \eta \)) into the estimated conditional mean function to obtain fitted values, which is called sampling effect. We use \( \hat{m}_{jt}(\hat{\eta}; \hat{\eta}) \) to denote the nonparametric estimates based on generated regressors, and \( m_{jt}(\eta; \eta) \equiv E[\hat{m}_{jt}(\hat{\eta}; \hat{\eta})] \). In those expressions the first \( \eta \) stands for the sampling effect, and the second one the projection effect. Correspondingly, \( \hat{m}_{jt}(\eta; \eta) \) and \( m_{jt}(\eta; \eta) \) stand for the infeasible (“oracle”) estimates based on true \( \eta \) and its expectation. Let \( \hat{y}_{jt}(\hat{\eta}) \) denote
the estimated bounds as

\[ \hat{y}_{jt}(\hat{\eta}) = \hat{\mu}_t + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\hat{\sigma}_t \]

and \( \tilde{y}_{jt}(\eta) \) denote the expectation of bounds based on infeasible (“oracle”) estimator \( \hat{y}_{jt}(\eta_0) = \mu_t(\eta_0) + \hat{m}_{jt}(\eta_0; \eta_0)\sigma_t(\eta_0) \) assuming true parameter \( \eta_0 \) were known,

\[ \tilde{y}_{jt} \equiv E(\hat{m}_{jt}(\eta_0)) = \mu_t(\eta_0) + m_{jt}(\eta_0; \eta_0)\sigma_t(\eta_0). \]

Then, the total error \( \hat{y}_{jt}(\hat{\eta}) - y_{jt} \) can be decomposed into two parts: variance term and bias term,

\[ \hat{y}_{jt} - y_{jt} = (\hat{y}_{jt} - \tilde{y}_{jt}) + (\tilde{y}_{jt} - y_{jt}). \]

For the variance term,

\[ \hat{y}_{jt} - \tilde{y}_{jt} = \hat{\mu}_t - \mu_t + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\hat{\sigma}_t - m_{jt}(\eta_0; \eta_0)\sigma_t, \quad (4.3.16) \]

which can be further decomposed and shown that our feasible two step estimator has the oracle property that the effect caused by the presence of parametrically generated regressors is asymptotically negligible for inference. Such result relies on the fact the nonparametric convergence rate in the second step is slower than the parametric convergence rate in the first step. In the decomposition, we have

\[ \hat{y}_{jt}(\hat{\eta}) - \tilde{y}_{jt} = \left( \hat{m}_{jt}(\eta_0; \eta_0) - m_{jt}(\eta_0; \eta_0) \right)\sigma_t(\eta_0) + \sigma_t(\hat{\eta})\frac{\partial m_{jt}(\hat{\eta}; \eta)}{\partial \eta}(\hat{\eta} - \eta_0) \]

part 1

part 2

151
\[ + \left[ 1 + \frac{\partial \hat{\mu}_t(\eta; \hat{\eta})}{\partial \mu_t} \sigma_t(\eta) + \frac{\partial \hat{\sigma}_t(\eta)}{\partial \sigma_t} \right] \begin{bmatrix} \hat{\mu}_t \\ \hat{\sigma}_t \end{bmatrix} (\hat{\eta} - \eta_0). \]

Equation (4.3.17) decomposes the sampling error into three parts: part 1 is the sampling error of infeasible estimator as if the true parameter \( \eta \) were known and used in the nonparametric estimation; part 2 is the nonparametric projection error caused by using generated regressors \( \hat{\mu}_t \) and \( \hat{\sigma}_t \) in the nonparametric estimation; part 3 accounts for the accumulated sampling error from the first step estimation by inserting estimates \( \hat{\mu}_t \) and \( \hat{\sigma}_t \) into the estimated conditional mean function. Both part 2 and 3, caused by the parameter uncertainty estimated in the first step estimation, converge to zero in parametric rate under some regularity conditions, while part 1 converges to zero in nonparametric rate. Therefore, part 1 is the dominating leading term, the effects of parameter uncertainty on the two step estimator vanish asymptotically, and therefore the feasible estimator \( \hat{m}_{jt}(\hat{\eta}; \hat{\eta}) \) is asymptotically equivalent to the infeasible estimator \( \hat{m}_{jt}(\eta; \eta) \). The following theorem summarizes the asymptotic properties of feasible two step estimator.

**Theorem 4.3.1 (Asymptotic Properties of Two Step Estimator).** Given the assumptions that

(i) the estimates \( \| \hat{\eta} - \eta_0 \| = O_p(T^{-1/2}) \) and \( |\hat{m}_{jt}(\eta; \eta) - m_{jt}(\eta)| = O_p(T^{-\alpha}) \) with the optimal convergence rate \( 0 < \alpha < 1/2 \) in a neighborhood of true parameter \( \eta_0 \),

(ii) the conditional mean \( \mu_t(\eta) \) and conditional volatility \( \sigma_t(\eta) \) of the underlying stochas-
tic process $y_t$ are bounded in probability, their derivatives $\partial \mu_t(\eta)/\partial \eta$ and $\partial \sigma_t(\eta)/\partial \eta$ exist and are also bounded in probability in a neighborhood of true parameter $\eta_0$ for all $t$.

(iii) for $\eta$ and $\eta^*$ in a neighborhood of true parameter $\eta_0$ for all $t$, the nonparametric estimates $\hat{m}_{jt}(\eta; \eta^*)$ given the estimates $\eta^*$ are bounded in probability, and their derivatives with respect to $\mu_t(\eta)$ and $\sigma_t(\eta)$ are of order $O_p(T^{\delta_1})$ with $0 \leq \delta_1 < 1/2 - \alpha$,

(iv) for $\eta$ and $\eta^*$ in a neighborhood of true parameter $\eta_0$ for all $t$, the derivative of nonparametric projection $\hat{m}_{jt}(\eta; \eta^*)$ in $\eta^*$ exists and are of order $O_p(T^{\delta_2})$ with $0 \leq \delta_2 < 1/2 - \alpha$,

we have that

(1) (Asymptotic Equivalence) the feasible two step estimator $\hat{y}_{jt}(\hat{\eta}) = \hat{\mu}_t + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\hat{\sigma}_t$ is asymptotic equivalent to the infeasible estimator $\hat{y}_{jt}(\eta) = \mu_t + \hat{m}_{jt}(\eta; \eta)\sigma_t$ in the sense that

$$|\hat{y}_{jt}(\hat{\eta}) - \hat{y}_{jt}(\eta)| = o_p(T^{-\alpha})$$

for all $t$ and $j = u, l$, and,

(2) (Asymptotic Normality) the limiting distribution of the feasible two step estimator is the same as that of infeasible estimator, i.e.,

$$T^{-\alpha}(\hat{y}_{jt}(\hat{\eta}) - y_{jt} - B_{jt}(\eta)) \overset{d}{\to} N[0, V_{jt}(\eta)]$$

for all $t$ and $j = u, l$. The asymptotic bias and variance for the estimator are $B_{jt}(\eta)$

153
and \( V_{jt}(\eta) = \sigma_t^2 \cdot \text{asy.var}(\tilde{m}_{jt}(\eta;\eta)) \) whose expressions depend on the nonparametric methods used in the second step.

In Theorem 4.3.1 we use high level assumptions to obtain the oracle properties of two step estimator so that any nonparametric methods (such as local polynomials and series estimators) are applicable in the second step. The four assumptions have intuitive explanations. Assumption (i) assumes that the estimates in the first step estimation is root-\( T \) consistent and converges faster than nonparametric estimator in the second step, which is a crucial assumption for the oracle properties. If the conditional mean and variance in the first step are generated nonparametrically and do not necessarily have faster convergence rate than the nonparametric estimator in the second step, the oracle property in Theorem 4.3.1 may not hold, see Mammen, Rothe and Schienle (2010) for more detail. Assumptions (ii), (iii) and (iv) assume that the nonparametric estimator \( \hat{m}_{jt}(\eta;\eta) \) changes smooth both in its variables \( \mu_t \) and \( \sigma_t \), which depend on the first \( \eta \), and in the projections on its variables \( \mu_t \) and \( \sigma_t \), which depend on the second \( \eta \). For different nonparametric methods employed in the second step, more primitive assumptions, though less transparent in explanations, exist to guarantee the oracle properties for the feasible two step estimator. Borrowing from Conrad and Mammen (2009), we give those primitive assumptions for the (kernel-based) local linear and local constant estimators in the appendix.
4.4 The Five-minute Low/High Stock Returns of Wells Fargo

We model the interval-valued time series of the high frequency five-minute low/high returns to the stock of Wells Fargo, trading in the New York Stock Exchange, from January 3rd to January 31st, 2011 for a total of 20 trading days. Both the stock price and the number of trades at the 5-minute frequency are also recorded. The returns are computed with respect to the last price of the previous five-minute period, that is,

\[ r_{ht} = \frac{P_{high,t} - P_{close,t-1}}{P_{close,t-1}} \times 10,000\% \]
\[ r_{lt} = \frac{P_{low,t} - P_{close,t-1}}{P_{close,t-1}} \times 10,000\% \]

where \( P_{high,t} \) and \( P_{low,t} \) are the highest and lowest price in the five-minute period \( t \), and \( P_{close,t-1} \) is the last price in the previous five-minute period \( t - 1 \). We remove any trades before 9:30 am (opening time) and after 4:00 pm (closing time) for a total of 1,560 observations. We plot the low/high returns and the number of trades in Figure 4.1 and present the descriptive statistics for the three series in Table 4.1.

The first and second column in Table 4.1 show the descriptive statistics of low and high returns. Although both low and high returns exhibit large volatilities and heavy tails, it is noticeable that low returns possess even larger variance and heavier tails than high returns do, which implies that the beneficiary and adversary shocks on the stock prices are asymmetric. In addition, the correlation of low and high returns is 0.392. The positive co-movements of low and high returns (with correlation 0.392) suggests that, in
Figure 4.1: Time Series Plots of 5-min Low/High Returns and Number of Trades

(a) 5-min Low/High Returns (in %)

(b) 5-min Number of Trades
<table>
<thead>
<tr>
<th>Statistics</th>
<th>Low Return</th>
<th>High Return</th>
<th>Center</th>
<th>Radius</th>
<th>Number of Trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>-536.8</td>
<td>-216.1</td>
<td>-337.4</td>
<td>1.464</td>
<td>103</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>-15.96</td>
<td>3.505</td>
<td>-5.267</td>
<td>7.03</td>
<td>810</td>
</tr>
<tr>
<td>Median</td>
<td>-9.084</td>
<td>8.997</td>
<td>0</td>
<td>10</td>
<td>1210</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>-3.509</td>
<td>15.79</td>
<td>5.058</td>
<td>14.46</td>
<td>1850</td>
</tr>
<tr>
<td>Maximum</td>
<td>251.1</td>
<td>324.2</td>
<td>287.7</td>
<td>275.5</td>
<td>17310</td>
</tr>
<tr>
<td>Mean</td>
<td>-12.44</td>
<td>12.1</td>
<td>-0.1725</td>
<td>12.27</td>
<td>1528</td>
</tr>
<tr>
<td>Variance</td>
<td>325.037</td>
<td>254.078</td>
<td>201.110</td>
<td>88.448</td>
<td>1449407</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.392</td>
<td>-0.133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>-7.301</td>
<td>3.117</td>
<td>-2.513</td>
<td>5.442</td>
<td>3.744</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>153.002</td>
<td>60.394</td>
<td>107.603</td>
<td>83.575</td>
<td>26.309</td>
</tr>
</tbody>
</table>

Table 4.1: Descriptive Statistics

a short period, either the beneficiary or adversary shock on the stock price dominates, inducing the low and high returns to increase or decrease in the same direction. The descriptive statistics for center and radius are reported in the third and fourth column of the table. The center and radius are naturally defined as the middle point and one half of the interval length.

To start with the estimation of the proposed two-step models in this empirical application, we simply use the center values of the low and high stock returns $r_{ct} = (r_{ht} + r_{lt})/2$ as a proxy for the underlying latent process, since the underlying process for the interval-valued time series is latent. In the first step, a AR(1)-GARCH(1,1) model is used to characterize and model the center values of the interval times series. The first step estimation produces estimated conditional means $\hat{\mu}_{t|t-1}$ and standard deviations $\hat{\sigma}_{t|t-1}$ capturing the underlying distributions that generate the extremes (low and high returns). Then, in the second step, we consider nonparametric regression models where the dependent variables are the extremes $r_{lt}$ and $r_{ht}$ with or without standardization by the estimated conditional means $\hat{\mu}_{t|t-1}$ and variances $\hat{\sigma}_{t|t-1}$. The regressors in the nonparametric regressions are the generated conditional mean $\hat{\mu}_{t|t-1}$, conditional stan-
standard deviation $\hat{\sigma}_{t|t-1}$, and number of trades $n_t$. In addition, we also consider the case in which the number of trades were unobservable and excluded from the nonparametric regression, so that the effect of number of trades on the accuracy of model fitting can be revealed through model comparison.

Our proposed models are summarized below,

- In the first step, obtain conditional mean $\hat{\mu}_{t|t-1}$ and standard deviation $\hat{\sigma}_{t|t-1}$ from the AR(1)-GARCH(1,1) model for the center values of the low and high returns, i.e., $r_{ct} \equiv (r_{lt} + r_{ht})/2$;

- In the second step, we consider nonparametric regressions with several different specifications listed below,

  - **Model P1**: standardize $r_{lt}$ and $r_{ht}$ with $\hat{\mu}_{t|t-1}$ and $\hat{\sigma}_{t|t-1}$ as $r_{jt}^* = (r_{jt} - \hat{\mu}_{t|t-1})/\hat{\sigma}_{t|t-1}$ for $j = l, h$, and consider the nonparametric regression

    \[ r_{jt}^* = m_j(\hat{\mu}_{t|t-1}, \hat{\sigma}_{t|t-1}, n_t) + e_{jt}, \quad \text{for } j = l, h. \]

  - **Model P2**: use standardized $r_{jt}^*$ and exclude the number of trades $n_t$ from the nonparametric regression

    \[ r_{jt}^* = m_j(\hat{\mu}_{t|t-1}, \hat{\sigma}_{t|t-1}) + e_{jt}, \quad \text{for } j = l, h. \]

  - **Model P3**: use standardized $r_{jt}^*$ and exclude the estimated conditional mean
\[ \hat{\mu}_{t|t-1} \text{ and standard deviation } \hat{\sigma}_{t|t-1} \text{ from the nonparametric regression} \]

\[ r_{jt}^* = m_j(n_t) + e_{jt}, \quad \text{for } j = l, h. \]

- **Model P4**: use \( r_{jt} \) without standardization and include both number of trades \( n_t \) and the estimated conditional mean \( \hat{\mu}_{t|t-1} \) and standard deviation \( \hat{\sigma}_{t|t-1} \) in the nonparametric regression

\[ r_{jt} = m_j(\hat{\mu}_{t|t-1}, \hat{\sigma}_{t|t-1}, n_t) + e_{jt}, \quad \text{for } j = l, h. \]

We compare our two-step estimators with the estimation techniques proposed into current literature. We implement the approaches suggested by Lima Neto and De Carvalho (2008, 2010), henceforth LNC, and we also estimate a location-scale model from which we construct interval estimates.

Given an interval-valued time series \( \{r_t\} = \{[r_{lt}, r_{ht}]\} \), we obtain the time series of the centers and the time series of the radius. LNC estimate the following system of equations

\[ r_{ct} = \beta_0^c + \beta_1^c r_{c,t-1} + \cdots + \beta_p^c r_{c,t-p} + \epsilon_t^c \quad \text{(4.4.18)} \]
\[ r_{rt} = \beta_0^r + \beta_1^r r_{r,t-1} + \cdots + \beta_p^r r_{r,t-p} + \epsilon_t^r. \quad \text{(4.4.19)} \]

Their center/range method (CRM) consists of estimating each equation by least squares; and their constrained center/range method (CCRM) consists of imposing the restriction \( \beta_j^r \geq 0, j = 0, \ldots, p \) on the equation of the radius to ensure that \( \hat{y}_{rt} \geq 0 \) and, therefore, \( \hat{y}_{lt} \leq \hat{y}_{ut} \). Then, the equation of the center is estimated by least squares and the constrained equation of the radius is estimated by adapting Lawson and Hanson’s
algorithm. Note that the LNC system of center/radius equations is a special case of a vector autoregressive system of lower/upper bound equations as follows,

\[
\begin{bmatrix}
  r_{lt} \\
  r_{ht}
\end{bmatrix}
= \begin{bmatrix}
  \beta_{le} \\
  \beta_{hc}
\end{bmatrix}
+ \sum_{j=1}^{p}
\begin{bmatrix}
  \beta_{11}^{(j)} & \beta_{12}^{(j)} \\
  \beta_{21}^{(j)} & \beta_{22}^{(j)}
\end{bmatrix}
\begin{bmatrix}
  y_{lt, t-j} \\
  y_{ht, t-j}
\end{bmatrix}
+ \begin{bmatrix}
  \varepsilon_{lt} \\
  \varepsilon_{ht}
\end{bmatrix}.
\]

(4.4.20)

Therefore, in addition to the LNC system of center/radius equations (4.4.18) and (4.4.19), we also include the vector autoregressive system of low/high equations into the group of models for comparison.

The second set of comparisons is with a location-scale model applied to the time series of centers, for which we estimate the GARCH(1,1) model whose the i.i.d. standardized error term \( \zeta_t \) follows a standard normal or Student-\( t \) density with \( \nu \) degrees of freedom.

Based on this model, we construct \((1 - \alpha)\)-probability intervals, which will depend on the distributional assumptions on \( \zeta_t \), i.e.,

\[
\begin{align*}
[\hat{r}_{lt}, \hat{r}_{ht}]_\alpha &= \left[ \hat{r}_{ct} - z_{\frac{\alpha}{2}} \hat{\sigma}_t, \hat{r}_{ct} + z_{\frac{\alpha}{2}} \hat{\sigma}_t \right] \\
[\hat{r}_{lt}, \hat{r}_{ht}]_\alpha &= \left[ \hat{r}_{ct} - t_{\frac{\nu}{2}} \hat{\sigma}_t \sqrt{\frac{\nu}{\nu - 2}}, \hat{r}_{ct} + t_{\frac{\nu}{2}} \hat{\sigma}_t \sqrt{\frac{\nu}{\nu - 2}} \right].
\end{align*}
\]

Since the original data \([r_{lt}, r_{ht}]\) are the observed extreme values of the process at time \( t \), we will stretch the estimated interval \([\hat{r}_{lt}, \hat{r}_{ht}]_\alpha\) to cover as much as 99% or 99.5% probability, so that \( \hat{r}_{lt} \) and \( \hat{r}_{ht} \) are far away into the tails of the distribution.

We summarize the eight methodologies that are to be compared with our proposed two-step estimators as follow,

- **CCRM**: Constrained Center and Range Method
• **CRM**: Center and Range Method

• **VAR**: Vector Autoregressive Method

• **GARCH-N (99%)**: 99% C.I. from GARCH(1,1) with conditional normality

• **GARCH-N (99.5%)**: 99.5% C.I. from GARCH(1,1) with conditional normality

• **GARCH-T (99%)**: 99% C.I. from GARCH(1,1) with conditional Student-t

• **GARCH-T (99.5%)**: 99.5% C.I. from GARCH(1,1) with conditional Student-t

We evaluate the performance of each estimation methodology according to four criteria: (i) Root Mean Squared Error (RMSE) for each of the upper and lower bounds, (ii) Coverage (CR) and Efficiency Rates (ER) of the estimated intervals (Rodrigues and Salish, 2011), (iii) Multivariate Loss Functions (MLF) for the vector of lower and upper bounds (Komunjer and Owyang, 2011), and (iv) Mean Distance Error (MDE) between the fitted and actual intervals (Arroyo et al., 2011).

For a sample of size $T$, let us call $\hat{y}_t = [\hat{y}_{lt}, \hat{y}_{ut}]$ the fitted values of the corresponding interval $y_t = [y_{lt}, y_{ut}]$ obtained by each methodology. These are the definitions of the four criteria:

(i) RMSE:

\[
RMSE_l = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{lt} - y_{lt})^2},
\]

\[
RMSE_u = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{ut} - y_{ut})^2}.
\]
(ii) CR and ER:

\[
CR = \frac{1}{T} \sum_{t=1}^{T} \frac{w(y_t \cap \hat{y}_t)}{w(y_t)} ,
\]

\[
ER = \frac{1}{T} \sum_{t=1}^{T} \frac{w(y_t \cap \hat{y}_t)}{w(\hat{y}_t)} .
\]

where \( y_t \cap \hat{y}_t \) is the intersection of the actual and fitted intervals, and \( w(\cdot) \) represents the range or width of the interval.

The coverage rate is the average proportion of the actual interval that is covered by the fitted interval, and the efficiency rate is the average proportion of the fitted interval that is covered by the actual interval. Both rates are between zero and one and a large rate means a better fit. However, given an actual interval, a wide fitted interval will imply a large coverage rate but a low efficiency rate, on the contrary, a tight fitted interval will imply a low coverage rate but a high efficiency rate. Therefore, we take into account the potential trade-off between the two rates by calculating an average of the two rates, i.e., \((CR + ER)/2\).

(iii) MLF: We implement the following multivariate loss function

\[
L_p(\tau, e) \equiv ( \| e \|_p + \tau' e ) \| e \|_p^{p-1}
\]

where \( \| \cdot \|_p \) is the \( l_p \)-norm, \( \tau \) is two-dimensional parameter vector bounded by the unit ball \( B_q \) in \( \mathbb{R}^2 \) with \( l_q \)-norm (where \( p \) and \( q \) satisfy \( 1/p + 1/q = 1 \)), and \( e = (e_l, e_u) \) is the bivariate residual interval \((\hat{y}_{lt} - y_{lt}, \hat{y}_{ut} - y_{ut})\). We consider two norms, \( p = 1 \) and \( p = 2 \) and their corresponding \( \tau \) parameter vectors within the unit balls \( B_\infty \) and \( B_2 \).
respectively,

\[
MLF_1 = \int_{\tau \in B_{\infty}} (|e_l| + |e_u| + \tau_1 e_l + \tau_2 e_u) d\tau
\]

\[
MLF_2 = \int_{\tau \in B_2} \left[ e_l^2 + e_u^2 + (\tau_1 e_l + \tau_2 e_u)(e_l^2 + e_u^2)^{1/2} \right] d\tau.
\]

(iv) MDE: Let \( D^q(\hat{y}_t, y_t) \) be a distance measure of order \( q \) between the fitted and the actual intervals, the mean distance error is defined as

\[
MDE^q(\{\hat{y}_t\}, \{y_t\}) = \left( \frac{\sum_{t=1}^T D^q(\hat{y}_t, y_t)}{T} \right)^{1/q}
\]

We consider two cases, \( q = 1 \) and \( q = 2 \), with a distance measure such as

\[
D(\hat{y}_t, y_t) = \frac{1}{\sqrt{2}} \sqrt{(\hat{y}_{lt} - y_{lt})^2 + (\hat{y}_{ut} - y_{ut})^2}.
\]
<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE</th>
<th>CR &amp; ER</th>
<th>MLF</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>CR</td>
<td>ER</td>
</tr>
<tr>
<td>CCRM</td>
<td>15.0914</td>
<td>14.0978</td>
<td>0.7635</td>
<td>0.6687</td>
</tr>
<tr>
<td>CRM</td>
<td>15.0914</td>
<td>14.0978</td>
<td>0.7635</td>
<td>0.6687</td>
</tr>
<tr>
<td>VAR</td>
<td>15.0211</td>
<td>14.0264</td>
<td>0.7651</td>
<td>0.6703</td>
</tr>
<tr>
<td>GARCH(1,1)-N (99%)</td>
<td>29.7162</td>
<td>29.8141</td>
<td>0.9681</td>
<td>0.3835</td>
</tr>
<tr>
<td>GARCH(1,1)-N (99.5%)</td>
<td>32.7702</td>
<td>32.9406</td>
<td>0.9752</td>
<td>0.3568</td>
</tr>
<tr>
<td>GARCH(1,1)-T (99%)</td>
<td>30.5207</td>
<td>30.7042</td>
<td>0.9779</td>
<td>0.3521</td>
</tr>
<tr>
<td>GARCH(1,1)-T (99.5%)</td>
<td>38.4883</td>
<td>38.8186</td>
<td>0.9869</td>
<td>0.2946</td>
</tr>
<tr>
<td>Model P1 (N-W)</td>
<td>13.9047</td>
<td>12.3535</td>
<td>0.7689</td>
<td>0.6815</td>
</tr>
<tr>
<td>Model P2 (N-W)</td>
<td>15.2977</td>
<td>13.7108</td>
<td>0.7680</td>
<td>0.6655</td>
</tr>
<tr>
<td>Model P3 (N-W)</td>
<td>16.2392</td>
<td>15.5449</td>
<td>0.7626</td>
<td>0.6703</td>
</tr>
<tr>
<td>Model P4 (N-W)</td>
<td>14.2463</td>
<td>13.0747</td>
<td>0.7566</td>
<td>0.6775</td>
</tr>
<tr>
<td>Model P1 (Local Linear)</td>
<td>12.2770</td>
<td>11.4120</td>
<td>0.7667</td>
<td>0.7055</td>
</tr>
<tr>
<td>Model P2 (Local Linear)</td>
<td>15.0293</td>
<td>13.7581</td>
<td>0.7655</td>
<td>0.6687</td>
</tr>
<tr>
<td>Model P3 (Local Linear)</td>
<td>16.4486</td>
<td>15.6407</td>
<td>0.7589</td>
<td>0.6765</td>
</tr>
<tr>
<td>Model P4 (Local Linear)</td>
<td>12.4550</td>
<td>10.3562</td>
<td>0.7690</td>
<td>0.7098</td>
</tr>
</tbody>
</table>

Table 4.2: Methodology Evaluation for 5-minute Low/High Interval Stock Returns of Wells Fargo Corporation (WFC)
In Tables 4.2 we compare the performance of the different estimation techniques in the modelling of the interval-valued time series of the five-minute low/high stock returns to the Wells Fargo. The numbers in boldface correspond to the losses of our proposed models (Model P1 and P4) outperforming those competing models when we consider the functions RMSE, MLF, and MDE, and to the maximum rates when we consider the weighted CR/ER rates. These are the main findings:

1. Across methods, Model P1 and P4 with both local constant and local linear smoothing are superior to CCRM and CRM, and these are far better than the GARCH models with confidence intervals. The classical methodology embedded in normal or fat-tail location-scale models is by far the worst performer across all evaluation functions and it is very inefficient on delivering an acceptable fitted interval as the efficiency rates (ER) shows.

2. Across our proposed four semiparametric models, either Model P1 or P4 has the best performance for both local constant and local linear smoothing; and the performances of both Model P2 and P3 resemble those of competing models. Note that Model P2 excludes the number of trades \( n_t \) from the nonparametric regression, and Model P3 excludes the estimated conditional mean \( \hat{\mu}_{t|t-1} \) and standard deviation \( \hat{\sigma}_{t|t-1} \). Such results imply that the information of number of trades \( n_t \) contributes to the estimation of low/high returns as extreme values.

3. In Figure 4.2, Figure 4.3, and Figure 4.4, we plot the actual low/high returns together with fitted ones from different models. The findings are better illustrated through the comparison among those figures. The estimates from the simple VAR model, shown in Figure 4.2a fail to capture the spikes of low/high returns, which is reasonable since
such model does not exploiting the dynamics of conditional volatility. The estimates from the confidence intervals derived from GARCH models exploiting the dynamics of conditional volatility, on the contrary, tend to exaggerate the low/high returns, as it is shown in Figure 4.2b where we only plot the estimates from GARCH(1,1)-Normal with 99% confidence intervals as an example. For the our proposed models, the estimates from Model P1 and P4 with local linear smoothing fit almost equally well by picking up some of the spikes and most of the dynamics in the low/high returns, shown in Figure 4.3a and Figure 4.4b respectively. After the number of trades $n_t$ are excluded, the estimates from Model P2 fail to capture those spikes in low/high returns. In contrast, Model P3 excludes the estimated conditional mean and volatility, and the only regressor in the nonparametric regression is the number of trades. Although the estimates, shown in Figure 4.4a, pick the spikes in low/high returns, however, some of the spikes are falsely captured. Therefore, such comparison among the estimates from the four models suggest that both conditional moments and number of trades are crucial in correctly picking up the spikes and dynamics of low/high returns.
Figure 4.2: Fitted Returns from Competing Models

(a) Fitted Low/High Returns from VAR Model

(b) Fitted Returns from GARCH(1,1)-Normal with 99% Confidence Intervals
(a) Fitted Returns from Model P1 with Local Linear Smoothing

(b) Fitted Returns from Model P2 with Local Linear Smoothing

Figure 4.3: Fitted Returns from Proposed Models (P1 and P2)
(a) Fitted Returns from Model P3 with Local Linear Smoothing

(b) Fitted Returns from Model P4 with Local Linear Smoothing

Figure 4.4: Fitted Returns from Proposed Models (P3 and P4)
4.5 Conclusion

In many disciplines, massive data sets are generated and stored in an aggregate format such as intervals or histograms. Existing models in the current literature for interval-valued time series mainly focus on the regressions of minimum and maximum (or center and radius) of intervals on its lag terms with no or restrictive assumption on the distribution of error terms. Such specification ignores the extreme nature of the lower and upper bounds of intervals.

In this paper, we assume that there is an underlying stochastic process for the interval-valued time series, and that the lower and upper bounds of the intervals are the realized extreme observations (minima and maxima) based on the random draws from the conditional densities of underlying stochastic process which is strictly stationary. The specification of this data generating process decomposes the analysis of interval-valued time series into two parts. The first part is the classical point-valued time series analysis for the underlying stochastic process, for which the time series of centers is used as a proxy. The second part is to model the conditional mean of the lower/upper bounds nonparametrically, since the conditional mean of extreme value is often highly nonlinear and intractable. Following this methodology, this paper combines the expertise of the two fields to analyze the interval-valued time series data, and propose a two step procedure for the estimation of interval-valued time series: construct and estimate a parametric model for the time series of centers, and obtain the estimated conditional moments of it; and then perform nonparametric regression of lower/upper bounds with the estimated conditional moments as the regressors. Bases on the literature on nonparametric regression with generated regressors, the effect of parameter uncertainty in the second step is
asymptotically negligible given some regularity conditions, and therefore, our two-step estimator has typical nonparametric convergence rate.

In the empirical application, we apply our proposed two step estimator to the data of five-minute low/high stock returns of Wells Fargo Corporation. In the first step, we obtain conditional mean and standard deviation from the AR(1)-GARCH(1,1) model for the center values of the low/high returns, and in the second step, we perform nonparametric regression (local constant and local linear smoothing) by regressing low/high returns on the number of trades and estimated conditional moments generated in the first step. We then compare our two-step estimator with its several competing approaches, including (constrained) center-radius model, VAR model and location-scale model with confidence intervals. The performance of each estimation methodology are evaluated according to four criteria: root mean squared errors, coverage and efficiency rates, multivariate loss functions, and mean distance error. The comparison shows that our proposed estimators are superior to the competing approaches, and that both estimated moments and number of trades are indispensable in capturing the spikes and dynamics of low/high returns correctly.
Bibliography


Appendix A

Estimation of \((\sigma_l, \sigma_u, \rho)\) in Two-step Estimation

Under the normality assumption, we have,

\[
m_{10} = E(\varepsilon_l | \varepsilon_u - \varepsilon_l \geq \Delta) = \frac{\rho \sigma_l \sigma_u - \sigma_l^2}{\sigma_m} \frac{\phi(\Delta/\sigma_m)}{1 - \Phi(\Delta/\sigma_m)} \tag{A.1}
\]

\[
m_{01} = E(\varepsilon_u | \varepsilon_u - \varepsilon_l \geq \Delta) = \frac{\sigma_u^2 - \rho \sigma_l \sigma_u}{\sigma_m} \frac{\phi(\Delta/\sigma_m)}{1 - \Phi(\Delta/\sigma_m)} \tag{A.2}
\]

\[
m_{20} = E(\varepsilon_u^2 | \varepsilon_u - \varepsilon_l \geq \Delta) = \sigma_u^2 + \frac{\sigma^2(\rho \sigma_u - \sigma_l)^2}{\sigma_m} \frac{\Delta}{1 - \Phi(\Delta/\sigma_m)} \tag{A.3}
\]

\[
m_{02} = E(\varepsilon_u^2 | \varepsilon_u - \varepsilon_l \geq \Delta) = \sigma_u^2 + \frac{\sigma_u^2 (\sigma_u - \rho \sigma_l)^2}{\sigma_m} \frac{\Delta}{1 - \Phi(\Delta/\sigma_m)} \tag{A.4}
\]

\[
m_{11} = E(\varepsilon_l \varepsilon_u | \varepsilon_u - \varepsilon_l \geq \Delta) = \rho \sigma_l \sigma_u + \frac{\sigma_l \sigma_u (\rho \sigma_u - \sigma_l)(\sigma_u - \rho \sigma_l) \Delta}{\sigma_m^2} \frac{\phi(\Delta/\sigma_m)}{1 - \Phi(\Delta/\sigma_m)} \tag{A.5}
\]

To easy the notation, \(\Delta\) denotes \(\Delta(y^{t-1}, \Delta \beta)\) and all the subscripts \(t\) are dropped. In the two-step estimation, \(\Delta\) and \(\sigma_m\) are consistently estimated. The first and second moments of \(\varepsilon_u\) and \(\varepsilon_l\), conditioning on the observability restriction, can be written as
\( \tilde{m}_{10}, \tilde{m}_{01}, \tilde{m}_{20}, \tilde{m}_{02}, \tilde{m}_{11} \), by plugging the estimates \( \hat{\Delta} \) and \( \hat{\sigma}_m \) into the expressions (A.1) - (A.5). Therefore, the parameters \((\sigma_l, \sigma_u, \rho)\) can be estimated by the simple method of moments as follows

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{lt}^2 = \frac{1}{T} \sum_{t=1}^{T} \{\tilde{m}_{20}(t) - |\tilde{m}_{10}(t)|^2\}
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{ut}^2 = \frac{1}{T} \sum_{t=1}^{T} \{\tilde{m}_{02}(t) - |\tilde{m}_{01}(t)|^2\}
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{lt} \hat{u}_{ut} = \frac{1}{T} \sum_{t=1}^{T} \{\tilde{m}_{11}(t) - |\tilde{m}_{10}(t)\tilde{m}_{01}(t)|\}
\]

where \( \hat{u}_{lt} \) and \( \hat{u}_{ut} \) are the residuals of the second step regression,

\[
\hat{u}_{lt} = y_{lt} - \hat{\beta}_{lt} - \sum_{j=1}^{p} \hat{\beta}_{11}^{(j)} y_{l,t-j} - \sum_{j=1}^{p} \hat{\beta}_{12}^{(j)} y_{u,t-j} - \hat{C}_l \hat{\lambda}_{l-1}
\]

\[
\hat{u}_{ut} = y_{ut} - \hat{\beta}_{uc} - \sum_{j=1}^{p} \hat{\beta}_{21}^{(j)} y_{l,t-j} - \sum_{j=1}^{p} \hat{\beta}_{22}^{(j)} y_{u,t-j} - \hat{C}_u \hat{\lambda}_{l-1}
\]
Appendix B

Proof of Theorem 1.3.2

(a) Consistency

We only consider the regression for lower bounds $y_{l,t}$, the same reasoning applies to the estimators for the upper bound model. From the equation (1.3.18),

$$
\begin{pmatrix}
\hat{\beta}_l \\
\hat{C}_l
\end{pmatrix} = \hat{\gamma}_l = \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}'y_l \\
= \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}' [\hat{H}\gamma_l + u_l] \\
= \gamma_l + \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}'u_l
$$

in which $u_l \equiv C_l \left( \Lambda - \hat{\Lambda} \right) + v_l$. Note that,

$$
\plim_{T \to \infty} \left( \hat{H}'\hat{H} \right)^{-1} \hat{H}'u_l = \plim_{T \to \infty} \left[ \frac{\hat{H}'\hat{H}}{T} \right]^{-1} \plim_{T \to \infty} \frac{\hat{H}'u_l}{T}
$$

Call $D \equiv \hat{H} - H = \epsilon' \otimes \left( \hat{\Lambda} - \Lambda \right)$, defining the row vector $\epsilon' \equiv (0, \cdots, 0, 1)$ taking the
value of 1 for the last element and 0 otherwise. Note that,

\[
\Lambda - \hat{\Lambda} = -\mathbf{J}(\Delta \beta^*) (\Delta \hat{\beta}^* - \Delta \beta_0^*)
\]

\[
\mathbf{D} = \mathbf{J}'(\Delta \beta^*) (\Delta \hat{\beta}^* - \Delta \beta_0^*)
\]

\[
\mathbf{u}_t = C_t (\Lambda - \hat{\Lambda}) + \mathbf{v}_t
\]

\[
= -C_t \mathbf{J}(\Delta \beta^*) (\Delta \hat{\beta}^* - \Delta \beta_0^*) + \mathbf{v}_t.
\]

Given assumptions (i) and (ii),

\[
\frac{1}{T} \mathbf{H}' \mathbf{H} = O_p(1)
\]

\[
\frac{1}{T} \mathbf{H}' \mathbf{J}(\Delta \beta^*) = O_p(1)
\]

\[
\frac{1}{T} \mathbf{J}'(\Delta \beta^*) \mathbf{J}(\Delta \beta^*) = O_p(1)
\]

and by construction \(\mathbf{v}_t\) is a martingale difference sequence with respect to information set \(\mathcal{F}_{t-1}\), and conditioning in the observability restriction \(y_{lt} \leq y_{ut}\).

\[
E(v_{lt}|\mathcal{F}_{t-1}) = E(\varepsilon_{lt} - E(\varepsilon_{lt}|\Delta \varepsilon_t \geq \sigma_m \Delta(y^{t-1}, \Delta \beta^*)|\mathcal{F}_{t-1}))
\]

\[
= E(\varepsilon_{lt}|\mathcal{F}_{t-1}) - C_t \lambda_{t-1} = 0
\]

and thus we have \(E(h_{lt-1}v_{lt}|\mathcal{F}_{t-1}) = 0\) for all \(t\). Given assumptions (iii) and (iv), and by the central limit theorem for martingale difference sequence, we have

\[
\frac{1}{\sqrt{T}} \mathbf{J}'(\Delta \beta^*) \mathbf{v}_l = O_p(1)
\]

\[
\frac{1}{\sqrt{T}} \mathbf{H}' \mathbf{v}_l \overset{d}{\to} N(0, \Psi_l).
\]

\(^1\mathcal{F}_{t-1}\) denotes the information set available at time \(t - 1\). Similarly, \(E(v_{mt}|\mathcal{F}_{t-1}) = 0\).
Then, we can prove that

\[
\frac{\hat{H}'\hat{H}}{T} = O_p(1),
\]

\[
\frac{\hat{H}'u_L}{T} = O_p(T^{-1/2}).
\]

For the first equation,

\[
\frac{1}{T} \hat{H}'\hat{H} = \frac{1}{T}(H' + D')(H + D)
\]

\[
= \frac{1}{T} H'H
\]

\[
+ \frac{1}{T} \ell' \otimes H'J(\Delta \beta^*) (\Delta \beta^* - \Delta \beta_0^*)' + \left( \frac{1}{T} \ell' \otimes H'J(\Delta \beta^*) (\Delta \beta^* - \Delta \beta_0^*) \right)'
\]

\[
+ \ell' \otimes (\Delta \beta^* - \Delta \beta_0^*) \frac{J'(\Delta \beta^*)J(\Delta \beta^*)}{T} (\Delta \beta^* - \Delta \beta_0^*)
\]

\[
= O_p(1) + O_p(T^{-1/2}) + O_p(T^{-1})
\]

\[
\gg \frac{1}{T} H'H
\]

(B.1)

For the second equation,

\[
\frac{\hat{H}'u}{T} = \frac{1}{T}(H' + D')[C_t(\Lambda - \hat{\Lambda}) + v_l]
\]

\[
= \frac{1}{T} H'v_L - C_t \frac{H'J(\Delta \beta^*)}{T} (\Delta \beta^* - \Delta \beta_0^*)
\]

\[
- C_t \ell \otimes (\Delta \beta^* - \Delta \beta_0^*)' \frac{J'(\Delta \beta^*)J(\Delta \beta^*)}{T} (\Delta \beta^* - \Delta \beta_0^*)
\]

\[
+ \ell \otimes (\Delta \beta^* - \Delta \beta_0^*) \frac{J'(\Delta \beta^*)v_l}{T}
\]

\[
= O_p(T^{-1/2}) + O_p(T^{-1/2}) + O_p(T^{-1}) + O_p(T^{-1})
\]

\[
= O_p(T^{-1/2})
\]

\[
\gg \frac{1}{T} H'v_l - C_t \frac{H'J(\Delta \beta^*)}{T} (\Delta \beta^* - \Delta \beta_0^*)
\]

(B.2)
Hence,
\[
\hat{\gamma}_l - \gamma_l = O_p(1)O_p(T^{-1/2}) = O_p(T^{-1/2})
\]

Therefore, the two-step estimator \( \hat{\gamma}_L \) is consistent, i.e., \( \text{plim}_{T \to \infty} \hat{\gamma}_l = \gamma_l \)

(b) Asymptotic Normality

Now we consider the asymptotic distribution of the two-step estimator,

\[
\sqrt{T}(\hat{\gamma}_l - \gamma_l) = \left( \frac{\hat{H}'\hat{H}}{T} \right)^{-1} \frac{\hat{H}'u_l}{\sqrt{T}}
\]

From equation (B.1) and (B.2), we have,
\[
\sqrt{T}(\hat{\gamma}_l - \gamma_l) \overset{p}{\to} \left( \frac{H'H}{T} \right)^{-1} \left( \frac{1}{\sqrt{T}}H'v_l - C_l \frac{H'J(\Delta\beta^*)}{T} \sqrt{T}(\Delta\beta^* - \Delta\beta_0^*) \right)
\overset{d}{\to} N(0, B\Xi B')
\]

where
\[
B = \text{plim}_{T \to \infty} \left( \frac{H'H}{T} \right)^{-1}
\]

\[
\text{var} \left( \frac{1}{\sqrt{T}}H'v_l + C_l \frac{H'(\Lambda - \hat{\Lambda})}{\sqrt{T}} \right)
= \frac{1}{T} E(H'v_lv_l'H) + C_l^2 E \left( \frac{H'J(\beta^*)}{T} S'J(\beta^*)H \right) + E \left( \frac{H'v_l(\Lambda - \hat{\Lambda})H'C_l}{T} \right) + E \left( \frac{H'(\Lambda - \hat{\Lambda})v_l'H}{T} C_l \right)
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} E(h_{t-1}h'_{t-1}v_{lt}^2) + C_t^2 \hat{Q}' \hat{Q} + M_{IT} + M'_{IT} \\
\to \Psi_t + C_t^2 \hat{Q}'_0 \hat{Q}_0 + M_{t0} + M'_{t0} \\
\equiv \Xi_t
\]

where the second equality holds because \( h_{t-1}v_{Lt} \) is a martingale difference sequence.
Appendix C

Proof of Proposition 1.3.1

We only need to prove (1.3.35), (1.3.36), and (1.3.38), since (1.3.37) and (1.3.39) can be derived similarly.

(i) Proof of (1.3.35):

$$\frac{1}{T} \sum_{t=1}^{T} \Delta v_t^2 = \frac{1}{T} \sum_{t=1}^{T} (\Delta y_t + z_{t-1} \Delta \hat{\beta} - \hat{\sigma}_m \lambda_{t-1})^2$$

$$= \frac{1}{T} \sum_{t=1}^{T} \Delta v_t^2 + \frac{1}{T} \sum_{t=1}^{T} z_{t-1} (\Delta \hat{\beta} - \Delta \hat{\beta}) (\Delta \hat{\beta} - \Delta \hat{\beta})' z_{t-1}$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \hat{\sigma}_m^2 (\lambda_{t-1} - \hat{\lambda}_{t-1})^2 + \frac{1}{T} \sum_{t=1}^{T} \lambda_{t-1}^2 (\sigma_m - \hat{\sigma}_m)^2$$

$$+ \frac{2}{T} \sum_{t=1}^{T} z_{t-1} (\Delta \hat{\beta} - \Delta \hat{\beta}) \Delta v_t + \frac{2}{T} \sum_{t=1}^{T} \hat{\sigma}_m (\lambda_{t-1} - \hat{\lambda}_{t-1}) \Delta v_t$$

$$+ \frac{2}{T} \sum_{t=1}^{T} \lambda_{t-1} (\sigma_m - \hat{\sigma}_m) \Delta v_t + \frac{2}{T} \sum_{t=1}^{T} z_{t-1} (\Delta \hat{\beta} - \Delta \hat{\beta}) (\lambda_{t-1} - \hat{\lambda}_{t-1}) \hat{\sigma}_m$$

$$+ \frac{2}{T} \sum_{t=1}^{T} z_{t-1} (\Delta \hat{\beta} - \Delta \hat{\beta}) \lambda_{t-1} (\sigma_m - \hat{\sigma}_m)$$
\[
+ \frac{2}{T} \sum_{t=1}^{T} \hat{\sigma}_m (\lambda_{t-1} - \hat{\lambda}_{t-1}) \lambda_{t-1} (\sigma_m - \hat{\sigma}_m)
\]

In the above expression, the first term is

\[
\frac{1}{T} \sum_{t=1}^{T} \Delta v_t^2 \rightarrow \frac{1}{T} \sum_{t=1}^{T} \text{var}(\Delta v_t | y_{t-1}) \rightarrow \text{E}(\text{var}(\Delta v_t | y_{t-1})). \quad (C.1)
\]

In (C.1), the first convergence in probability is because of the Law of Large Numbers for mixing sequence. The second convergence in probability follows because of the ergodic theorem, since the assumptions 1 and 5 on the stationarity and mixing properties of \{Y_t\} imply its ergodicity. Therefore, we only need to prove that the rest of the terms in the summation converges to zero in probability. In the rest of the proof, we will be using the following property extensively \(\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)\).

For the second term,

\[
\text{vec} \left( \frac{1}{T} \sum_{t=1}^{T} z_{t-1} (\overline{\Delta \beta} - \Delta \beta)(\overline{\Delta \beta} - \Delta \beta)' z_{t-1}' \right) = \frac{1}{T} \sum_{t=1}^{T} (z_{t-1}' \otimes z_{t-1}) \text{vec} \left( (\overline{\Delta \beta} - \Delta \beta)(\overline{\Delta \beta} - \Delta \beta)' \right) \rightarrow 0
\]

since

\[
\frac{1}{T} \sum_{t=1}^{T} (z_{t-1} \otimes z_{t-1}) = O_p(1)
\]

\(\overline{\Delta \beta} - \Delta \beta \rightarrow 0\)

because of assumption (i) in Theorem 1.3.2 and result (a) in Theorem 1.3.1.
For the third term,

$$\frac{1}{T} \sum_{t=1}^{T} \sigma_m^2 (\lambda_{t-1} - \hat{\lambda}_{t-1})^2 = \sigma_m^2 \frac{1}{T} \sum_{t=1}^{T} j_{t-1} (\hat{\Delta}\beta^* - \Delta\beta^*) (\hat{\Delta}\beta^* - \Delta\beta^*) j_{t-1}'$$

where $j_{t-1}$ is the $t$-th row of Jacobian matrix $J(\Delta\beta^*)$, and therefore,

$$\text{vec} \left( \frac{1}{T} \sum_{t=1}^{T} j_{t-1} (\hat{\Delta}\beta^* - \Delta\beta^*) (\hat{\Delta}\beta^* - \Delta\beta^*) j_{t-1}' \right) = \frac{1}{T} \sum_{t=1}^{T} j_{t-1} \otimes j_{t-1} \text{vec} \left( (\hat{\Delta}\beta^* - \Delta\beta^*) (\hat{\Delta}\beta^* - \Delta\beta^*) j_{t-1}' \right) \overset{p}{\rightarrow} 0,$$

given that

$$\frac{1}{T} \sum_{t=1}^{T} (j_{t-1} \otimes j_{t-1}) = O_p(1)$$

$$\hat{\Delta}\beta^* - \Delta\beta^* \overset{p}{\rightarrow} 0$$

$$\hat{\sigma}_m^2 \overset{p}{\rightarrow} \sigma_m^2$$

because of assumption (ii) in Theorem 1.3.2 and result (a) in Theorem 1.3.1.

The proofs for the rest terms are omitted here, since similar proof technique applies to the rest of the summation terms. Their convergence to zero in probability relies on assumptions in Theorem 1.3.2 and the results in Theorem 1.3.1.

Therefore, we prove (1.3.35),

$$\frac{1}{T} \sum_{t=1}^{T} \Delta v_t^2 \overset{p}{\rightarrow} \frac{1}{T} \sum_{t=1}^{T} \Delta v_t^2 \overset{p}{\rightarrow} \text{E}(\text{var}(\Delta v_t | y^{t-1})).$$
(ii) Proof of \((1.3.36)\).

Let \(\beta_l^c\) denote \(\beta_l(C_l)\) defined in \((1.3.33)\).

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{lt}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_{lt} - z_{t-1} \beta_l^c - C_l \hat{\lambda}_{t-1})^2 \\
= \frac{1}{T} \sum_{t=1}^{T} \left[ z_{t-1}(\beta_l - \beta_l^c) + C_l(\lambda_{t-1} - \hat{\lambda}_{t-1}) + v_{lt} \right]^2 \\
= \frac{1}{T} \sum_{t=1}^{T} v_{lt}^2 + \frac{1}{T} \sum_{t=1}^{T} z_{t-1}(\beta_l - \beta_l^c)(\beta_l - \beta_l^c)'z_{t-1}' + \frac{1}{T} \sum_{t=1}^{T} C_l^2(\lambda_{t-1} - \hat{\lambda}_{t-1})^2 \\
= \frac{1}{T} \sum_{t=1}^{T} z_{t-1}(\beta_l - \beta_l^c)v_{lt} + \frac{1}{T} \sum_{t=1}^{T} z_{t-1}(\beta_l - \beta_l^c)(\lambda_{t-1} - \hat{\lambda}_{t-1})C_l \\
+ \frac{1}{T} \sum_{t=1}^{T} C_l(\lambda_{t-1} - \hat{\lambda}_{t-1})v_{lt}.
\]

In the above expression, for the first term in the summation, we have

\[
\frac{1}{T} \sum_{t=1}^{T} v_{lt}^2 \overset{p}{\rightarrow} \frac{1}{T} \sum_{t=1}^{T} \text{var}(v_{lt}|y^{t-1}) \overset{p}{\rightarrow} E(\text{var}(v_{lt}|y^{t-1}))
\]

because of the Law of Large Numbers for mixing sequences and the ergodic theorem for \(\{Y_t\}\). The rest of the terms in the summation converges to zero in probability by similar arguments as those in the proof of \((1.3.35)\).

(iii) Proof of \((1.3.38)\).

Given \((1.3.25)\), \((1.3.35)\), \((1.3.36)\), and the continuous mapping theorem, \((1.3.38)\) holds.
Appendix D

Proof of Theorem 4.3.1

Proof of (1):

We conduct the Taylor expansion of feasible two step estimates $\hat{y}_{jt}(\hat{\eta})$ around the true value $\eta_0$,

$$\hat{y}_{jt}(\hat{\eta}) = \mu_t(\hat{\eta}) + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\sigma_t(\hat{\eta})$$

$$= \mu_t(\eta_0) + \hat{m}_{jt}(\eta_0, \eta_0)\sigma_t(\eta_0)$$

$$= \frac{\partial \mu_t}{\partial \eta}(\hat{\eta} - \eta_0) + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\frac{\partial \sigma_t}{\partial \eta}(\hat{\eta} - \eta_0)$$

$$= \sigma_t(\hat{\eta}) \left[ \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \eta} + \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \eta} \right] (\hat{\eta} - \eta_0), \quad (D.1)$$

given the differentiability of $\mu_t(\eta)$, $\sigma_t(\eta)$ and $m_{jt}(\eta; \eta)$ in assumptions (ii) – (iv) of Theorem 4.3.1. Note that the oracle estimator $\tilde{y}_{jt}(\eta_0) \equiv \mu_t(\eta_0) + \hat{m}_{jt}(\eta_0, \eta_0)\sigma_t(\eta_0)$.

Rearranging terms in (D.1), we have

$$\hat{y}_{jt}(\hat{\eta}) - \tilde{y}_{jt}(\eta_0) = \frac{\partial \mu_t}{\partial \eta}(\hat{\eta} - \eta_0) + \hat{m}_{jt}(\hat{\eta}; \hat{\eta})\frac{\partial \sigma_t}{\partial \eta}(\hat{\eta} - \eta_0)$$
\[
\sigma_t(\hat{\eta}) \left[ \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \eta} + \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \eta} \right] (\hat{\eta} - \eta_0) = \sigma_t(\hat{\eta}) \left[ 1 + \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \mu_t} \right] \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \mu_t} \sigma_t(\eta) \sigma_t(\eta_0) (\hat{\eta} - \eta_0) \\
+ \sigma_t(\hat{\eta}) \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \eta} (\hat{\eta} - \eta_0) = o_p(T^{-\alpha})
\] (D.2)

given the parametric convergence rate of \( \hat{\eta} \) in assumption (i) and the orders of magnitude for \( \mu_t(\eta), \sigma_t(\eta), \hat{m}_{jt}(\eta; \eta) \) and their derivatives with respect to \( \eta \) in a neighborhood of the true \( \eta_0 \) in assumptions (ii) – (iv). Such decomposition proves (1) in Theorem 4.3.1.

Proof of (2):

Note that the expectation of infeasible estimator \( \hat{y}_{jt}(\eta_0) \) given the true \( \eta_0 \) is denoted as

\[
\bar{y}_{jt} \equiv E(\hat{y}_{jt}(\eta_0)) = \mu_t(\eta_0) + m_{jt}(\eta_0; \eta_0) \sigma_t(\eta),
\] (D.3)

then subtracting (D.3) from (D.1), we have

\[
\hat{y}_{jt}(\hat{\eta}) - \bar{y}_{jt} = \left[ \frac{\hat{m}_{jt}(\eta_0; \eta_0) - m_{jt}(\eta_0; \eta_0)}{\partial \eta_0} \right] \sigma_t(\eta_0) + \sigma_t(\hat{\eta}) \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \eta} (\hat{\eta} - \eta_0) \\
+ \left[ 1 + \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \mu_t} \right] \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \mu_t} \sigma_t(\eta) \sigma_t(\eta_0) (\hat{\eta} - \eta_0) \\
+ \left[ \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \mu_t} \right] \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \mu_t} \sigma_t(\eta) \sigma_t(\eta_0) (\hat{\eta} - \eta_0) \\
+ \left[ \frac{\partial \hat{m}_{jt}(\eta; \hat{\eta})}{\partial \sigma_t} \right] \frac{\partial \hat{m}_{jt}(\hat{\eta}; \eta)}{\partial \sigma_t} \sigma_t(\eta) \sigma_t(\eta_0) (\hat{\eta} - \eta_0)
\] (part 3)

Note that part 1, the variance term for infeasible estimator, is of order \( O_p(T^{-\alpha}) \), and that part 2 and 3 is of order \( o_p(T^{-\alpha}) \) which are asymptotically negligible. Therefore, such decomposition shows that the feasible two step estimator has the same limiting distribution as the infeasible estimator.
Appendix E

Primitive Conditions for

Theorem 4.3.1

Conrad and Mammen (2009) consider the nonparametric regression on latent covariates that are generated by initial parametric and nonparametric estimators. Our model is different from theirs in two aspects. First, the estimation of our model consists of estimating two parts (4.2.10) and (4.2.11) sequentially by plugging the regressors generated in the first step into the nonparametric regression model in the second step. In the contrast, in Conrad and Mammen (2009) the latent covariates depends on the the model itself, therefore, their estimator is obtained iteratively. Second, the latent regressors in our model is generated parametrically, while their initial estimator of latent regressors has both parametric and nonparametric components. It turns out that the oracle property in Theorem 4.3.1 is a special case of Theorem 2 in Conrad and Mammen (2009), and therefore we can adopt the primitive conditions for the oracle property of kernel-based nonparametric estimators (local linear and Nadaraya-Watson) with minor changes to
accommodate our model specification:

(i) The underlying process \( h_t = (\mu_t, \sigma_t, n_t) \) is stationary and \( \beta \)-mixing with mixing coefficients \( \beta(j) \leq cv^j \) for constants \( c > 0 \) and \( 0 < v < 1 \). The density \( f_{ht} \) of \( h_t \) is Lipschitz continuous and bounded away from 0 on a compact set \( I \equiv I_\mu \times I_\sigma \times I_n \).

The joint density of \( h_t \) and \( h_{t+s} \) is bounded on \( I_1 \times I_2 \), uniformly in \( s \);

(ii) The estimator \( \hat{\eta}_s \) satisfies \( \| \hat{\eta} - \eta_0 \| = o_p(T^{-\delta_\eta}) \) for constants \( 0 < \delta_\eta < 1/2 \);

(iii) For \( \eta_k (k = 1, 2) \) with \( \| \eta_k - \eta_0 \| \leq T^{\delta_\eta} \), we assume that

\[
| \hat{h}_t(\eta_1) - \hat{h}_t(\eta_2) | \leq V_T \| \eta_1 - \eta_2 \| + R_T,
\]

where \( V_T \) is a random variable with \( V_T = O_p(T^{\rho_\eta}) \) and \( R_T = o_p(T^{-\rho_0}) \) with constants \( 0 \leq \rho_\eta < \delta_\eta - \gamma \) and \( 0 \leq \rho_0 < (1 + \gamma)/2 \), where \( T^{-\gamma} \) is the order of the bandwidth of the kernel smoothing;

(iv) For \( \epsilon > 0 \) it holds with a constant \( C > 0 \) that

\[
H(\epsilon, \| \cdot \|, \{ \eta : \| \eta - \eta_0 \| \leq T^{-\delta_\eta} \}) \leq C \epsilon^{-1/2} T^{-\rho_\eta/2},
\]

where, for a set \( A \), \( H(\epsilon, \| \cdot \|, A) = \log N(\epsilon, \| \cdot \|, A) \) is the entropy of \( A \), i.e., \( N(\epsilon, \| \cdot \|, A) \) is the number of balls with radius \( \epsilon \) that are necessary to cover \( A \);

(v) It holds that \( E[\exp(\rho |\varepsilon_{jt}|)|\mathcal{F}_{t-1}] < C \) almost surely for \( \rho > 0 \) small enough and \( h_t \in I_1 \) with a constant \( C < \infty \);

(vi) The kernel \( K \) has bounded support (\([-1, 1]\), say) and a continuous derivative. The bandwidth \( b \) is of order \( T^{-\gamma} \) for a constant \( \gamma \) with \( 0 < \gamma < 1/3 \);
(vii) There exists $\delta_1 > 0$ such that for $\theta_k$ ($k = 1, 2$) with $||\eta_1 - \eta_0|| < \delta_1$ it holds that the (multivariate) process $(h_t, \hat{h}_t(\eta_1), \hat{h}_t(\eta_2))$ is $\beta$-mixing with mixing coefficients $\beta(j) < cv^j$ for constants $c > 0, 0 < v < 1$;

(viii) We assume that

$$\frac{1}{T} \sum_{t=1}^{T} E \left[ (\hat{h}_t(\eta) - h_t)K_b(\hat{h}_t(\eta) - x) \right] = o(T^{-2\gamma})$$

uniformly for $(\eta, x) \in \mathcal{G}_T$, where $\mathcal{G}_T$ is the set of tuples $(\eta, x)$ with $||\eta - \eta_0|| \leq T^{-\delta_0}$ and $x \in I$;

(ix) (for Nadaraya-Watson smoothing) We assume that

$$\frac{1}{T} \sum_{t=1}^{T} E \left[ (\hat{h}_t(\eta) - x)K_b(\hat{h}_t(\eta) - x) - (h_t - x)K_b(h_t - x) \right] = o(T^{-2\gamma})$$

uniformly for $(\eta, x) \in \mathcal{G}_T$. 