Monte Carlo Examination of Static and Dynamic Student $t$ Regression Models

by

Remi Paczkowski

Dissertation submitted to the Faculty of
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Agricultural and Applied Economics

Approved:
Dr. Anya McGuirk, Chair
Dr. Paul Hoepner
Dr. Daniel Taylor
Dr. Christine Anderson-Cook
Dr. Lester Myers, Department Head
(in lieu of Dr. Paul Driscoll)

Key words: Static Student $t$ Regression Model, Dynamic Student $t$ Regression Model, Student $t$ Autoregressive Model, Monte Carlo experiment, Maximum Likelihood estimation.

December, 1997
Blacksburg, Virginia

Copyright 1997, Remi Paczkowski
Monte Carlo Examination of Static and Dynamic Student $t$ Regression Models

by

Remi Paczkowski

(ABSTRACT)

This dissertation examines a number of issues related to Static and Dynamic Student $t$ Regression Models.

The Static Student $t$ Regression Model is derived and transformed to an operational form. The operational form is then examined in a series of Monte Carlo experiments. The model is judged based on its usefulness for estimation and testing and its ability to model the heteroskedastic conditional variance. It is also compared with the traditional Normal Linear Regression Model.

Subsequently the analysis is broadened to a dynamic setup. The Student $t$ Autoregressive Model is derived and a number of its operational forms are considered. Three forms are selected for a detailed examination in a series of Monte Carlo experiments. The models’ usefulness for estimation and testing is evaluated, as well as their ability to model the conditional variance. The models are also compared with the traditional Dynamic Linear Regression Model.
Acknowledgments

I would like to take this opportunity to express my deepest gratitude to my major professor Anya McGuirk, a wonderful human being, whose good nature and patience were tested many times during this project. I would like to thank her especially for the invaluable guidance in my research, the countless hours spent discussing the subject, and for clarifying many of my misconceptions. Finally, I am forever indebted to her for the encouragement and support she provided when I sometimes lacked faith.

I would like to thank other members of my committee: Paul Hoepner, Dan Taylor, Christine Anderson-Cook, and Paul Driscoll for their comments on an earlier draft. Their valuable feedback improved significantly the quality of the presentation.

I would like to express my deep gratitude to professor Aris Spanos, who, although not on my committee, always took time to meet with me. Without many helpful suggestions provided by professor Spanos this project could not have been completed.

My sincere thanks go to fellow graduate students. In particular I would like to thank Peter Ammermann, from the discussions with whom I greatly benefited. I would also like to thank Joe Schleich, for allowing me to use his office to do the editorial work. Finally, I would like to gratefully acknowledge the friendship of Eduard Zaloshnja, Bob Dawson, Godfrey Gibbison, Sonali Mitra, Zana Kruja, Xiaofeng Zhou, Nihal Attapattu and Line Carpentier. My years at Tech wouldn’t be the same without you.

Last, but not least, I would like to thank my roommates: Tony Dziepak, Paul Smith, and Sergio Harding for their invaluable help in navigating the treacherous field of English grammar.
# Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgments ............................................................................................................... iii

Table of Contents ................................................................................................................ iv

List of Tables ....................................................................................................................... vii

List of Figures ..................................................................................................................... ix

Chapter 1: Introduction ....................................................................................................... 1

1.1 Problem Statement, Objectives, and Methods ......................................................... 4
1.2 Justification of the Study ........................................................................................ 6
1.3 Literature of the Subject ....................................................................................... 6
1.4 Overview of the Dissertation ............................................................................... 7

Chapter 2: The Static Student $t$ Regression Model ....................................................... 8

2.1 Derivation of the Static Student $t$ Regression Model ............................................ 9
2.2 Operational Form of the Model ........................................................................... 14
2.3 An Interesting Extension - the Static Student $t$ Regression Model with Dynamic Heteroskedasticity ........................................................................ 16

Chapter 3: Monte Carlo Examination of the Static Student $t$ Regression Model ........... 19

3.1 Design of the Monte Carlo Experiment .................................................................. 19
3.2 Data Generation ...................................................................................................... 23
3.3 Maximum Likelihood Estimation .......................................................................... 24
3.4 Results ....................................................................................................................... 25
  3.4.1 Estimates of $\beta_1$ and $\beta_2$ ........................................................................... 25
  3.4.2 Estimates of $\beta_0$ .......................................................................................... 37
Chapter 3: Estimation of Variance Components

3.4.3 Estimates of $\sigma^2$ ................................................................. 39
3.4.4 Estimates of $\Sigma_{11}^{-1}$ ............................................................ 43
3.4.5 The Product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ .......................................... 46
3.4.6 Modeling the Conditional Variance .......................................... 50

3.5 Conclusions .................................................................................. 51

Chapter 4: The Dynamic Student $t$ Regression Model ........................................... 53

4.1 Derivation of the Student $t$ Autoregressive Model ....................................... 53
4.2 Operational Form of the Model .................................................................. 58
4.3 Alternative Forms of the Log-Likelihood Function ........................................ 68
  4.3.1 Direct Estimation of $\beta$ ............................................................ 68
  4.3.2 Direct Estimation of $(\Sigma_t)^{-1}$ .................................................. 71
  4.3.3 The Hybrid Approach ................................................................... 74
4.4 The Case of the Markov Variance ................................................................ 75
4.5 The Adapted Static Model ........................................................................ 77

Chapter 5: Monte Carlo Examination of Dynamic Student $t$ Regression Models ...... 78

5.1 Design of the Monte Carlo Experiment ..................................................... 78
5.2 Data Generation ................................................................................... 82
5.3 Results ............................................................................................... 83
  5.3.1 Estimates of $\beta_1$ and $\beta_2$ ...................................................... 83
  5.3.2 Estimates of $\beta_0$ ...................................................................... 95
  5.3.3 Estimates of $\sigma^2$ ................................................................. 100
  5.3.4 Estimates of $\Sigma_{22}^{-1}$ and $(\Sigma_t^1)^{-1}$ .................................... 105
  5.3.5 The Product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ and $\hat{\sigma}^2 (\hat{\Sigma}_t^1)^{-1}$ ............. 112
  5.3.6 Modeling the Conditional Variance .............................................. 119
5.4 Conclusions ....................................................................................... 120

Chapter 6: Conclusions ............................................................................. 122
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>125</td>
</tr>
<tr>
<td>Appendix 1: Derivation of the Conditional Student ( t ) Density</td>
<td>128</td>
</tr>
<tr>
<td>Appendix 2: Delta Method</td>
<td>131</td>
</tr>
<tr>
<td>Appendix 3: Graphs of the Empirical Distributions of Estimates from the Static Model</td>
<td>133</td>
</tr>
<tr>
<td>A3.1 Graphs of Empirical Distributions for ( \nu=4 )</td>
<td>133</td>
</tr>
<tr>
<td>A3.2 Graphs of Empirical Distributions for ( \nu=8 )</td>
<td>139</td>
</tr>
<tr>
<td>Appendix 4: Graphs of the Empirical Distributions of Estimates from the Dynamic Models</td>
<td>145</td>
</tr>
<tr>
<td>A4.1 Graphs of Empirical Distributions for ( \nu=4 )</td>
<td>145</td>
</tr>
<tr>
<td>A4.2 Graphs of Empirical Distributions for ( \nu=8 )</td>
<td>158</td>
</tr>
<tr>
<td>Appendix 5: Descriptive Statistics for the Empirical Distributions of the Estimated Standard Errors</td>
<td>171</td>
</tr>
<tr>
<td>Vita</td>
<td>174</td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>Selected descriptive statistics of the empirical distribution of ML estimates of $\beta_1$ and $\beta_2$</td>
<td>26</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Selected descriptive statistics of the empirical distribution of OLS estimates of $\beta_1$ and $\beta_2$</td>
<td>29</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Selected descriptive statistics of the empirical distribution of ML and OLS estimates of $\beta_0$</td>
<td>39</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Selected descriptive statistics of the empirical distributions of $\sigma^2$ and $\sigma$ estimated by ML</td>
<td>43</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Selected descriptive statistics of the empirical distributions of $\sigma^2$ estimated by OLS</td>
<td>43</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Selected descriptive statistics of the empirical distributions of $(\hat{\Sigma}^{-1}<em>{22})</em>{11}$ and $(\hat{\Sigma}^{-1}<em>{22})</em>{31}$</td>
<td>46</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Selected descriptive statistics of the empirical distributions of $\hat{\sigma}^2(\hat{\Sigma}^{-1}<em>{22})</em>{11}$ and $\hat{\sigma}^2(\hat{\Sigma}^{-1}<em>{22})</em>{31}$</td>
<td>48</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Selected descriptive statistics of the empirical distributions of ML estimates of $\beta_1$ and $\beta_2$</td>
<td>88</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Selected descriptive statistics of the empirical distributions of OLS estimates of $\beta_1$ and $\beta_2$</td>
<td>89</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Selected descriptive statistics of the empirical distribution of ML estimates of $\beta_0$</td>
<td>99</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Selected descriptive statistics of the empirical distribution of OLS estimates of $\beta_0$</td>
<td>99</td>
</tr>
</tbody>
</table>
Table 5.5. Selected descriptive statistics of the empirical distributions of ML and OLS estimates of $\sigma^2$........................................................................................................... 104

Table 5.6. Selected descriptive statistics of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$ and $(\hat{\Sigma}_{22}^{-1})_{2,3}$ (ML-A)........................................................................................................... 106

Table 5.7. Selected descriptive statistics of the empirical distribution of $(\hat{\Sigma}_{1}^{-1})_{i,i}$ and $(\hat{\Sigma}_{1}^{-1})_{i,i-2}$ (ML-B, ML-S).............................................................. 109

Table 5.8. Selected descriptive statistics of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}_{22})_{2,2}$ and $\hat{\sigma}^2(\hat{\Sigma}_{22})_{2,3}$ (ML-A).................................................................................................... 113

Table 5.9. Selected descriptive statistics of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}_{1})_{i,i}^{-1}$ and $\hat{\sigma}^2(\hat{\Sigma}_{1})_{i,i-2}^{-1}$ (ML-B, ML-S).................................................................................. 116

Table A5.1. Selected descriptive statistics for the estimated standard errors of $\beta_1$ and $\beta_2$
- Maximum Likelihood.......................................................................................... 172

Table A5.2. Selected descriptive statistics for the estimated standard errors of $\beta_1$ and $\beta_2$
- OLS and HCSE................................................................................................. 173
List of Figures

Figure 1.1. t-plots and kernel density estimates .................................................... 3
Figure 1.2. Kernel density estimates and t-distribution ............................................. 4
Figure 3.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=6$ .... 27
Figure 3.2. Empirical distribution of ML and OLS standard errors, $\nu=6$ ................. 30
Figure 3.3. Empirical distributions of the ML and OLS t-statistics, $\nu=6$ ................. 33
Figure 3.4. Empirical power curves for tests of hypotheses about $\beta_1$ at 5% significance level, $\nu=6$ ................................................................. 35
Figure 3.5. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$ .... 38
Figure 3.6. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=6$ ...... 42
Figure 3.7. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{11}$ and $(\hat{\Sigma}_{22}^{-1})_{21}$ (ML). ................................................................. 44
Figure 3.8. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{11}$ and $\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{21}$, $\nu=6$ ................................................................. 49
Figure 3.9. Actual and predicted conditional variance from SSTM and NLRM .... 51
Figure 5.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=6$ ....... 86
Figure 5.2. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=6$ ....... 87
Figure 5.3. Kernel density estimates of the empirical distribution of standard errors for $\hat{\beta}_1$, $\nu=6$ ................................................................. 90
Figure 5.4. 100 estimates of $\beta_1$, $\nu=6$ .................................................................................. 92
Figure 5.5. Empirical power curves for tests of hypotheses about $\beta_1$ at 5% significance level, $\nu=6$ .................................................................................. 94
Figure 5.6. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$ ........ 97
Figure 5.7. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$ ........ 98
Figure 5.8. Kernel density estimates of the empirical distribution of $\sigma^2$, $\nu=6$........... 102
Figure 5.9. Kernel density estimates of the empirical distribution of $\sigma^2$, $\nu=6$........... 103
Figure 5.10. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{12})_{2,2}$ and $(\hat{\Sigma}^{-1}_{22})_{2,1}$, $\nu=6$ .................................................................................. 107
Figure 5.11. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{i,i})^{i}$ and $(\hat{\Sigma}^{-1}_{i,i})^{i-1}$, $\nu=6$ .................................................................................. 110
Figure 5.12. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{i,i})^{i}$ and $(\hat{\Sigma}^{-1}_{i,i})^{i-1}$, $\nu=6$ .................................................................................. 111
Figure 5.13. Kernel density estimates of the empirical distribution of $\sigma^2(\hat{\Sigma}^{-1}_{12})_{2,2}$ and $\sigma^2(\hat{\Sigma}^{-1}_{22})_{2,1}$ (ML-A), $\nu=6$ .................................................................................. 114
Figure 5.14. Kernel density estimates of the empirical distribution of $\sigma^2(\hat{\Sigma}^{-1}_{i,i})^{i}$ and $\sigma^2(\hat{\Sigma}^{-1}_{i,i})^{i-1}$ (ML-B), $\nu=6$ .................................................................................. 117
Figure 5.15. Kernel density estimates of the empirical distribution of $\sigma^2(\hat{\Sigma}^{-1}_{i,i})^{i}$ and $\sigma^2(\hat{\Sigma}^{-1}_{i,i})^{i-1}$ (ML-S), $\nu=6$ .................................................................................. 118
Figure 5.16. Actual and predicted conditional variance ........................................... 119

Figure 5.17. Window OLS and ASSTM estimates of the conditional variance ...... 120

Figure A3.1.1. Kernel density estimates of the empirical distribution of $\beta_1$, $\nu$=4 .... 134

Figure A3.1.2. Kernel density estimates of the empirical distribution of $\beta_1$, $\nu$=4 ...... 135

Figure A3.1.3. Kernel density estimates of the empirical distribution of $\sigma^2$, $\nu$=4....... 136

Figure A3.1.4. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1})_{i1}$ and
$(\hat{\Sigma}^{-1})_{21}$, $\nu$=4 ..................................................................................... 137

Figure A3.1.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}^{-1})_{i1}$
and $\hat{\sigma}^2(\hat{\Sigma}^{-1})_{21}$, $\nu$=4 ..................................................................................... 138

Figure A3.2.1. Kernel density estimates of the empirical distribution of $\beta_1$, $\nu$=8 ...... 140

Figure A3.2.2. Kernel density estimates of the empirical distribution of $\beta_0$, $\nu$=8....... 141

Figure A3.2.3. Kernel density estimates of the empirical distribution of $\sigma^2$, $\nu$=8....... 142

Figure A3.2.4. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1})_{i1}$ and
$(\hat{\Sigma}^{-1})_{21}$ (ML), $\nu$=8 ..................................................................................... 143

Figure A3.2.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}^{-1})_{i1}$
and $\hat{\sigma}^2(\hat{\Sigma}^{-1})_{21}$ (ML), $\nu$=8 ..................................................................................... 144

Figure A4.1.1. Kernel density estimates of the empirical distribution of $\beta_1$, $\nu$=4 ...... 146

Figure A4.1.2. Kernel density estimates of the empirical distribution of $\beta_0$, $\nu$=4 ...... 147

Figure A4.1.3. Kernel density estimates of the empirical distribution of $\beta_0$, $\nu$=4...... 148
Figure A4.1.4.  Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $v=4$ ....... 149
Figure A4.1.5.  Kernel density estimates of the empirical distribution of $\hat{\alpha}^2$, $v=4$ ....... 150
Figure A4.1.6.  Kernel density estimates of the empirical distribution of $\hat{\alpha}^2$, $v=4$ ....... 151
Figure A4.1.7.  Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$ and  
$(\hat{\Sigma}_{22}^{-1})_{2,1}$, $v=4$ ........................................................................................................ 152
Figure A4.1.8.  Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_{ii}^{-1})_{i,i}$ and  
$(\hat{\Sigma}_{ii}^{-1})_{i,i-2}$, $v=4$ ........................................................................................................ 153
Figure A4.1.9.  Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_{ii}^{-1})_{i,i}$ and  
$(\hat{\Sigma}_{ii}^{-1})_{i,i-2}$, $v=4$ ........................................................................................................ 154
Figure A4.1.10. Kernel density estimates of the empirical distribution of $\hat{\alpha}^2(\hat{\Sigma}_{22}^{-1})_{2,2}$ 
and $\hat{\alpha}^2(\hat{\Sigma}_{22}^{-1})_{2,1}$, $v=4$ ........................................................................................................ 155
Figure A4.1.11. Kernel density estimates of the empirical distribution of $\hat{\alpha}^2(\hat{\Sigma}_{ii}^{-1})_{i,i}$ 
and $\hat{\alpha}^2(\hat{\Sigma}_{ii}^{-1})_{i,i-2}$, $v=4$ ........................................................................................................ 156
Figure A4.1.12. Kernel density estimates of the empirical distribution of $\hat{\alpha}^2(\hat{\Sigma}_{ii}^{-1})_{i,i}$ and  
$\hat{\alpha}^2(\hat{\Sigma}_{ii}^{-1})_{i,i-2}$, $v=4$ ........................................................................................................ 157
Figure A4.2.1.  Kernel density estimates of the empirical distribution of $\hat{\beta}_i$, $v=8$ ....... 159
Figure A4.2.2.  Kernel density estimates of the empirical distribution of $\hat{\beta}_i$, $v=8$ ....... 160
Figure A4.2.3.  Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $v=8$ ....... 161
Figure A4.2.4.  Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $v=8$ ....... 162
Figure A4.2.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=8$ ........ 163
Figure A4.2.6. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=8$ ........ 164
Figure A4.2.7. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{22})_{2,2}$ and
$(\hat{\Sigma}^{-1}_{22})_{i,i}, \nu=8$ ................................................................. 165
Figure A4.2.8. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{\tau})_{i,i}$ and
$(\hat{\Sigma}^{-1}_{\tau})_{i,i-2}, \nu=8$ ........................................................................... 166
Figure A4.2.9. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}^{-1}_{\tau})_{i,i}$ and
$(\hat{\Sigma}^{-1}_{\tau})_{i,i-2}, \nu=8$ ........................................................................... 167
Figure A4.2.10. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}^{-1}_{22})_{2,2}$
and $\hat{\sigma}^2(\hat{\Sigma}^{-1}_{22})_{i,i}, \nu=8$ ................................................................. 168
Figure A4.2.11. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}^{-1}_{\tau})_{i,i}$ and
$\hat{\sigma}^2(\hat{\Sigma}^{-1}_{\tau})_{i,i-2}, \nu=8$ ........................................................................... 169
Figure A4.2.12. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}^{-1}_{\tau})_{i,i}$
and $\hat{\sigma}^2(\hat{\Sigma}^{-1}_{\tau})_{i,i-2}, \nu=8$ ........................................................................... 170
Chapter 1

Introduction

It is a generally accepted fact, that many real-life data series exhibit characteristics of non-normal distributions. In particular, it has been known since the early 1960s, that financial data\(^1\) tend to exhibit a number of characteristics inconsistent with normality.\(^2\) An example supporting this claim is presented in Figure 1.1. The graphs included there show daily changes in corn futures closing prices, and weekly log differences of the exchange rate between the Canadian dollar (CND) and the US dollar (USD). The graph of an independent and identically distributed normal series with mean 0 and variance 1 (NIID(0,1)) is included for comparison.

Both real-life data series have a constant mean, but their variance exhibits a number of differences when compared with the NIID series. The observations tend to be more concentrated around the mean, than the NIID data, and the outliers are relatively bigger. Additionally, patterns in the conditional variance of the real-life data are evident. Big changes tend to be followed by big changes, and small changes are followed by small changes, creating a clustering effect. This particular characteristic of the real-life data contrasts strongly with the NIID series, whose variation exhibits no discernible pattern. A look at the kernel density graphs, included in Figure 1.1, confirms the above observations.\(^3\) The distribution of real-life data is more peaked than the normal, reflecting a higher concentration of observations around the mean.\(^4\) It also exhibits fat tails reflecting relatively bigger outliers. In fact, kernel density graphs shown in Figure 1.1 bear striking resemblance to the \(t\)-distribution. This last observation is confirmed by the graphs in Figure 1.2. These graphs depict the same kernel density estimates, that were shown in Figure 1.1. Superimposed are the contours of the \(t\)-distribution with the same mean and variance as the empirical data. The degrees of freedom parameter of the superimposed \(t\)-distribution was set

---

\(^1\) In this dissertation, the term “financial data” is understood broadly. It includes prices of securities, interest rates, exchange rates, prices of futures contracts, etc.

\(^2\) See Mandelbrot (1963).

\(^3\) Normal kernel density with bandwidth equal to \(1.06\sigma^{0.2}\), where \(\sigma\) is the standard deviation of the data and \(t\) is the number of observations. See Silverman (1986).

\(^4\) The dashed line in the graphs represents the contour of the normal distribution with the same mean and variance as the data being shown in the graph.
equal to 3.4, in the graph showing estimated kernel density for the daily changes in closing prices of corn futures. In the graph showing kernel density for weekly log-differences in CND/USD exchange rate, this parameter was set equal to 5. In both cases the $t$-distribution provides an almost perfect approximation of the actual distribution of the data. This last fact is consistent with a number of studies of financial data, where tests favored the Student $t$ distribution over the normal (see Spanos (1993)).
Figure 1. 1. t-plots and kernel density estimates.

The dashed line represents the contour of the normal distribution with the same mean and variance as the data whose kernel density is shown in the graph.
1.1. Problem Statement, Objectives, and Methods.

The above examples suggest, that, at least for some financial data, assuming normality may be inappropriate. If normality is assumed, but the data are actually distributed $t$, the following assumptions underlying both the Normal Linear Regression Model (NLRM) and the Dynamic Linear Regression Model (DLRM) are violated:  

- weak (and strong) exogeneity of the conditioning variables;
- normality of the joint distribution;
- homoskedasticity of the conditional variance.

In this context two issues deserve special consideration:

- the consequences of fitting models which assume normality to Student $t$ data;
- the possibility of improvement over the regular OLS, based on utilizing all the information contained in the data.

The consequences of dealing with the non-normal/heteroskedastic data have been explored already in the econometric literature (see Spanos (1986) for example). The most obvious consequence is that the “traditional” Maximum Likelihood Method (i.e. the one assuming normality) is no longer valid. In this case the researcher can still resort to the Least

---

6 See Spanos (1986) for a comprehensive discussion of the full set of assumptions underlying both models.
Squares Method, which does not require any distributional assumptions. If the data are distributed $t$, the OLS estimator of regression coefficients remains unbiased and consistent.$^7$ However, due to the presence of heteroskedasticity, the OLS estimator of the variance-covariance matrix of regression coefficients becomes inconsistent.$^8$ Because of that, no meaningful testing is possible on the results of OLS. A solution to this problem was proposed by White (1980). In this influential paper he argued, that, although there is no consistent estimator of the conditional variance, a consistent estimator of the variance-covariance matrix of regression coefficients is available. The formula proposed by White gives rise to the Heteroskedasticity-Corrected Standard Errors (HCSEs). Testing is still asymptotically justified, if the regular OLS standard errors are replaced with HCSEs.

Following up on the possibility, that exploiting all the information contained in the data can result in an improvement over the OLS, Spanos (1990) and Spanos (1994) proposed a family of static and dynamic Student $t$ regression models. This family has the potential to outperform the traditional normal/homoskedastic formulations, when the distribution of the data exhibits bell-shaped symmetry and leptokurticity. The biggest advantage offered by the Student $t$ models comes in the ability to model the conditional variance in a statistically coherent fashion. Under the Student $t$ approach the conditional mean and variance are modeled jointly, which opens up the possibility to outperform not only the traditional OLS but the ARCH/GARCH formulations as well.$^9$

Despite the fact that some of the models proposed by Spanos have been successfully applied to modeling real-life phenomena (see Robertson and Spanos (1991), Spanos (1993), and McGuirk, Robertson and Spanos (1993)), a comprehensive evaluation of the performance of different Student $t$ regression models has not yet been conducted. Of particular interest is the examination the following issues:

1. Usefulness of the Student $t$ models and the implied Maximum Likelihood Estimators (MLEs) for estimation and testing.

$^9$ For details regarding ARCH and GARCH see Engle (1982) and Bollerslev (1986) respectively.
2. The ability of the Student $t$ models to model the conditional variance accurately.
3. The comparison of the Student $t$ models with the traditional NLRM and DLRM.

An attempt to explore the above issues through a series of Monte Carlo experiments will be undertaken in this dissertation.

1.2. Justification of the Study.

The importance of evaluation of performance of the Student $t$ regression models can be fully appreciated when their potential applications are considered. One area where the Student $t$ models show great promise seems to be the analysis of speculative prices. The examples presented previously, showed, that, at least in some instances, the $t$ distribution provides an almost perfect approximation of the distribution of financial data. In this case, the big payoff from using the Student $t$ model comes in the ability to model the conditional variance. This is important, because a large part of financial analysis is concerned with managing risk, which is measured by the variance of the prices of securities.\textsuperscript{10} Thus, a model capable of capturing and predicting the changes in the conditional variance is an extremely valuable tool for both the financial researchers and the traders.

1.3. Literature of the Subject.

The approach adopted in this dissertation is that of probabilistic reduction, originally proposed by Spanos (1986). In this approach operational models are derived from the joint distribution of all the observables, by imposing reduction assumptions falling into three categories:\textsuperscript{11}

- distributional assumptions;
- memory assumptions;
- homogeneity assumptions.

\textsuperscript{10} For details see any handbook dealing with portfolio analysis, for example Elton and Gruber (1991).
\textsuperscript{11} For an extensive discussion of the probabilistic reduction approach, as well as the underlying assumptions, see Spanos (1986).
This assures that all aspects of the model are treated in a statistically coherent fashion. In particular, the marginal and conditional distributions retain the connection to the underlying joint distribution.

The probabilistic reduction approach remains vastly underutilized in the context of non-normal distributions. Except for the precious few works of Spanos, McGuirk, and Robertson virtually no research has been done in this area.

The theoretical foundations of the Student $t$ models, as defined in this dissertation, were laid in Spanos (1990) and Spanos (1994). The derivations in these path-breaking papers are similar to the derivations in Chapter 2 and Chapter 4. Because of that, no formal review will be presented, and the two papers will be quoted as needed. For a review of the statistical literature pertaining to the class of elliptically-contoured distributions the reader is referred to Chmielewski (1981).

One of the models proposed by Spanos, the Student $t$ Autoregressive model (STAR), was applied in modeling the dynamics of a number of exchange rates (Spanos (1993), and McGuirk, Robertson and Spanos (1993)) and the volatility of the US interest rates (Robertson (1992)). In these studies a comprehensive battery of misspecification tests indicated that the STAR model was able to account for the probabilistic features of the data. Additionally, the STAR model was found to dominate rival ARCH/GARCH formulations on misspecification grounds.


The remaining part of the dissertation is organized as follows. In Chapter 2 the Static Student $t$ Regression Model is derived. It is then transformed to an operational form, which is later used in a series of Monte Carlo experiments. The results of these experiments and their analysis are contained in Chapter 3. In Chapter 4 the Dynamic Student $t$ Regression Model is derived and a number of its operational forms are considered. Three models are selected for a detailed examination in a series of Monte Carlo experiments. The results of these experiments and their analysis are presented in Chapter 5. Finally, Chapter 6 contains the conclusions drawn from the examination of selected models.
Chapter 2

The Static Student $t$ Regression Model

The Static Student $t$ Regression Model (SSTRM) was first proposed by Spanos (1994) as an extension of the Normal Linear Regression Model (NLRM). The non-normal models arise naturally within the framework of the probabilistic reduction approach (see Spanos (1986)), when the normal density is replaced by some other form of density. The obvious starting point for any extensions of the NLRM is the elliptical family, which is the parent family of the normal distribution. Just like the normal, other members of the elliptical family have linear conditional means, but their conditional variances are heteroskedastic (when they exist). Therefore, they lend themselves naturally to modeling the phenomena characterized by linear conditional means and heteroskedastic conditional variances.

Within the probabilistic reduction framework, the choice of a particular model is governed by the probabilistic features of the phenomenon being modeled. In this context, examination of the graphs of raw data, as well as kernel density estimates of the marginal and bivariate distributions, is of paramount importance. Visual inspection of the data can provide valuable insight enabling the modeler to make educated guesses regarding the nature of the relationship and the correct form of the distribution. When the data exhibit constant mean, no memory, and bell-shaped symmetry combined with such features as constant variance, 95% of observations located between $\pm 2$ standard deviations, and no discernible pattern of variation, the NLRM is likely a good choice. On the other hand, if the bell-shaped symmetry is coupled with fat tails and excessive peakedness around the mean, the Student $t$ model makes a good candidate. The choice, however, is always tentative, and should be verified through a battery of misspecification tests (see McGuirk, Driscoll and Alwang (1993) and Robertson (1992)).

---

1 See Kelker (1970).
2 This approach, originally due to Spanos (1986), is gaining wider acceptance. For example see Goldberger (1991).
2.1. **Derivation of the Static Student \( t \) Regression Model.**

Let \( \{Z_t, t = 1, \ldots, T\} \), where \( Z_t = (Y_t, X_t)' \), \( Y_t: 1 \times 1, X_t: k \times 1 \), be independent and identically distributed (IID) according to

\[
Z_t \sim St_{k+1}(\mu, \Sigma, \nu),
\]

where \( St \) denotes the Student \( t \) distribution, \( \mu \) is the vector of means, \( \Sigma \) is the scale matrix, and \( \nu \) denotes the degrees of freedom.

The joint distribution of \( Z_T \), which is poly-\( t \), can be reduced as follows

\[
D(Z_T; \Phi) = \prod_{t=1}^{T} D(Z_t; \Psi_t) = \prod_{t=1}^{T} D(Z_t; \Psi) = \prod_{t=1}^{T} D(Y_t|X_t; \Psi_1)D(X_t; \Psi_2),
\]

where \( \Phi \) and \( \Psi \) denote the set of parameters characterizing each distribution. The first equality in the above decomposition follows from independence, the second equality is the consequence of “identically distributed” assumption, and the last one uses sequential conditioning (Bayes theorem). When dealing with normal IID models, \( D(X_t; \Psi_2) \) can be dropped because of weak exogeneity. In the present case, however, the weak exogeneity no longer holds, and ignoring the “marginal” part results in a significant loss of information.

The densities \( D(X_t) \) and \( D(Y_t|X_t) \) take the form

---

3 Scale matrix is related to the variance-covariance matrix through \( V = \frac{\nu}{\nu - 2} \Sigma \), where \( V \) is the variance-covariance matrix.

4 \( Z_T = vec[Z_1 \quad Z_2 \quad \ldots \quad Z_T] \). The subscript/superscript convention indicates both the range and the order of elements being stacked in a column. For an explanation of the vec operator see footnote 16.

5 See Drèze (1977) and references therein.


7 See Spanos (1994).

8 See Appendix 1 for derivation of the conditional density. The formulas 3.4.39 and 3.4.42 in Poirier (1995) p. 127 are wrong.
\[
D(X_i) = \frac{\Gamma(\frac{v + k}{2})}{(\pi v)^{\frac{k}{2}} \Gamma(\frac{v}{2})} |\Sigma_{22}|^{\frac{1}{2}} \left[ 1 + \frac{1}{v} (X_i - \mu_2)' \Sigma_{22}^{-1} (X_i - \mu_2) \right]^{\frac{v + k}{2}},
\]
(2.2)

\[
D(Y_i|X_i) = \frac{\Gamma(\frac{v + k + 1}{2})}{(\pi v)^{\frac{k}{2}} \Gamma(\frac{v}{2})} \left| \sigma^2 \left[ 1 + \frac{1}{v} (X_i - \mu_2)' \Sigma_{22}^{-1} (X_i - \mu_2) \right] \right|^{\frac{1}{2}} \times
\]
\[
\times \left[ 1 + \frac{1}{v} \left( Y_i - \beta_0 - \beta' X_i \right) \right] \left\{ \sigma^2 \left[ 1 + \frac{1}{v} (X_i - \mu_2)' \Sigma_{22}^{-1} (X_i - \mu_2) \right] \right\}^{-1} \left( Y_i - \beta_0 - \beta' X_i \right) \right|^{\frac{v + k + 1}{2}},
\]
(2.3)

where

\[\Sigma\] has been partitioned into \[\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}\], such that \[\frac{v}{v - 2} \sigma_{11}\] is the variance of \(Y\),

\[\frac{v}{v - 2} \sigma_{21}\] is the vector of covariances between \(Y\) and the \(X\)'s, and \[\frac{v}{v - 2} \Sigma_{22}\] is the variance-covariance matrix of the \(X\)'s;

\[\sigma^2 = \sigma_{11} - \sigma_{12} \Sigma_{22}^{-1} \sigma_{21} ;\]

\[\mu_2 = \text{E}(X_i) ;\]

\[\beta = \Sigma_{22}^{-1} \sigma_{21} - \text{vector of regression coefficients, which does NOT include a constant};\]

\[\beta_0 - \text{constant term}.\]

The parameters of interest are \(\beta, \mu_2, \sigma^2, \) and \(\Sigma_{22}^{-1}\).

The mean of the conditional distribution in (2.3) is given by\(^{10}\)

\[\text{E}(Y_i|X_i) = \beta_0 + \beta' X_i ,\]

\(^9\) \(\beta_0 = \mu_1 - \beta' \mu_2\) and the sample mean of \(Y\) can be used as an estimate of \(\mu_1\) in this calculation.

\(^{10}\) See Zellner (1971).
and the conditional variance takes the form\(^{11}\)

\[
\text{var}(Y_t | X_t) = \frac{v}{v + k - 2} \sigma^2 \left[ 1 + \frac{1}{v} (X_t - \mu_2)' \Sigma^{-1}_{22} (X_t - \mu_2) \right].
\]

It should be observed that the conditional mean takes the same form as in the NLRM, but the conditional variance is heteroskedastic.

Ordinarily, equations (2.2) and (2.3) would be substituted into (2.1) producing the likelihood function of the SSTRM. The logarithmic form of the likelihood function would then be differentiated, and the first order conditions solved, to obtain the estimators of the parameters of interest.\(^{12}\) This approach, although feasible, will not be followed here. Instead, an easier formulation will be developed, starting from a reparametrization of the joint density.

Consider again the reduction in (2.1), this time taking only the first two steps

\[
D(Z_t^1; \Phi) = \prod_{i=1}^{T} D(Z_i; \Psi_i) = \prod_{t=1}^{T} D(Z_t; \Psi).
\]

(2.4)

Since \(Z_t\) is distributed as multivariate-\(t\), its density takes the form\(^{13}\)

\[
D(Z_t) = \frac{\Gamma \left( \frac{v + k + 1}{2} \right)}{\sqrt{(\pi v) \Gamma \left( \frac{v}{2} \right)}} \left| \Sigma \right|^{-\frac{1}{2}} \left[ 1 + \frac{1}{v} (Z_t - \mu)' \Sigma^{-1} (Z_t - \mu) \right]^{-\frac{v + k + 1}{2}}.
\]

This formulation does not involve all of the parameters of interest. However, using the results stated in Searle (1982) p.258 and p.260, the above equation can be rewritten as

\[
D(Z_t) = \frac{\Gamma \left( \frac{v + k + 1}{2} \right)}{\sqrt{(\pi v) \Gamma \left( \frac{v}{2} \right)}} \left( \sigma^2 \right)^{-\frac{1}{2}} \left| \Sigma_{22} \right|^{-\frac{1}{2}} \left[ 1 + \frac{1}{v} (X_t - \mu_2)' \Sigma^{-1}_{22} (X_t - \mu_2) + \frac{1}{v \sigma^2} (Y_t - \beta_0 - \beta'X_t)^2 \right]^{-\frac{v + k + 1}{2}},
\]

\(^{11}\) See Zellner (1971).

\(^{12}\) First order conditions turn out to be non-linear and have to be solved iteratively using a numerical procedure.

\(^{13}\) See Johnson and Kotz (1972).
where the notational convention is the same as in (2.2) - (2.3), and all the parameters of interest appear explicitly. Substituting this into (2.4) produces the likelihood function (LF) for the Static Student $t$ Regression Model

$$LF = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{v+k+1}{2}\right)}{\left(\pi v\right)^{\frac{T}{2}} \Gamma\left(\frac{v}{2}\right)} \left(\sigma^2\right)^{-\frac{T}{2}} \left|\Sigma_{22}\right|^{-\frac{1}{2}} \left[1 + \frac{1}{v} (X_t - \mu_2) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right)\right]^{-\frac{v+k+1}{2}}. \tag{2.5}$$

The logarithmic form of (2.5) is better suited for maximization

$$LLF = T \ln \left(\frac{\Gamma\left(\frac{v+k+1}{2}\right)}{\left(\pi v\right)^{\frac{T}{2}} \Gamma\left(\frac{v}{2}\right)} \left|\Sigma_{22}\right|^{-\frac{1}{2}} \left[1 + \frac{1}{v} (X_t - \mu_2) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right)\right]^{-\frac{v+k+1}{2}}\right) - \frac{T}{2} \ln \sigma^2 - \frac{T}{2} \ln \left|\Sigma_{22}\right| - \frac{v+k+1}{2} \sum_{t=1}^{T} \ln \left[1 + \frac{1}{v} (X_t - \mu_2) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right)\right],$$

where LLF denotes the log-likelihood function. The first order conditions (FOC) for this formulation take the form

$$\frac{\partial LLF}{\partial \beta} = \frac{v+k+1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} \left(\mu_1 - \mu_2\right) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right) = 0,$$

$$\frac{\partial LLF}{\partial \mu_2} = \frac{v+k+1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} \left[\left(X_t - \mu_2\right) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right)\right] = 0,$$

$$\frac{\partial LLF}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{v+k+1}{2v} \frac{1}{\sigma^4} \sum_{t=1}^{T} \frac{1}{\gamma_t} u_t^2,$$

$$\frac{\partial LLF}{\partial \text{vech}\Sigma_{22}^{-1}} = \frac{T}{2} \left(\text{vec}\Sigma_{22}^{-1}\right) \left(\Sigma_{22}^{-1}(X_t - \mu_2) \left(\Sigma_{22}^{-1}(X_t - \mu_2) + \frac{1}{v\sigma^2}(Y_t - \beta_0 - \beta X_t)^2\right)\right) = 0,$$

where

14 The proportionality has been replaced by an equality sign since it results in no loss of information.
15 The substitution of $\mu_1 - \beta \mu_2$ for $\beta_0$ was made in this calculation.
\[
\gamma_t = 1 + \frac{1}{\nu} (X_t - \mu_2)' \Sigma_{22}^{-1} (X_t - \mu_2) + \frac{1}{\nu \sigma^2} (Y_t - \beta_0' - \beta' X_t)^2,
\]

\[
u_t = Y_t - \beta_0 - \beta' X_t,
\]

\(G\) is a selector matrix which transforms \(\text{vech}\Sigma_{22}^{-1}\) into \(\text{vec}\Sigma_{22}^{-1}\), and \(\otimes\) denotes a Kronecker product.\(^{16}\)

The above first order conditions are non-linear and have no closed form solution. To get some idea about the form of the implied Maximum Likelihood Estimators (MLEs), Spanos (1994), in a similar case, sets the FOC equal to zero, and solves holding the value of \(\gamma_t\) fixed. The resulting MLEs are of GLS type. He claims the properties of unbiasedness and consistency for them. Additionally, Goldberger (1991) states, that, under certain conditions, MLE is a Best Asymptotically Normal (BAN) estimator, whose asymptotic distribution is \(N(\theta, \frac{\phi^2}{n})\), where \(N\) denotes the normal distribution, \(\theta\) is the true population parameter, \(\phi^2 = \frac{1}{E \left[ \partial^2 \text{LLF}(X, \theta) \right]}\), and \(n\) is the number of observations.

The fact that the first order conditions are non-linear, and have to be solved iteratively using a numerical procedure, gives rise to some new difficulties. In such a situation, any numerical algorithm is likely to encounter problems with the positive definiteness of \(\hat{\Sigma}_{22}^{-1}\) as well as the positivity of \(\hat{\sigma}^2\). The solutions to these problems will be developed in the next section.

---

\(^{16}\) The vec operator transforms a matrix into a single vector by stacking the columns of the matrix one underneath another. The vech operator performs the same operation as vec, but starts each column at its diagonal element. Thus, in case of a symmetric square matrix (like a variance-covariance matrix) applying the vech operator amounts to extracting functionally independent elements and arranging them in a column vector. For any symmetric matrix \(X\), \(\text{vec}X\) and \(\text{vech}X\) are related via \(\text{vech}X = H \text{vec}X\) and \(\text{vec}X = G \text{vech}X\), where \(H\) and \(G\) are selector matrices. For further details see Henderson and Searle (1979).
Let us start tackling the problems outlined at the end of section 2.1 by considering the positivity of $\hat{\sigma}^2$.\(^{17}\) The obvious way to proceed, is to utilize the fact, that the square of any number, different from zero, is positive. Thus, if we can estimate $\hat{\sigma}$, then $\hat{\sigma}^2$ will be positive automatically. Subsequently, $\hat{\sigma}$ can be redefined as the (positive) square root of $\hat{\sigma}^2$.\(^{18}\) This suggests, that taking the relevant derivative with respect to $\sigma$, rather than $\sigma^2$, should solve the problem.

Extending the same argument to a matrix case, suggests factorizing $\Sigma_{22}^{-1}$ into a product of two matrices, say

$$\Sigma_{22}^{-1} = A^\prime A,$$

where $A$ can be a symmetric matrix or a Cholesky factorization. In the following argument it is assumed to be a symmetric matrix.\(^{19}\)

The log-likelihood function, reproduced below for easy reference, takes the form

$$\text{LLF} = \Gamma \left( \frac{v + k + 1}{2} \right) \frac{2}{(\pi v)^{v/2} \Gamma(v/2)} \left( \frac{T}{2} \ln \sigma^2 + \frac{T}{2} \ln |A^\prime A| - \frac{T}{2} \ln \sigma^2 + \frac{T}{2} \ln |A^\prime A| \right)$$

$$- \frac{v + k + 1}{2} \sum_{t=1}^{T} \ln \left[ 1 + \frac{1}{v} \left( X_t - \mu_2 \right)^\prime A^\prime A (X_t - \mu_2) + \frac{1}{v \sigma^2} (Y_t - \beta_0 - \beta^\prime X_t)^2 \right],$$

where the substitution of $A^\prime A$ for $\Sigma_{22}^{-1}$ has been made.

---

\(^{17}\) The case of $\sigma^2=0$ implies an exact functional relationship, and, therefore, is uninteresting. Because of that, it is dropped from the analysis.

\(^{18}\) This procedure is valid because of the invariance property of the MLEs.

\(^{19}\) The argument generalizes easily to the Cholesky case, when vech, $G$, and $H$, are replaced with vecp, $P$, and $R$ (for their definitions see Chapter 4). Both forms involve the same number of distinct elements in $A$, and neither one offers any advantage in precision of the estimates.
Differentiating the LLF with respect to $\beta$, $\mu_2$, $\sigma$, and $A$ produces the estimable form of first order conditions\(^{20}\)

\[
\frac{\partial \text{LLF}}{\partial \beta} = \frac{v + k + 1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} \frac{u_t}{\sigma^2} (X_t - \mu_2)' ,
\]

(2.7)

\[
\frac{\partial \text{LLF}}{\partial \mu_2} = \frac{v + k + 1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} \left[ \left( X_t - \mu_2 \right)' A' A - \frac{u_t}{\sigma^2} \beta' \right] ,
\]

(2.8)

\[
\frac{\partial \text{LLF}}{\partial \sigma} = -\frac{T}{\sigma} - \frac{v + k + 1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} u_t^2 ,
\]

(2.9)

\[
\frac{\partial \text{LLF}}{\partial \text{vech} A} = T \left[ \text{vec}(A' A)^{-1} \right]' G H (A' \otimes I_k) G - \frac{v + k + 1}{v} \sum_{t=1}^{T} \frac{1}{\gamma_t} \left( X_t - \mu_2 \right)' \otimes \left( X_t - \mu_2 \right)' G H (A' \otimes I_k) G ,
\]

(2.10)

where $H$ is a selector matrix transforming $\text{vec} \Sigma_{22}^{-1}$ into $\text{vech} \Sigma_{22}^{-1}$, $I_k$ is a $k \times k$ identity matrix, and all other symbols are defined as before.

The model is now operational. Equations (2.7) - (2.10) can be set equal to zero, and solved by computer solvers, in order to maximize (2.6).

One final thought concerns the standard errors. When this system of equations is solved numerically for Maximum Likelihood estimates of $\beta$, $\mu_2$, $\sigma$ and $A$, asymptotic standard errors for these estimates can be obtained from the inverse of the final Hessian. However, the estimates of interest are the elements of $\hat{\Sigma}_{22}^{-1}$, and their standard errors, not the elements of $\hat{A}$. Relying on the invariance property of Maximum Likelihood estimators, $\hat{\Sigma}_{22}^{-1}$ can be calculated as

$$
\hat{\Sigma}_{22}^{-1} = \hat{A}' \hat{A} .
$$

---

\(^{20}\) The substitution of $\mu_1 - \beta' \mu_2$ for $\beta_0$ was made in this calculation.
Its standard errors can be derived using the $\delta$-method (see Appendix 2). The equation of the estimated variance-covariance matrix ($\text{vec} \Sigma^{-1}$) of the distinct elements of $\Sigma_{22}$ is given by\(^{21}\)

\[
\text{vec}(\text{vech} \Sigma_{22}^{-1}) = \left[ H(I_{k^2} + K_{k,k})(I_k \otimes \hat{\Delta} G) \right] \left[ H(I_{k^2} + K_{k,k})(I_k \otimes \hat{\Delta} G) \right]',
\]

where $\hat{\Delta}$ is the estimated variance-covariance matrix of the distinct elements of $A$, and $K_{k,k}$ is a commutation matrix.\(^{22}\)

2.3. An Interesting Extension - the Static Student $t$ Regression Model with Dynamic Heteroskedasticity.

An interesting extension of the Static Student $t$ Regression Model was proposed by Spanos (1990). Utilizing the fact, that, in case of non-normal distributions, non-correlation does not coincide with independence, he proceeded to specify a Static Student $t$ Regression Model with a dynamic heteroskedastic variance.

Let \(\{Z_t, t = 1, ..., T\}\), where $Z_t = (Y_t, X_t')'$, $Y_t$: $1 \times 1$, $X_t$: $k \times 1$, be non-correlated and jointly distributed as\(^{23}\)

\[
Z_t \sim \text{St}_{T(k+1)} \left( I_T \otimes \mu, I_T \otimes \Sigma, \nu \right).
\]

Since $Z_t$'s are no longer independent, the reduction has to rely on sequential conditioning

\[
D(Z_t^T) = D(Z_1) \prod_{t=2}^{T} D(Z_t | Z_{t-1}^T)
\]

\[
= D(Z_1) \prod_{t=2}^{T} D(Y_t | X_t, Z_{t-1}) D(X_t | Z_{t-1})
\]

\(^{21}\) First order $\delta$-method.

\(^{22}\) A commutation matrix $K$ consists of rearranged rows of an identity matrix, such that $\text{vec} X = K_{m,n} \text{vec} X'$, for any matrix $X_{mn}$. The subscript on $K$ denotes the dimensionality of $X$. For further details see Magnus and Neudecker (1988).

\(^{23}\) In this section, the order of $Z_t$'s is reversed, for reasons that will become apparent in a moment.
Kelker (1970) proves, that, for elliptically-contoured multivariate distributions, regression functions are linear, and the vector of regression coefficients is given by

$$\beta = \Omega_{22}^{-1} \omega_{21},$$

where $\Omega_{22}$ and $\omega_{21}$ refer to the standard way of partitioning of the scale matrix $\Omega$.\(^{24}\) Using the above fact we can deduce that the vector of regression coefficients for $Y_t | X_t, Z_{t-1}^T$ is given by\(^{25}\)

$$\beta_t = \left[ \begin{array}{c} \Sigma_{22} \\ 0_{(t-1)(k+1) \times k} \end{array} \right]^{-1} \left[ \begin{array}{c} \sigma_{21} \\ 0_{(t-1)(k+1) \times 1} \end{array} \right] = \left[ \begin{array}{c} \Sigma_{22}^{-1} \sigma_{21} \\ 0_{(t-1)(k+1) \times 1} \end{array} \right] = \left[ \begin{array}{c} \beta \end{array} \right],$$

where $\beta$ is the vector of regression coefficients corresponding to the elements of $X_t$ and $0_{(t-1)(k+1) \times 1}$ is the vector of regression coefficients for $Z_{t-1}^T$ (all equal 0). Using the fact that

$$\left[ \begin{array}{c} \beta' \\ 0_{1 \times (t-1)(k+1)} \end{array} \right] \left[ \begin{array}{c} X_t \\ Z_{t-1}^T \end{array} \right] = \beta'X_t,$$

the zeros corresponding to $Z_{t-1}^T$ can be dropped, leaving us with $\beta$ - a vector of usual static regression coefficients. In a similar way we can deduce, that the conditional scale matrix, here reduced to a scalar, takes the form

$$\Sigma_{Y_t | X_t, Z_{t-1}^T} = \left\{ \begin{array}{ccc} \sigma_{11} - \left[ \begin{array}{c} \sigma_{12} \\ 0_{1 \times (t-1)(k+1)} \end{array} \right] \left[ \begin{array}{c} \Sigma_{22} \\ 0_{(t-1)(k+1) \times k} \end{array} \right]^{-1} \left[ \begin{array}{c} \sigma_{21} \\ 0_{(1\times (t-1)(k+1) \times 1)} \end{array} \right] \times \\
1 + \frac{1}{\nu} \left[ (X_t - \mu_2)' (Z_{t-1}^T - 1_{t-1} \otimes \mu)' \right] \left[ \begin{array}{c} \Sigma_{22} \\ 0_{(t-1)(k+1) \times k} \end{array} \right]^{-1} \left[ \begin{array}{c} X_t - \mu_2 \\ 1_{t-1} \otimes \mu \end{array} \right] \right\} \times$$

$$= \sigma^2 \left[ 1 + \frac{1}{\nu} (X_t - \mu_2)' \Sigma_{22}^{-1} (X_t - \mu_2) + \frac{1}{\nu} \sum_{i=1}^{(t-1)} (Z_i - \mu)' \Sigma_{22}^{-1} (Z_i - \mu) \right].$$

\(^{24}\) $\Omega$ is the scale matrix of the joint distribution of $Y_t^T$. In the present case $\Omega = \Omega_{1T}^T = I_{1T} \otimes \Sigma$.

\(^{25}\) In the calculations below, the order of $Z$'s has been reversed, i.e. $Z_{t-1}^T$ has been replaced by $Z_{t-1}^T$. This makes the notation compatible with the standard way of partitioning of $\Omega_{1T}$, where $\Omega_{22}$ is thought of as the conditioning part. It can be easily verified that this change has no impact on the final results, but it makes the calculations much more intuitive.
Thus, the distribution of \((Y_t|X_t,Z_{t-1}^i)\) can be compactly written as

\[
(Y_t|X_t,Z_{t-1}^i) \sim \text{St}(\beta_0 + \beta'X_t, \sigma^2 \left[ 1 + \frac{1}{v}(X_t - \mu_2)' \Sigma_{22}(X_t - \mu_2) + \frac{1}{v} \sum_{i=1}^{i} (Z_i - \mu)' \Sigma^{-1}(Z_i - \mu) \right], v + k + (t - 1)(k + 1)).
\]

(2.11)

The distribution of \((X_t|Z_{t-1}^i)\) can be obtained in the same way

\[
(X_t|Z_{t-1}^i) \sim \text{St}(\mu_2, \sigma^2 \left[ 1 + \frac{1}{v} \sum_{i=1}^{i} (Z_i - \mu)' \Sigma^{-1}(Z_i - \mu) \right], v + (t - 1)(k + 1)).
\]

Spanos (1990) uses (2.11) to propose the following model

\[
E(Y_t|X_t,Z_{t-1}^i) = \beta_0 + \beta'X_t,
\]

\[
\text{var}(Y_t|X_t,Z_{t-1}^i) = \frac{\sigma^2}{v + k + (t - 1)(k + 1) - 2} \times \left[ 1 + \frac{1}{v}(X_t - \mu_2)' \Sigma_{22}(X_t - \mu_2) + \frac{1}{v} \sum_{i=1}^{i} (Z_i - \mu)' \Sigma^{-1}(Z_i - \mu) \right].
\]

This model has a static mean and a dynamic heteroskedastic variance. The form of the variance is related to the ARCH/GARCH formulation. For a more comprehensive discussion of this fact see the original work of Spanos (1990).
Chapter 3

Monte Carlo Examination of the Static Student $t$ Regression Model

In Chapter 1 it was pointed out that in real life many data series exhibit characteristics of non-normal distributions. In particular this applies to some financial data, which tend to be distributed Student $t$.\footnote{See Spanos (1990) and references therein.} Questions were also raised regarding the appropriateness of estimating the Normal Linear Regression Model (NLRM) in these instances, since the presence of heteroskedasticity violates its underlying assumptions. A proper way of handling data which is distributed Student $t$ was outlined in Chapter 2. Following Spanos (1994) the Static Student $t$ Regression Model (SSTRM) was derived. By construction, this model is capable of accommodating the probabilistic structure of the Student $t$ data.

At this stage it is only natural to ask, what are the practical implications of failing to follow the proper methodology in the case when the Static Student $t$ Regression Model is appropriate. The question is an important one, since ignoring the non-normality is a generally accepted way of handling the problem. The theoretical consequences of such conduct were briefly discussed in section 1.1. To study this question in more detail, together with some other related issues, a Monte Carlo experiment was performed.

3.1. Design of the Monte Carlo Experiment.

The Monte Carlo experiment was conducted with two sets of objectives in mind. First, the experiment was designed to illustrate the performance of the SSTRM under a variety of conditions. In this context, the examination of the following issues was deemed particularly interesting:

1. The ability of the Maximum Likelihood estimator (MLE) to estimate the underlying parameters accurately.
2. The ability of the estimated standard errors to describe accurately the dispersion of the estimates. This issue has obvious implications for testing.

3. The ability of the SSTRM to model the conditional variance accurately.

The second objective of the Monte Carlo experiment was a comparison of the SSTRM with the benchmark of NLRM. This comparison was motivated by the common practice of ignoring non-normality and proceeding with the estimation of NLRM, even though its underlying assumptions are not valid. This approach is herein referred to as the naive OLS approach.\(^2\) The experiment was designed to examine how much better one could do by utilizing all the information contained in the data, versus doing the naive OLS.

Some researchers go a step beyond the naive OLS approach, and acknowledge the presence of heteroskedasticity in the Student \(t\) data. To tackle the problem, they usually replace the regular OLS standard errors with the standard errors corrected for heteroskedasticity (HCSEs) proposed by White (1980). Therefore, the Monte Carlo experiment also examined the performance of White’s standard errors in describing the dispersion of the empirical distribution.

To propose a specific design for the experiment a number of arbitrary decisions had to be made. The final setup represented a compromise between the desired depth of the investigation and feasibility. The key choices and their justification are discussed below.

The true underlying characteristics of the data were given by

\[
\begin{pmatrix}
Y_t \\
X_t
\end{pmatrix} \sim \text{St}_3(\mu, \Sigma, \nu) = \text{St}_3\left(\begin{bmatrix} 40 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 50 & 8 & 3 \\ 8 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}, \nu \right).
\]

This implies the following values for the parameters of interest\(^3\)

\(^2\) This term is intended to describe a situation when the researcher does not have any idea that the data is non-normal and proceeds with the OLS estimation. Whether non-normality is overlooked, or ignored on purpose, does not make any difference for the results.

\(^3\) See section 2.1 for an explanation of the relationship between the parameters of interest and the primary parameters of the distribution in (3.1).
The scale matrix was purposefully chosen to yield round values for the parameters of interest. The columns of $X$ were set mutually orthogonal to avoid confounding the results by the effects of collinearity. However, in view of the results in Chapter 5, an examination of the impact of collinearity on the performance of the SSTRM might constitute an interesting extension of this study.

In Chapter 2 it was claimed that the MLE is a Best Asymptotically Normal (BAN) estimator. Therefore its performance in smaller and bigger samples is of interest. Additionally, when the SSTRM is estimated, the asymptotic standard errors for the parameter estimates are calculated from the inverse of the final Hessian. To gain insight into the accuracy of inferences drawn using these asymptotic results, it is essential to examine how well the estimated standard errors approximate the true standard errors for different sample sizes. Sample sizes of 50, 100, 200, and 500 were chosen to study the above issues. The largest size of 500 was adopted for comparability with the dynamic case (see Chapter 5), where hardware limitations precluded using a bigger sample size.

Another interesting question is how the performance of the SSTRM varies with the degrees of freedom. The performance of the SSTRM when the value of $\nu$ is low is most interesting, since many financial data series seem to possess this characteristic.\(^4\) Unfortunately, it also entails more difficulties in generating the data (see section 3.2). The generation is much easier for values of $\nu$ greater than 7, but then the distribution looks very much like normal. As a compromise the following cases were chosen to be studied: $\nu = 4, 6, \text{ and } 8$.

For each combination of sample size and degrees of freedom, 500 samples of data were generated. The number 500 was chosen as a compromise between the amount of time required to conduct a single run and the number of repetitions sufficiently big to give a reasonably good idea of the distribution of the estimator. Some preliminary runs were conducted using 1000

\[^4\text{See Spanos (1990).}\]
repetitions. This resulted in a running time of up to 22 hours in some hard-to-generate cases. Cutting the number of repetitions in half reduced the maximum running time to about 12 hours, with the average between 4 and 6 hours. At the same time, the 1000-repetition empirical distributions looked remarkably similar to the 500-repetition distributions.

For each sample generated, the SSTRM was estimated using Maximum Likelihood (ML), and the NLRM was estimated using OLS. The following data were archived for further analysis:

- Maximum Likelihood estimation:
  - estimates of the parameters;
  - asymptotic standard errors computed from the inverse of the final Hessian;
  - asymptotic standard errors for the elements of $\hat{\Sigma}_{22}^{-1}$ calculated using the first order $\delta$-method;
  - elements of the product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$;
  - standard errors for the elements of $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ calculated using the first order $\delta$-method;
  - final value of the log-likelihood function;
  - elements of $\hat{A}$;
  - estimates of $\Sigma_{22}$ and $\sigma$;

- OLS estimation:
  - estimates of the parameters (regression coefficients and the constant term);
  - standard errors of the estimated parameters;
  - HCSEs;
  - $\hat{\sigma}^3$.

The analysis of the results is presented in section 3.4.

The experiment was conducted using three computers equipped with Intel’s Pentium processors. The processors ran at 75, 90, and 120 MHz. The first two computers were equipped
with 64 MB of RAM and the third one with 32 MB. The computer code for the experiment was written in Gauss-386i, version 3.2.13 (see Gauss Manual).

3.2. Data Generation.

The data for the experiment were generated as follows. First, three raw series of Student $t$ random numbers with mean 0, variance 1, and $\nu$ degrees of freedom were generated. Next, the desired dependence structure was imposed on raw data through

$$\begin{bmatrix} Y_{3 \times 1} \\ X \end{bmatrix} = 1_T \otimes \mu + J \text{chol}\left( \frac{v}{\nu - 2} \Sigma \right),$$

where $J$ is a $T \times 3$ matrix containing the aforementioned 3 series of data, and $\text{chol}(.)$ refers to the upper-triangular Cholesky factorization. The multiplication by $\frac{v}{\nu - 2}$ is necessary, since the original setup (see (3.1)) uses the scale matrix, rather than the variance-covariance matrix. Specification in terms of the scale matrix is preferable, because $\Sigma_{22}^{-1}$ is estimated for the models in Chapter 2. Estimates of $\Sigma_{22}^{-1}$ are directly comparable to the inverse of the relevant part of the design matrix.

The raw Student $t$ numbers were generated using the algorithm proposed by Dagpunar (1988). The algorithm performs well, when the degrees of freedom parameter is equal to 7 or more. For $\nu < 7$ the resulting random series frequently are skewed, and have problems tracking the implied value of kurtosis, which is given by

$$\alpha_4 = 3 + \frac{6}{\nu - 4}; \quad \nu > 4.$$

(3.2)

To correct for this, the skewness and kurtosis of each series of raw numbers were tested, and regions of “acceptance” were established. Maximum allowable skewness was set to $\pm 0.1$. The tolerance for kurtosis was set to $\pm 0.5$ around the value implied by the degrees of freedom (see (3.2)).
The relationship in (3.2) breaks down for $\nu \leq 4$. Using a p-p plot to examine the numbers generated with $\nu = 4$, and varying regions of acceptance for kurtosis, it was found that setting the allowable kurtosis range to 7.5 - 8.5 results in a high proportion of good draws. $^5$ Consequently, in the experiment involving $\nu = 4$, the acceptance boundaries were set to 7.5 and 8.5 respectively.

Dagpunar’s algorithm uses numbers from a uniform distribution as an input, which gives the user an ability to control the seed. Taking advantage of this fact, the following procedure was chosen in order to enable easy reproduction of the data from any given run. The initial seed was set to $2^{11} - 1$. The machine generated a series of Student $t$ random numbers, which was tested against the skewness-kurtosis criteria. If the series did not meet the criteria, another series was drawn. The generation process continued, until the machine had three series, which were acceptable. Then, the starting seed for the next run was set equal to the previous starting seed plus 2.

3.3. Maximum Likelihood Estimation.

The Maximum Likelihood estimation was carried out using the Maxlik procedure, version 4.0.15 (see Maximum Likelihood 4.0 Manual), which is a part of the Gauss package. To decrease the time of computation, the analytical expressions for the vector of first derivatives were supplied to the procedure. Their correctness was verified through a comparison with the numerical gradient. The comparison produced a 7-digit accuracy.

Setting the optimization algorithm to the method due to Boyden, Fletcher, Goldfarb, and Shanno produced the best results. With all other settings at their default values, the convergence was usually achieved in 28 - 35 iterations.$^6$

The procedure provided asymptotic standard errors for the parameter estimates. The asymptotic standard errors were calculated from the inverse of the final Hessian. Maxlik computed the final Hessian numerically as a matrix of first derivatives of the analytical gradient

---

$^5$ See Spanos (forthcoming) for a discussion of the usefulness of p-p plots in determining the correct value of the degrees of freedom parameter in Student $t$ models.

$^6$ Convergence criterion set to a default value of $1 \times 10^{-12}$. 
(whose formula was supplied to the procedure).

3.4. Results.

The results of the experiment are reported below. The attention is restricted to $\beta, \beta_0, \sigma^2$, and $\Sigma_{22}^{-1}$. The estimates of $\mu_2$ are omitted from the analysis, because it is highly unlikely that a researcher would estimate a Student $t$ model just to obtain an estimate of this parameter. A much more convenient estimator of $\mu_2$ is available (the sample mean), which is unbiased and consistent.

3.4.1. Estimates of $\beta_1$ and $\beta_2$.

Descriptive statistics for the empirical distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$, estimated using Maximum Likelihood, are reported in Table 3.1. Table 3.2 contains the descriptive statistics for their OLS counterparts. Additionally, normal kernel density estimates of the empirical distribution of $\hat{\beta}_1$ for various sample sizes and $\nu=6$ are shown in Figure 3.1. Graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 3. In order to enable easy comparisons, graphs of the corresponding empirical distributions of ML and OLS estimates are shown side by side. The dashed line in the pictures represents the contour of the normal density, with the same mean and variance as the data whose distribution is shown in the graph. The density estimates for $\hat{\beta}_2$ look similar and are not shown.

The pictures in Figure 3.1 indicate that the empirical distribution of $\hat{\beta}_1$, estimated by Maximum Likelihood ($\hat{\beta}_1^{ML}$), is fairly close to normal, with some evidence of leptokurticity for smaller sample sizes. The distribution appears symmetric, with the exception of the graph for the sample size of 500. Its mode (and mean) is always close to the underlying true value ($\beta_1=4$). As the sample size increases, the variance of the empirical distribution goes down. (Notice the change of scale in subsequent pictures in Figure 3.1). A look at Table 3.1 confirms the inferences made from the graphs. The mean estimate converges to the underlying true value, and

---

7 $\beta_1$ and $\beta_2$ are respectively the first and second element of the vector $\beta$.
8 Normal kernel density with bandwidth equal to $1.06\sigma T^{-0.2}$, where $\sigma$ is the standard deviation and $T$ is the number of observations. See Silverman (1986).
the standard deviation of the empirical distribution gets smaller, as the sample size increases. Thus, \((\hat{\beta}_1)_{ML}\) appears to be a consistent estimator of \(\beta_1\). Additionally, it should be noticed, that the empirical standard deviation is significantly lower, than the mean of the asymptotic standard errors computed from the inverse of the final Hessian (see Table 3.1 and Figure 3.2).\(^9\) As the sample size increases, both the mean of the estimated standard errors and the empirical standard deviation tend to 0, and the difference between them gets smaller. Nevertheless, it should be kept in mind, that, in this experiment, the mean of the estimated standard errors overstated the actual variability by some 200%-400%.

Table 3.1. Selected descriptive statistics of the empirical distribution of ML estimates of \(\beta_1\) and \(\beta_2\).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Sample Size**</th>
<th>Empirical Estimate</th>
<th>Empirical SD***</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Estimated Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Empirical SD</td>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td><strong>v=4</strong></td>
<td>50</td>
<td>4.0381</td>
<td>0.1977</td>
<td>0.2167</td>
<td>3.6581</td>
<td>0.3453</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0125</td>
<td>0.1312</td>
<td>-0.1638</td>
<td>3.3186</td>
<td>0.2414</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0106</td>
<td>0.0887</td>
<td>-0.1167</td>
<td>3.3742</td>
<td>0.1711</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0072</td>
<td>0.0608</td>
<td>-0.0664</td>
<td>3.0179</td>
<td>0.1080</td>
</tr>
<tr>
<td></td>
<td><strong>v=6</strong></td>
<td>50</td>
<td>4.0206</td>
<td>0.1479</td>
<td>-0.0142</td>
<td>3.8586</td>
<td>0.3338</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0127</td>
<td>0.0903</td>
<td>0.1332</td>
<td>3.2083</td>
<td>0.2353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0029</td>
<td>0.0636</td>
<td>0.0213</td>
<td>3.2686</td>
<td>0.1662</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0015</td>
<td>0.0394</td>
<td>0.1974</td>
<td>2.7829</td>
<td>0.1048</td>
</tr>
<tr>
<td></td>
<td><strong>v=8</strong></td>
<td>50</td>
<td>4.0243</td>
<td>0.1087</td>
<td>0.0895</td>
<td>4.0929</td>
<td>0.3229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0166</td>
<td>0.0722</td>
<td>0.2154</td>
<td>3.6419</td>
<td>0.2297</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0071</td>
<td>0.0468</td>
<td>0.1594</td>
<td>3.1194</td>
<td>0.1628</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0016</td>
<td>0.0301</td>
<td>-0.0031</td>
<td>2.9274</td>
<td>0.1030</td>
</tr>
</tbody>
</table>

| \(\beta_2\) | **v=4** | 50 | 3.0572 | 0.2466 | -0.0476 | 3.0456 | 0.5142 | 0.1852 | 3.9448 |
|           |    | 100 | 3.0227 | 0.1747 | 0.0185 | 3.4026 | 0.5861 | -0.0234 | 3.1198 |
|           |    | 200 | 3.0119 | 0.1260 | -0.1038 | 3.1304 | 0.2489 | 0.1069 | 3.1453 |
|           |    | 500 | 3.0047 | 0.0841 | -0.0203 | 3.0999 | 0.1560 | 0.1682 | 3.1808 |
|           | **v=6** | 50 | 3.0760 | 0.1969 | 0.3177 | 4.5512 | 0.4905 | 0.1488 | 4.8122 |
|           |    | 100 | 3.0401 | 0.1184 | -0.0472 | 3.2499 | 0.3414 | 0.0629 | 3.3203 |
|           |    | 200 | 3.0159 | 0.0832 | -0.0312 | 3.2138 | 0.2400 | 0.0431 | 3.0988 |
|           |    | 500 | 3.0028 | 0.0552 | 0.0610 | 2.9374 | 0.1507 | -0.0028 | 2.8566 |
|           | **v=8** | 50 | 3.0651 | 0.1421 | 0.4997 | 4.7163 | 0.4725 | -0.1696 | 3.4211 |
|           |    | 100 | 3.0314 | 0.0846 | 0.3739 | 3.4009 | 0.3312 | -0.0028 | 3.1795 |
|           |    | 200 | 3.0173 | 0.0616 | 0.0747 | 3.0550 | 0.2332 | -0.2243 | 3.4573 |
|           |    | 500 | 3.0053 | 0.0388 | -0.0084 | 2.5849 | 0.1472 | 0.0683 | 3.2699 |

---

\(^9\) Descriptive statistics in Table 3.1 and the pictures in Figure 3.2 show that the empirical distribution of the asymptotic standard errors is not particularly skewed. Because of that, the mean is used to characterize their behavior.
Maximum Likelihood estimates of $\beta_1, \nu=6$  

OLS estimates of $\beta_1, \nu=6$

50 observations

100 observations

200 observations

500 observations

Figure 3.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, \nu=6$.\textsuperscript{10}

\textsuperscript{10} The labels under the graphs refer to the size of the Student $t$ data set that was used in estimation. They should NOT be interpreted as the number of estimates giving rise to the curves shown in the graphs. Each curve was generated using 500 estimates.
The set of graphs on the right-hand side (RHS) of Figure 3.1 shows kernel density estimates of the empirical distribution of \( \hat{\beta}_1 \) estimated by OLS. The distribution is skewed to the right, but its mode and mean remain close to the true value (\( \beta_1=4 \)).\(^{11}\) The graphs also show strong evidence of leptokurticity. Despite the skewness, the distribution is very compact with only a fraction of the dispersion exhibited by the Maximum Likelihood estimator. The difference in variance between the empirical distribution of MLE and OLS estimator (OLSE) becomes more pronounced as the sample size increases. The numbers in Table 3.2 support the conclusions reached by visual inspection of kernel density graphs. As the sample size gets bigger, the mean of the empirical distribution of OLSE converges to the true \( \beta_1 \), and the variance goes down. This confirms the consistency. It should be noticed, that, despite the skewed nature of the distribution of the OLS estimator, its empirical standard deviation is smaller than the corresponding standard deviation in the Maximum Likelihood case. This time, the empirical distribution of the estimated standard errors is skewed and leptokurtic. Its mode and mean understate the true variability (see Figure 3.2 and Table 3.2). It comes somewhat as a surprise to see that the standard errors corrected for heteroskedasticity overshoot the mark so badly. However, as the sample size increases, the HCSEs get closer to the empirical standard deviation, confirming their consistency. Additionally, it should be noticed, that, in this experiment, they seem to offer a reasonable approximation of the asymptotic standard errors computed from the inverted Hessian.

\(^{11}\) In each graph, the mean of the superimposed normal distribution was set equal to the mean of the data shown in this graph.
Table 3.2. Selected descriptive statistics of the empirical distribution of OLS estimates of $\beta_1$ and $\beta_2$.

<table>
<thead>
<tr>
<th>Parm. Name</th>
<th>DF</th>
<th>Sample Size'</th>
<th>Empirical Distribution</th>
<th>Estimated Standard Errors</th>
<th>HCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean Estimate</td>
<td>SD</td>
<td>Skewness</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>v=4</td>
<td>50</td>
<td>4.0275</td>
<td>0.0862</td>
<td>1.3909</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0155</td>
<td>0.0410</td>
<td>1.9692</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0076</td>
<td>0.0179</td>
<td>1.3669</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0036</td>
<td>0.0068</td>
<td>1.4576</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>v=6</td>
<td>50</td>
<td>4.0305</td>
<td>0.0804</td>
<td>0.8004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0129</td>
<td>0.0342</td>
<td>2.7236</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0076</td>
<td>0.0178</td>
<td>1.1250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0028</td>
<td>0.0063</td>
<td>1.1803</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>v=8</td>
<td>50</td>
<td>4.0289</td>
<td>0.0701</td>
<td>0.8883</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.0140</td>
<td>0.0324</td>
<td>0.9896</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.0064</td>
<td>0.0159</td>
<td>0.1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.0029</td>
<td>0.0061</td>
<td>1.0777</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>v=4</td>
<td>50</td>
<td>3.0804</td>
<td>0.1174</td>
<td>1.9716</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>3.0449</td>
<td>0.0612</td>
<td>1.7344</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>3.0203</td>
<td>0.0255</td>
<td>1.5358</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>3.0075</td>
<td>0.0098</td>
<td>1.2973</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>v=6</td>
<td>50</td>
<td>3.0869</td>
<td>0.1251</td>
<td>1.7498</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>3.0404</td>
<td>0.0530</td>
<td>2.1164</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>3.0193</td>
<td>0.0253</td>
<td>1.3567</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>3.0078</td>
<td>0.0102</td>
<td>2.1942</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>v=8</td>
<td>50</td>
<td>3.0804</td>
<td>0.1045</td>
<td>1.5445</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>3.0439</td>
<td>0.0526</td>
<td>1.4786</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>3.0194</td>
<td>0.0268</td>
<td>1.4896</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>3.0073</td>
<td>0.0088</td>
<td>1.4974</td>
</tr>
</tbody>
</table>

Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
Figure 3.2. Empirical distribution of ML and OLS standard errors, $v=6$. 
The experiment confirmed that both MLE and OLSE are consistent estimators of the true regression coefficients. But it also revealed that the distribution of the OLS estimator is skewed and leptokurtic. Additionally, the experiment indicated that both MLE and OLSE have problems accurately describing the true variability of the estimates in finite samples, although they tend to err in different directions. The implications of these findings for testing will be considered next.

A researcher testing a hypothesis related to his estimates can make two basic types of mistakes. These mistakes are known as type-I and type-II errors. Type-I error refers to a situation when the hypothesis under consideration is true, but despite that, it is rejected. The probability of this type of error is known as the size of the test. Type-II error refers to a situation when the hypothesis is false, but it does not get rejected. There is a tradeoff between the two types of errors, which presents a dilemma. The traditional way of solving this problem is to fix the probability of a type-I error at some level (called a level of significance, usually set at 5%) and rely on a test that minimizes the probability of a type-II error.

When the usual $t$-test is performed, choosing a 5% significance level corresponds to a critical value of $\pm 1.96$. Since 5% of the Student $t$ distribution lies outside of this range, the true hypothesis is rejected 5% of the time. However, to be certain that the rejection of the true hypothesis happens only 5% of the time, the researcher has to make sure, that the test statistic actually follows the Student $t$ distribution. If the test statistic is not distributed Student $t$, the actual probability of a type-I error is different than the intended 5%. In this situation the actual probability of making a type-I error is known as the actual size of the test. The significance level chosen by the researcher (i.e. 5%) is referred to as the nominal size.

It should also be kept in mind that the $t$-tests performed on the results of OLS are justified for finite samples, only if normality is assumed. The ML $t$-tests are usually justified asymptotically in a more general setup. However, it is a common practice to proceed with the $t$-test regardless of the sample size and the distributional features of the data. To gain an idea about the

\[ \text{Two sided } t\text{-test when } \nu=\infty. \text{ For } \nu=50 \text{ the critical value equals 2.009 and for } \nu=100 \text{ it is 1.984.} \]
consequences of such approach, the actual size of the $t$-test, and its power curve, have to be examined.

To examine the actual size of the $t$-test, for both Maximum Likelihood and OLS, we have to analyze the empirical distributions of their “$t$-statistics”.\textsuperscript{13} The empirical distributions of these $t$-statistics, for various sample sizes for $\nu=6$, are shown in Figure 3.3. Distributions for $\nu=4$ and $\nu=8$ look similar and are not shown. The distribution of the $t$-statistic for the ML estimates is bell-shaped and reasonably symmetric. However, it should be noticed, that the entire empirical distribution is contained inside the range (-1.96;1.96). Thus, even though the nominal size of the test is set at 5%, there is virtually no possibility of a type-I error, when Maximum Likelihood results are used for testing.

On the other hand, the distribution of the OLS $t$-statistic is skewed and leptokurtic and some portion of it lies outside of the interval corresponding to the 5% significance level. In this case the actual probability of a type-I error is equal to 23.0%, 15.6%, 18.6%, and 16.6% for the sample sizes of 50, 100, 200, and 500 respectively.

\textsuperscript{13} Calculated as $\frac{\hat{\beta}_1 - 4}{\text{SE}(\hat{\beta}_1)}$ for $H_0: \beta_1=4$. SE(.) in this expression denotes the estimated standard error.
Figure 3.3. Empirical distributions of the ML and OLS $t$-statistics, $\nu=6$. 
Another important characteristic of a test is its ability to reject false hypotheses. This is known as the power of the test. The empirical power curves for the \( t \)-tests performed on the results of Maximum Likelihood, OLS, and OLS with White’s standard errors, to test hypotheses about \( \beta_1 \), are shown in Figure 3.4. Sample sizes of 50 and 500 were chosen for this illustration. These empirical power curves were obtained using the following method. First, the range of hypothetical values for \( \beta_1 \) was chosen. For the sample size of 50, this range was set to \( 4 \pm 1.2 \), and for the sample of 500, it was set at \( 4 \pm 0.3 \).\(^{14}\) Next, 100 points were chosen within the range, such that the distances between the points were equal. The points represented hypothetical values of \( \beta_1 \). For each of the 100 hypothetical values of \( \beta_1 \), the \( t \)-statistics for 500 estimates were calculated using the formula

\[
t = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}.
\]

Finally, for each hypothetical value of \( \beta_1 \), the percentage of the \( t \)-statistics lying outside the interval \((-2.01;2.01)\) for the sample of 50, and \((-1.96;1.96)\) for the sample of 500, was calculated.

\(^{14}\) There is no special justification for the ranges that were chosen, other than the fact that they are sufficiently wide to include the portion of the power curve, which is of interest, for all three tests.
The graphs confirm previously reached conclusions regarding the type-I error. The power curve of the ML $t$-test approaches zero at the true value of $\beta_1 (=4)$.\textsuperscript{15} Thus, as was noted before, the true hypothesis is extremely unlikely to be rejected with this technique. On the other hand, the power curve of the OLS $t$-test always lies above 5%, indicating that the true hypothesis would have been rejected much more frequently than 5% of the time. However, when White’s standard

\textsuperscript{15} Strictly speaking, power at $\beta_1=4$ is not defined, since $H_0$ is true.
errors are used, instead of the regular OLS standard errors, the actual size of the test approaches 0% at $\beta_1=4$.

The graphs also reveal differences in power between the three methods. Specifically, the graphs indicate, that the OLS $t$-test has much more power than either the ML or White’s $t$-test, for all false hypotheses. In fact, for bigger samples, the OLS $t$-test reaches 100% rejection rate for false hypotheses before Maximum Likelihood $t$-test rejects anything. The power of the OLS $t$-test using White’s standard errors is even lower than that of Maximum Likelihood $t$-test. As an interesting point, it should be noticed, that the power curve of the OLS $t$-test is asymmetric. Reflecting the skewed nature of the distribution of OLS estimates, the OLS $t$-test shows more power on the left.

In conclusion, the comparison of Maximum Likelihood and OLS $t$-tests did not produce a clear winner. All of the examined $t$-tests exhibited a difference between the actual and the nominal size of the test, although they tended to err in different directions. Specifically, in this particular experiment, the ML $t$-test never rejected the true hypothesis, despite its 5% nominal level of significance. It also showed considerably less power than OLS $t$-test against false hypotheses. The price paid for the increased power of the OLS $t$-test, is the rejection of the true hypothesis 15.6%-23% of the time, depending on the sample size. This can be lowered to 0% by replacing the regular OLS standard errors with HCSEs. However, this type of testing was found to have even less power than ML.
3.4.2. Estimates of $\beta_0$.

The empirical distribution of $\hat{\beta}_0$, for $\nu=6$ and various sample sizes, is shown in a series of graphs in Figure 3.5.\(^{16}\) Graphs for $\nu=4$ and $\nu=8$ have been relegated to Appendix 3, since they are similar to the graphs in Figure 3.5.

The graphs reveal that, for most smaller samples, the empirical distribution of $(\hat{\beta}_0)_{ML}$ is skewed to the left and somewhat leptokurtic. However, for samples of 200 and 500, the distribution becomes normal.\(^{17}\) Its mode and mean converge to the true value of the constant term ($\beta_0=10$) as the sample size increases. Additionally, the variance of the distribution decreases as the sample size gets bigger. Thus, $(\hat{\beta}_0)_{ML}$ seems to be a consistent estimator of $\beta_0$.

Descriptive statistics reported in Table 3.3 confirm these findings.

The empirical distribution of the OLS estimates of $\beta_0$ is strongly leptokurtic and skewed to the left (see Figure 3.5). Descriptive statistics in Table 3.3 show that the mean of the distribution converges to the true value ($\beta_0=10$), and the variance decreases, as the sample size gets bigger. Thus, $(\hat{\beta}_0)_{OLS}$ seems consistent. As an interesting point, it should be noticed, that, even though the distribution of the OLS estimates is skewed, it is still more compact than the distribution of Maximum Likelihood estimates.

In conclusion, both MLE and OLSE appear to be consistent estimators of $\beta_0$. The OLS estimator is more precise, despite the skewed nature of its distribution.

\(^{16}\) Maximum Likelihood estimate of $\beta_0$ was calculated as $\hat{\beta}_0 = \hat{\mu}_1 - \hat{\beta}'\hat{\mu}_2$, relying on the invariance property of MLEs. Sample mean of $Y$ was used as an estimate of $\mu_1$.

\(^{17}\) It can be verified using the data from Table 3.3, that, for the sample size of 200 and 500, the distribution comfortably passes the Bera-Jarque test for normality (see Bera and Jarque (1982)) at 5% significance level, for all values of $\nu$. 
Maximum Likelihood estimates of $\beta_0$, $\nu=6$

OLS estimates of $\beta_0$, $\nu=6$

50 observations

100 observations

200 observations

500 observations

Figure 3.5. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$. 
Table 3.3. Selected descriptive statistics of the empirical distribution of ML and OLS estimates of $\beta_0$.

<table>
<thead>
<tr>
<th>DF</th>
<th>Sample Size</th>
<th>Maximum Likelihood</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Estimate</td>
<td>Empirical SD</td>
</tr>
<tr>
<td>$\nu=4$</td>
<td>50</td>
<td>9.6551</td>
<td>1.3148</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9.8880</td>
<td>0.8997</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.9137</td>
<td>0.5937</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.9472</td>
<td>0.4019</td>
</tr>
<tr>
<td>$\nu=6$</td>
<td>50</td>
<td>9.7168</td>
<td>1.0320</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9.8391</td>
<td>0.5987</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.9489</td>
<td>0.4221</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.9844</td>
<td>0.2746</td>
</tr>
<tr>
<td>$\nu=8$</td>
<td>50</td>
<td>9.7221</td>
<td>0.7228</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9.8145</td>
<td>0.4863</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.9202</td>
<td>0.3208</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.9798</td>
<td>0.1955</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

3.4.3. Estimates of $\sigma^2$.

The next series of pictures (see Figure 3.6) shows the empirical distribution of Maximum Likelihood and OLS estimates of $\sigma^2$ for $\nu=6$. Again, the graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 3. Descriptive statistics for the empirical distributions of ML and OLS estimates are summarized in Table 3.4 and Table 3.5.

At the outset, it should be stressed, that the two sets of estimates are not exactly comparable, because they do not measure the same thing. The usual $\hat{\sigma}^2$, from the OLS regression ($\hat{\sigma}^2_{\text{OLS}}$), represents an estimate of a homoskedastic conditional variance, whose formula is given by

$$\sigma^2_{\text{OLS}} = \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21},$$

where the variance-covariance matrix is denoted by $V$ to avoid confusion with the scale matrix, and the vectors/matrices on the RHS denote the partitions of $V$ in an obvious notation. On the other hand, $\hat{\sigma}^2_{\text{ML}}$ is an estimate of just one part of a heteroskedastic conditional variance, given by

$$\hat{\omega}_i^2 = \frac{\nu}{\nu + k - 2} \sigma^2_{\text{ML}} \left[ 1 + \frac{1}{\nu} (X_i - \mu) \right] \Sigma_{22}^{-1} (X_i - \mu),$$

where $V = \Sigma_{22}$.
where
\[ \sigma_{ML}^2 = \sigma_{11} - \sigma_{12} \Sigma_{21} \sigma_{21}, \quad (3.4) \]
and the notational convention for the partitions of the scale matrix is the same as in Chapter 2.

The expressions (3.3) and (3.4) are related via \( \sigma_{OLS}^2 = \frac{V}{V-2} \sigma_{ML}^2 \). This relationship stems from the fact that \( V = \frac{V}{V-2} \Sigma \). Clearly, if \( \hat{\sigma}_{OLS}^2 \) is a good estimator of \( \sigma_{OLS}^2 \), it will not be a good estimator of \( \sigma_{ML}^2 \). However, it is of interest to see, how well \( \hat{\sigma}_{ML}^2 \) can estimate \( \sigma_{ML}^2 \), and how misled a researcher would be by \( \hat{\sigma}_{OLS}^2 \) (i.e. by assuming that the true joint distribution is normal).

The graphs in Figure 3.6 indicate a slight skewness of the distribution of ML estimates of \( \sigma^2 \). However, it can be verified using the data from Table 3.4, that, for \( V=6 \), the distribution comfortably passes the Bera-Jarque skewness-kurtosis test for normality at the 5% significance level for all sample sizes.\(^{18}\) For \( V=4 \) and \( V=8 \), it also passes this test for the sample size of 500. The graphs show that the mode of the distribution initially moves toward the true value of 9, as the sample size increases, but eventually it overshoots the mark. This effect is less pronounced for higher values of \( V \). For the sample size of 500, the hypothesis, that the difference between the mean estimate and the true value of \( \sigma^2 \) is insignificant, is strongly rejected for \( V = 4, 6, \) and \( 8 \) (\( t \)-statistics 58.9, 21.9, and 30.3 respectively; 5% critical value \( \pm 1.96 \)). Thus, \( \hat{\sigma}_{ML}^2 \) appears to be a biased and inconsistent estimator of \( \sigma^2 \).

Descriptive statistics reported in Table 3.4 confirm the conclusions reached by visual examination of the graphs of kernel density. In addition to the characteristics of the empirical distribution of \( \hat{\sigma}_{ML}^2 \), the table contains also a summary of the relevant descriptive statistics for \( \hat{\sigma}_{ML}^2 \). These are included, because the original operational model in Chapter 2 was specified in terms of \( \sigma \). A look at the mean estimate column for \( \hat{\sigma}_{ML} \) reveals that it suffers from the same problem as \( \hat{\sigma}_{ML}^2 \). Thus, \( \hat{\sigma}_{ML}^2 \) may be a biased and inconsistent estimator of \( \sigma_{ML}^2 \). Additionally, \(^\text{18}\) See Bera and Jarque (1982).
the empirical distribution of the estimated standard errors of $\hat{\sigma}_{ML}$ is skewed in most cases. The mean of this distribution overstates the empirical standard deviation. However, as the sample size increases, the difference between the mean estimated standard error and the empirical standard deviation becomes smaller.

The empirical distribution of the OLS estimates of $\sigma^2$ is skewed to the left and strongly leptokurtic (see Figure 3.6). As the sample size increases, the mode of the distribution converges to $\frac{v}{v-2} \sigma^2_{ML}$, as expected. The distribution of the OLS estimates is much more concentrated around this value than the distribution of ML estimates around its own mean. Descriptive statistics of the empirical distribution of OLS estimates can be found in Table 3.5.

The practical consequences of the above findings vary between ML and OLS. In case of Maximum Likelihood, they are actually less serious, than it may seem at the first glance. However, the discussion of that fact will be postponed, until the distribution of $\hat{\sigma}^2\hat{\Sigma}^{-1}_{22}$ is introduced.

\[19\] In the case under consideration $\nu=4, 6, 8$ and $\sigma^2_{ML} = 9$, which implies $\sigma^2_{OLS} = 18, 13.5$ and 12 respectively.
Figure 3.6. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=6$. 

Maximum Likelihood estimates of $\sigma^2$, $\nu=6$ 

OLS estimates of $\sigma^2$, $\nu=6$ 

50 observations 

100 observations 

200 observations 

500 observations
Table 3.4. Selected descriptive statistics of the empirical distributions of $\sigma^2$ and $\sigma$ estimated by ML.

<table>
<thead>
<tr>
<th>DF</th>
<th>Samp. Size*</th>
<th>$\sigma^2$ Mean</th>
<th>$\sigma^2$ SD</th>
<th>$\sigma^2$ Skewness</th>
<th>$\sigma^2$ Kurtosis</th>
<th>$\sigma$ Empirical Distribution Mean</th>
<th>$\sigma$ Empirical Distribution SD</th>
<th>$\sigma$ Empirical Distribution Skewness</th>
<th>$\sigma$ Empirical Distribution Kurtosis</th>
<th>$\sigma$ Estimated Standard Errors Mean</th>
<th>$\sigma$ Estimated Standard Errors SD</th>
<th>$\sigma$ Estimated Standard Errors Skewness</th>
<th>$\sigma$ Estimated Standard Errors Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu=4$</td>
<td>50</td>
<td>7.9722</td>
<td>1.1313</td>
<td>0.2396</td>
<td>3.5671</td>
<td>2.8164</td>
<td>0.2007</td>
<td>-0.0227</td>
<td>3.4515</td>
<td>0.3585</td>
<td>0.0668</td>
<td>3.9575</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9.0532</td>
<td>0.8494</td>
<td>-0.0869</td>
<td>3.5536</td>
<td>3.0055</td>
<td>0.1421</td>
<td>-0.2791</td>
<td>3.8378</td>
<td>0.2675</td>
<td>-0.3868</td>
<td>4.6132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.7063</td>
<td>0.5948</td>
<td>-0.0285</td>
<td>3.1552</td>
<td>3.1140</td>
<td>0.0957</td>
<td>-0.1276</td>
<td>3.1667</td>
<td>0.1947</td>
<td>-0.1881</td>
<td>3.1102</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>10.1617</td>
<td>0.4414</td>
<td>0.1125</td>
<td>2.9539</td>
<td>3.1870</td>
<td>0.0692</td>
<td>0.0494</td>
<td>2.9416</td>
<td>0.1257</td>
<td>0.0612</td>
<td>2.9256</td>
<td></td>
</tr>
<tr>
<td>$\nu=6$</td>
<td>50</td>
<td>8.2635</td>
<td>0.8030</td>
<td>0.0534</td>
<td>3.4879</td>
<td>2.8712</td>
<td>0.1402</td>
<td>-0.1238</td>
<td>3.4149</td>
<td>0.3440</td>
<td>-0.1647</td>
<td>4.4132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>8.7189</td>
<td>0.5698</td>
<td>-0.0345</td>
<td>3.0117</td>
<td>2.9512</td>
<td>0.0967</td>
<td>-0.1338</td>
<td>3.0398</td>
<td>0.2485</td>
<td>-0.1010</td>
<td>3.0581</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.0543</td>
<td>0.4299</td>
<td>-0.1237</td>
<td>2.6865</td>
<td>3.0082</td>
<td>0.0716</td>
<td>-0.1835</td>
<td>2.7075</td>
<td>0.1782</td>
<td>-0.2528</td>
<td>3.0870</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.2634</td>
<td>0.2690</td>
<td>-0.0911</td>
<td>2.9224</td>
<td>3.0433</td>
<td>0.0442</td>
<td>-0.1236</td>
<td>2.9186</td>
<td>0.1137</td>
<td>-0.0534</td>
<td>3.1760</td>
<td></td>
</tr>
<tr>
<td>$\nu=8$</td>
<td>50</td>
<td>8.8643</td>
<td>0.6403</td>
<td>-0.3466</td>
<td>3.2452</td>
<td>2.9753</td>
<td>0.1085</td>
<td>-0.4679</td>
<td>3.4486</td>
<td>0.3430</td>
<td>-0.4815</td>
<td>3.5529</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9.0639</td>
<td>0.4369</td>
<td>-0.3092</td>
<td>3.4740</td>
<td>3.0097</td>
<td>0.0729</td>
<td>-0.4183</td>
<td>3.6085</td>
<td>0.2444</td>
<td>-0.5657</td>
<td>4.0122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.1906</td>
<td>0.2953</td>
<td>-0.3593</td>
<td>3.0160</td>
<td>3.0312</td>
<td>0.0506</td>
<td>-0.4071</td>
<td>3.0718</td>
<td>0.1737</td>
<td>-0.5537</td>
<td>3.7866</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.2836</td>
<td>0.2092</td>
<td>-0.1600</td>
<td>2.9976</td>
<td>3.0467</td>
<td>0.0344</td>
<td>-0.1935</td>
<td>3.0212</td>
<td>0.1102</td>
<td>-0.1284</td>
<td>2.9590</td>
<td></td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

Table 3.5. Selected descriptive statistics of the empirical distributions of $\sigma^2$ estimated by OLS.

<table>
<thead>
<tr>
<th>DF</th>
<th>Sample Size</th>
<th>Mean Estimate</th>
<th>Empirical SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu=4$</td>
<td>50</td>
<td>18.9542</td>
<td>0.8918</td>
<td>-0.1493</td>
<td>11.9799</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>18.4342</td>
<td>0.4334</td>
<td>-2.7590</td>
<td>28.2714</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>18.2185</td>
<td>0.1701</td>
<td>-0.9551</td>
<td>7.0164</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>18.0862</td>
<td>0.0688</td>
<td>-0.3413</td>
<td>5.1457</td>
</tr>
<tr>
<td>$\nu=6$</td>
<td>50</td>
<td>14.1079</td>
<td>0.6038</td>
<td>-1.0457</td>
<td>10.2709</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>13.8074</td>
<td>0.2799</td>
<td>-0.9771</td>
<td>7.1049</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>13.6679</td>
<td>0.1242</td>
<td>-1.1192</td>
<td>7.3353</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>13.5647</td>
<td>0.0507</td>
<td>-1.4694</td>
<td>8.6649</td>
</tr>
<tr>
<td>$\nu=8$</td>
<td>50</td>
<td>12.5592</td>
<td>0.4776</td>
<td>-1.0557</td>
<td>6.0915</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>12.2899</td>
<td>0.2208</td>
<td>-1.3678</td>
<td>8.3848</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>12.1403</td>
<td>0.1278</td>
<td>-3.7983</td>
<td>34.9683</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>12.0570</td>
<td>0.0371</td>
<td>-0.9740</td>
<td>5.5916</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

3.4.4. Estimates of $\Sigma^{-1}_{22}$.

Graphs of the empirical distributions of ML estimates of two elements of the inverted scale matrix $\Sigma^{-1}_{22}$, for $\nu=6$, are presented in Figure 3.7. Graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 3. The elements chosen for this illustration are: $(\Sigma^{-1}_{22})_{11} = 0.5$ - the first element on the main diagonal and $(\Sigma^{-1}_{22})_{21} = 0$ - the off-diagonal element. The graphs for the estimates of $(\Sigma^{-1}_{22})_{22}$ look similar to the graphs for $(\Sigma^{-1}_{22})_{11}$, and are not shown.

20 In case of OLS, the problem of estimating the scale matrix does not arise, for obvious reasons.
Maximum Likelihood estimates of \((\Sigma_{22}^{-1})_{11}, \nu=6\)

Maximum Likelihood estimates of \((\Sigma_{22}^{-1})_{31}, \nu=6\)

50 observations

50 observations

100 observations

100 observations

200 observations

200 observations

500 observations

500 observations

Figure 3.7. Kernel density estimates of the empirical distribution of \((\hat{\Sigma}_{22}^{-1})_{11}\) and \((\hat{\Sigma}_{22}^{-1})_{31}\) (ML).
The graphs in Figure 3.7 indicate, that the distribution of \( (\mathbf{\Sigma}^{-1})_{11} \) is skewed to the right, and shows evidence of leptokurticity for smaller sample sizes. As the sample size increases, the mode of the empirical distribution shifts to the left. This phenomenon is reminiscent of the problems with the mode of the distribution of estimates of \( \sigma^2 \). This time, however, the mode shifts in the opposite direction. This suggests taking a closer look at the distribution of the product \( \hat{\sigma}^2 \mathbf{\Sigma}^{-1} \) - a strategy which will be pursued in a moment, and later again in Chapter 5. The empirical distribution of \( (\hat{\mathbf{\Sigma}})^{-1} \) shows evidence of skewness and leptokurticity for smaller samples. As the sample size increases, the distribution tends towards normal. Its mode is close to the true underlying value of 0 and does not exhibit any shifting. The variance of the distributions of \( (\hat{\mathbf{\Sigma}})^{-1} \) and \( (\hat{\mathbf{\Sigma}})^{-1} \) decreases as the sample size gets bigger.

Descriptive statistics for the empirical distributions of \( \hat{\mathbf{\Sigma}}^{-1} \) and \( \hat{\mathbf{\Sigma}}^{-1} \) are reported in Table 3.6. A look at the column of mean estimates for \( \hat{\mathbf{\Sigma}}^{-1} \) reveals that the mean suffers from the same shifting problem, that was earlier detected for the mode. The hypothesis that the mean of the distribution is equal to the underlying true value (=2) is strongly rejected for the sample size of 500 for all values of \( \nu \). Thus, the Maximum Likelihood estimator of \( \mathbf{\Sigma}^{-1} \) may be biased and inconsistent. Table 3.6 also contains selected characteristics of the estimated standard errors for \( \hat{\mathbf{\Sigma}}^{-1} \) and \( \hat{\mathbf{\Sigma}}^{-1} \). The estimated standard errors were calculated using the first order \( \delta \)-method (see Chapter 2). The distribution of the estimated standard errors is skewed and its mean overstates the actual standard deviation. However, the difference between the mean estimated standard error and the empirical standard deviation gets smaller as the sample size increases. This confirms the asymptotic validity of the \( \delta \)-method.

---

21 \( t \)-statistics -68.16 for \( \nu = 4 \), -26.89 for \( \nu = 6 \), and -37.27 for \( \nu = 8 \). These values are so much bigger than the 5% critical value (±1.96), that, despite the skewed nature of the distribution, it can be safely concluded that the hypothesis is rejected.

22 A more precise result can be derived using the second order approximation (see Spanos (forthcoming)).
Table 3.6. Selected descriptive statistics of the empirical distributions of \( \hat{\Sigma}_1^{22} \) and \( \hat{\Sigma}_2^{22} \)

<table>
<thead>
<tr>
<th>Parm. Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>Empirical Distribution</th>
<th>Estimated Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean Estimate</td>
<td>SD</td>
</tr>
<tr>
<td>( \hat{\Sigma}_1^{22} )</td>
<td>v=4</td>
<td>50</td>
<td>0.5562</td>
<td>0.0783</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.4905</td>
<td>0.0442</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.4629</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.4430</td>
<td>0.0187</td>
</tr>
<tr>
<td>( \hat{\Sigma}_2^{22} )</td>
<td>v=6</td>
<td>50</td>
<td>0.5326</td>
<td>0.0536</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.5106</td>
<td>0.0335</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.4945</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.4828</td>
<td>0.0143</td>
</tr>
<tr>
<td>( \hat{\Sigma}_3^{22} )</td>
<td>v=8</td>
<td>50</td>
<td>0.4990</td>
<td>0.0318</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.4884</td>
<td>0.0237</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.4859</td>
<td>0.0165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.4825</td>
<td>0.0105</td>
</tr>
</tbody>
</table>

Size of the Student \( t \) data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

First order \( \delta \)-method.

3.4.5. The product \( \hat{\sigma}^2 \hat{\Sigma}^{-1} \).

Figure 3.8 shows kernel density estimates of the empirical distribution of \( \hat{\sigma}^2 \hat{\Sigma}^{-1} \) for \( \nu = 6 \). Again the graphs for \( \nu = 4 \) and \( \nu = 8 \) can be found in Appendix 3. The true values of these expressions are the elements of \( \sigma^2 \Sigma^{-1} \), which equals

\[
9 \times \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4.5 & 0 \\ 0 & 9 \end{bmatrix}.
\]

The graphs in Figure 3.8 indicate that the empirical distribution of \( \hat{\sigma}^2 \hat{\Sigma}^{-1} \) is skewed to the right, although for \( \nu = 8 \) it becomes fairly symmetric for sample sizes of 200 and 500 (see Appendix 3). Its mode and mean can be seen converging to the underlying true value. Descriptive statistics in Table 3.7 support the above findings. The mean of the empirical

---

23 Distribution of the product of estimates (NOT estimates of the product).
distribution of $\hat{\sigma}^2 (\hat{\Sigma}^{-1})_{11}$ converges to $\sigma^2 (\Sigma^{-1})_{11}$, and the variance decreases as the sample size gets bigger. This time, there seems to be no overshooting problem. Oddly enough, the convergence process is faster for lower values of $\nu$.

The empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}^{-1})_{21}$ is somewhat skewed and leptokurtic for smaller sample sizes (see Figure 3.8). However, for bigger samples, the distribution is close to normal. Descriptive statistics in Table 3.7 show that the mean of the distribution converges to the underlying true value as the sample size increases, and the variance goes down.

In conclusion, the product $\hat{\sigma}^2 \hat{\Sigma}^{-1}$ appears to be a consistent estimator of $\sigma^2 \Sigma^{-1}$.

Estimated standard errors reported in Table 3.7 were calculated using the first order $\delta$-method. In this calculation the elements of $\hat{\Sigma}^{-1}$ were treated as random variables and $\hat{\sigma}^2$ as a constant. For most sample sizes and degrees of freedom, the distribution of the estimated standard errors is skewed. As in the previous cases, their mean overstates the actual variability, but the magnitude of overstatement gets smaller as the sample size increases.

---

24 It can be verified using the data in Table 3.7, that, for the sample size of 500, the distribution comfortably passes the Bera-Jarque skewness-kurtosis test for normality (see Bera and Jarque (1982)) at 5% significance level for all values of $\nu$.

25 This assumption was adopted to make the calculation of the standard errors tractable.
Table 3.7. Selected descriptive statistics of the empirical distributions of $\tilde{\sigma}^2(\bar{X}_{11})$ and $\tilde{\sigma}^2(\bar{X}_{21})$.

<table>
<thead>
<tr>
<th>Parm. Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>Empirical Distribution</th>
<th>Estimated Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean Estimate</td>
<td>SD</td>
</tr>
<tr>
<td>$\tilde{\sigma}^2(\bar{X}_{11})$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>4.4372</td>
<td>0.9120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.4402</td>
<td>0.5736</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.4929</td>
<td>0.4122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.5013</td>
<td>0.2681</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>4.4062</td>
<td>0.6525</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.4543</td>
<td>0.4398</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.4779</td>
<td>0.3034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.4724</td>
<td>0.1831</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>4.3376</td>
<td>0.4480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>4.4280</td>
<td>0.3163</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>4.4656</td>
<td>0.2087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>4.4793</td>
<td>0.1403</td>
</tr>
<tr>
<td>$\tilde{\sigma}^2(\bar{X}_{21})$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>-0.0752</td>
<td>0.6361</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.0302</td>
<td>0.4123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.0115</td>
<td>0.2715</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.0026</td>
<td>0.1780</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>-0.1056</td>
<td>0.4441</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.0489</td>
<td>0.2807</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.0252</td>
<td>0.1905</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.0150</td>
<td>0.1151</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>-0.0969</td>
<td>0.3157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.0377</td>
<td>0.2162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.0224</td>
<td>0.1385</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.0101</td>
<td>0.0903</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

** First order $\delta$-method, $\tilde{\sigma}^2$ treated as constant.
Figure 3.8. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}_{21}^{-1})_{11}$ and $\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{21}$, $\nu=6$. 
3.4.6. Modeling the Conditional Variance.

The findings in section 3.4.5 go a long way towards salvaging the usefulness of the SSTRM in modeling of the conditional variance. To see why, recall that the Student $t$ conditional variance is given by

$$\text{var}(Y_t | X_t) = \frac{\nu}{\nu + k - 2} \left[ \sigma^2 + \frac{1}{\nu} (X_t - \mu)^\prime \Sigma^{-1}_{22} (X_t - \mu) \right],$$

where $\sigma^2$ has been brought inside the square bracket.

The above expression contains two terms involving $\sigma^2$. In the first term, $\sigma^2$ appears by itself, which means that the Maximum Likelihood estimator of this component of $\text{var}(Y_t | X_t)$ will suffer from the problems described earlier in this chapter. In the second term, $\sigma^2$ is accompanied by $\Sigma^{-1}_{22}$; the combination which Maximum Likelihood estimated reasonably well. It is important to notice that the part estimated reliably is associated with the $X$’s. Taken together, the above observations imply, that, even though the estimated form of the conditional variance can miss the mark, it should be off by roughly the same amount for different values of the conditioning variables. Thus, the predictions of the variability should be able to track the pattern of the heteroskedastic conditional variance fairly well.

To illustrate this point, a series of 200 observations of $(Y_t, X_t)'$ was drawn using the procedure described in section 3.2. The true distribution of this series was the same as in (3.1). Both the SSTRM and the NLRM were estimated for these data. The results of the estimation were used to generate predictions of the conditional variance, which are shown in Figure 3.9. For better readability only the first 100 observations are shown.
As the graph illustrates, SSTRM is able to mimic the pattern of the conditional variance very well. The distance between the actual and predicted variance is fairly constant, as expected. On the other hand, NLRM by construction ignores the dependence of the conditional variance on the conditioning variables, and attempts to fit a straight line. Thus, the Static Student $t$ Regression Model offers a vast improvement in modeling of the heteroskedastic conditional variance, compared to the traditional NLRM.

3.5. Conclusions.

In this chapter an effort was undertaken to evaluate the performance of the Static Student $t$ Regression Model. The questions of interest included the model’s usefulness for estimation and testing, its ability to model the conditional variance accurately, and its relative performance when compared to the traditional NLRM.

The Monte Carlo experiment confirmed that both MLE and OLSE are consistent estimators of regression coefficients, and of a constant term, when the data are distributed Student $t$. OLS estimator exhibited a lower empirical variance, despite the fact that its distribution was skewed. The empirical distributions of ML estimates of $\sigma^2$ and $\Sigma_{22}^{-1}$ exhibited characteristics which indicate that the MLEs of these parameters might be biased and inconsistent. However, the product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ appears to be a consistent estimator of $\sigma^2 \Sigma_{22}^{-1}$.
The power of the traditional $t$-test, used to test hypotheses about $\beta_1$, was also examined. In this case, the experiment looked at ML, regular OLS, and OLS with HCSEs. All three methods exhibited discrepancies between the nominal and the actual size of the test. Additionally, differences in power between the methods were uncovered, as expected. Specifically, it was found that the OLS $t$-test reaches 100% rejection rate for false hypotheses much faster than the ML $t$-test. Unfortunately, it also rejects the true hypothesis 15.6%-23% of the time, depending on the sample size. Its empirical power curve is asymmetric, showing more power towards $H_0$’s involving values lower than the actual $\beta_1$. On the other hand, the ML $t$-test virtually never rejects the true $H_0$, but it also has a higher probability of a type-II error. It was found, that, in bigger samples, OLS $t$-test reaches 100% rejection rate for false $H_0$’s, before ML $t$-test starts rejecting anything. The probability of a type-I error for OLS $t$-test can be lowered to 0% by using White’s standard errors. However, this combination was found to have even less power than the ML $t$-test. Thus, none of the three methods was found superior to the others, and the choice among them should depend on the particular set of circumstances.

Finally, the SSTRM proved capable of modeling the conditional variance accurately. The predictions generated by the SSTRM were found to follow the pattern of the true conditional variance very closely. On the other hand, NLRM by construction is incapable of modeling the changes in the conditional variance, and generates identical predictions for all observations.

The above findings were largely invariant to changes in the value of the degrees of freedom parameter.
Chapter 4

The Dynamic Student $t$ Regression Model

In this chapter the scope of the analysis is broadened to include leptokurtically distributed variables exhibiting correlation over time. Traditionally, this type of situation can be handled in two ways. The first approach ignores the non-normality and calls for estimation of a DLRM. In the second approach, heteroskedasticity of the conditional variance is recognized and an ARCH/GARCH formulation is estimated. Still another way of treating variables correlated over time, whose distribution is characterized by leptokurticity, was suggested by Spanos (1990). In this path-breaking paper he proposed a series of models based on the Student $t$ distribution, which can account for the probabilistic features of the data. The following investigation concentrates on one of the models proposed by Spanos - the Student $t$ Autoregressive model with dynamic heteroskedasticity (STAR). This particular model was selected because empirical studies indicated that it provides accurate description of the behavior of some speculative prices (see Robertson (1992), Spanos (1993), and McGuirk, Robertson and Spanos (1993)).

4.1. Derivation of the Student $t$ Autoregressive Model.

The following derivation of the STAR model draws heavily on Spanos (1986) and Spanos (1990). The reader is therefore referred to these sources for further details.

Let $\{Y_t, t = 1, \ldots, T\}$ be a stochastic process with the following characteristics:

- $\{Y_t, t = 1, \ldots, T\}$ is a stationary process;
- the conditional mean of $Y_t$ can be described as Markov of order $k$;
- the joint distribution of all the $Y$’s is multivariate-$t$, compactly written as

$$Y_T \sim St_T (\mu_{T}, \Sigma_{T}, v) ,$$

where $Y_T^1$ denotes a vector $(Y_1, \ldots, Y_T)^t$.

$^1$ The ordering of the variables in $Y_T^1$ follows a convention similar to the usual treatment of time-series data, where newer observations are added successively at the bottom end of the series. This notation proves advantageous for
- the scale matrix $\Sigma_1^T$ is a Toeplitz matrix whose elements die out exponentially as the distance from the main diagonal increases (asymptotic independence).

Under the above assumptions $\Sigma_1^T$ has the following structure

$$
\Sigma_1^T = \begin{bmatrix}
\sigma_0 & \sigma_1 & \sigma_2 & \ldots & \sigma_{T-1} \\
\sigma_1 & \sigma_0 & \sigma_1 & \ldots & \sigma_{T-2} \\
\sigma_2 & \sigma_1 & \sigma_0 & \ldots & \sigma_{T-3} \\
& & & \ddots & \\
\sigma_{T-1} & \sigma_{T-2} & \sigma_{T-3} & \ldots & \sigma_0
\end{bmatrix},
$$

where $\sigma_i$'s are such that $\frac{\nu}{\nu - 2} \sigma_i = \text{cov}(y_t, y_{t-i})$; $t = 1, \ldots, T$; $i = 0, 1, \ldots, t-1$. It should be noticed that every leading submatrix of $\Sigma_1^T$ has the same type of structure as $\Sigma_T^1$.

In the absence of independence, the reduction process has to rely on sequential conditioning. It takes the form

$$
D(Y^1_T) = D(Y^1_1) \prod_{t=2}^T D(Y^1_t | Y^1_{t-1}) = D(Y^1_{2k+1}) \prod_{t=2k+2}^T D(Y^1_t | Y^1_{t-1}).
$$

(4.1)

The specific expressions for $D(Y^1_{2k+1})$ and $D(Y_t | Y^1_{t-1})$ can be obtained in the same manner as in the static case

$$
D(Y^1_{2k+1}) = \frac{\Gamma\left(\frac{\nu + 2k + 1}{2}\right)}{(\pi \nu)^{\frac{\nu}{2}}} |\Sigma_{2k+1}^1|^{-\frac{1}{2}} \int \left[1 + \frac{1}{\nu} \left(Y^1_{2k+1} - \mu_{2k+1}\right) \left(\Sigma_{2k+1}^1\right)^{-1} \left(Y^1_{2k+1} - \mu_{2k+1}\right)^T\right]^{-\frac{\nu + 2k + 1}{2}} dY_{2k+1}.
$$

(4.2)

the derivations in this chapter, but it has consequences for the partitioning of the scale matrix. The traditional expression $\Sigma_{2i}^T \sigma_{2i}$, where $\Sigma_{2i}$ is the scale matrix for the vector of conditioning variables, now becomes $\Sigma_{1i}^T \sigma_{1i}$.

2 The decomposition can be carried out further, up to $D(Y_t) \prod_{i=2}^T D(Y_i | Y^1_{t-1})$. However, it is convenient to stop at $t=2k+1$, for reasons which will become apparent in section 4.2, when the structure of $(\Sigma_{1i}^T)^{-1}$ is explained.
\[
D(Y_t | Y_{t-1}) = \frac{\Gamma \left( \frac{V + t}{2} \right)}{(\pi V)^{\frac{N+1}{2}}} \left| \sigma^2 \left[ 1 + \frac{1}{V} (Y_t^i - 1, \ldots, Y_t^i) \left( \Sigma_{t-1}^i \right)^{-1} \left( Y_t^i - 1, \ldots, Y_t^i \right) \right] \right|^{\frac{1}{2}} \times \\
\times \left[ 1 + \frac{1}{V} (Y_t - \beta_0 - \beta Y_{t-1})^i \left( \sigma^2 \left[ 1 + \frac{1}{V} (Y_{t-1}^i - 1, \ldots, Y_{t-1}^i) \left( \Sigma_{t-1}^i \right)^{-1} \left( Y_{t-1}^i - 1, \ldots, Y_{t-1}^i \right) \right] \right)^{-1} (Y_t - \beta_0 - \beta Y_{t-1}) \right]^{\frac{V+1}{2}}, \quad t > 2k+1,
\]

where

\[
\beta = (\Sigma_{t-1}^{-k})^{-1} \sigma_{t-1}^{-k} = \begin{bmatrix}
\sigma_0 & \sigma_1 & \cdots & \sigma_{k-1} \\
\sigma_1 & \sigma_0 & \cdots & \sigma_{k-2} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{k-1} & \sigma_{k-2} & \cdots & \sigma_0
\end{bmatrix}^{-1} \begin{bmatrix}
\sigma_k \\
\sigma_{k-1} \\
\vdots \\
\sigma_1
\end{bmatrix},
\]

\[
\sigma^2 = \sigma_0 - \sigma_{t-1,1}^i \left( \Sigma_{t-1}^i \right)^{-1} \sigma_{t-1,1}^i,
\]

\[
\mu = \text{E}(y_t),
\]

\[
\beta_0 - \text{constant term}.
\]

The parameters of interest are \(\beta\), \(\mu\), \(\sigma^2\), and functionally independent elements of \(\left( \Sigma_{t-1}^i \right)^{-1}\). The mean of (4.3) is given by

\[
\text{E}(Y_t | Y_{t-1}^i) = \beta_0 + \beta^i Y_{t-1}^i,
\]

---

\(^3\) The subscript/superscript in \(\Sigma_{t-1}^{t-k}\) and \(\sigma_{t-1,1}^{t-k}\) refer to \(y_i\)’s, NOT to \(i\) in \(\sigma_i\). Thus, \(\frac{V}{V-2} \Sigma_{t-1}^{t-k}\) is the variance-covariance matrix for \((y_{t-k}, \ldots, y_{t-1})^t\) and \(\frac{V}{V-2} \sigma_{t-1,1}^{t-k}\) is the vector of covariances between \((y_{t-k}, \ldots, y_{t-1})^t\) and \(y_t\). Notice that \(\Sigma_{t-1}^{t-k}\) and \(\sigma_{t-1,1}^{t-k}\) are identical for all \(t>k\).

\(^4\) At first glance, it may seem that since \(\left( \Sigma_{t-1}^i \right)^{-1}\) and \(\sigma_{t-1,1}^i\) change with \(t\), \(\sigma_i\)’s should vary as well (i.e. \(\sigma_i\) would be more appropriate). However, under our assumptions, \(\left( \Sigma_{t-1}^i \right)^{-1}\) and \(\sigma_{t-1,1}^i\) possess a particular structure, which results in all \(\sigma_i\)’s being identical. This point will be explained in more detail in section 4.2.

\(^5\) \(\beta_0\) is defined as \(\mu(1 - \beta^i 1_k)\).

\(^6\) See Zellner (1971).
and the conditional variance takes the form

\[
\text{var}(Y_t | Y_{t-1}) = \frac{v}{v + (t-1)\sigma^2} \left[ 1 + \frac{1}{v} (Y_{t-1}^i - 1_{t-1}\mu) \left( \Sigma_{t-1}^i \right)^{-1} (Y_{t-1}^i - 1_{t-1}\mu) \right],
\]

and is heteroskedastic. It should be noticed, that the conditional mean involves only \(k\) lags, but there is no cutoff lag for the conditional variance. For an extensive discussion of a number of interesting issues related to this model, including a comparison with ARCH/GARCH formulations, the reader is referred to the original work of Spanos. Plugging equations (4.2) and (4.3) into (4.1) produces an explicit expression for the STAR(\(k\)) likelihood function (LF)

\[
\text{LF} = \frac{\Gamma(\frac{v + 2k + 1}{2})}{(\pi v)^{\frac{v}{2}}} \left| \Sigma_{2k+1}^i \right|^{\frac{1}{2}} \left[ 1 + \frac{1}{v} (Y_{2k+1}^i - 1_{2k+1}\mu) \left( \Sigma_{2k+1}^i \right)^{-1} (Y_{2k+1}^i - 1_{2k+1}\mu) \right]^{\frac{v + 2k + 1}{2}} \times
\]

\[
\times \prod_{t=2k+2}^{T} \frac{\Gamma(\frac{v + t}{2})}{(\pi v)^{\frac{v}{2}}} \left| \Sigma_{t}^i \right|^{\frac{1}{2}} \left[ 1 + \frac{1}{v} (Y_{t}^i - 1_{t}\mu) \left( \Sigma_{t}^i \right)^{-1} (Y_{t}^i - 1_{t}\mu) \right]^{\frac{1}{2}} \times
\]

\[
\times \left[ 1 + \frac{1}{v} (Y_t - \beta_0 - \beta Y_{t-k}^i)^\prime \left( \sigma^2 \left[ 1 + \frac{1}{v} (Y_{t}^i - 1_{t}\mu) \left( \Sigma_{t}^i \right)^{-1} (Y_{t}^i - 1_{t}\mu) \right] \right)^{-1} (Y_t - \beta_0 - \beta Y_{t-k}^i) \right]^{\frac{v + 1}{2}}.
\]

As in the static case, the logarithmic form of (4.4) is better suited for maximization. It should also be noticed, that, given the assumption of asymptotic independence, the effect of initial conditions will be inconsequential for large \(t\) (see Spanos (1986), p.528). Therefore, when \(t\) is large relative to \(k\), dropping the first part of (4.4) will make very little difference. Hence, the short version of the log-likelihood function (LLF) takes the form

\[
\text{LLF} = \ln \left( \frac{\Gamma(\frac{v + T}{2})}{(\pi v)^{\frac{v}{2}}} \right) - \frac{T - (2k + 1)}{2} \ln \sigma^2 - \frac{1}{2} \sum_{t=2k+2}^{T} \ln \left[ 1 + \frac{1}{v} (Y_{t}^i - 1_{t}\mu) \left( \Sigma_{t}^i \right)^{-1} (Y_{t}^i - 1_{t}\mu) \right] - \sum_{t=2k+2}^{T} \frac{v + t}{2} \ln \left[ 1 + \frac{1}{v} (Y_t - \beta_0 - \beta Y_{t-k}^i)^\prime \left( \sigma^2 \left[ 1 + \frac{1}{v} (Y_{t}^i - 1_{t}\mu) \left( \Sigma_{t}^i \right)^{-1} (Y_{t}^i - 1_{t}\mu) \right] \right)^{-1} (Y_t - \beta_0 - \beta Y_{t-k}^i) \right].
\]

\(7\) See Zellner (1971).

\(8\) \(k\) refers to the number of lags in Markov mean. In the expression below, the proportionality has been replaced by an equality sign, since it results in no loss of information.
The first order conditions for (4.5) take the form\(^9\)

\[
\frac{\partial \text{LLF}}{\partial \beta} = \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{\gamma_t \delta_t \sigma^2} u_t (Y_{t,k}^t - \mathbf{1}_t \mu)' , 
\]

(4.6)

\[
\frac{\partial \text{LLF}}{\partial \mu} = \frac{1}{v} \sum_{t=2k+2}^{T} \frac{1}{\delta_t} \left[ Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right]' \left( \Sigma_{t-1}^{1} \right)^{-1} \mathbf{1}_{t-1} + 
+ \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{\gamma_t \delta_t \sigma^2} \left[ u_t (1 - \beta' \mathbf{1}_k) - \frac{1}{\delta_t} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' \left( \Sigma_{t-1}^{1} \right)^{-1} \mathbf{1}_{t-1} u_t^2 \right], 
\]

(4.7)

\[
\frac{\partial \text{LLF}}{\partial \sigma^2} = -\frac{T - (2k + 1)}{2} \frac{1}{\sigma^2} + \frac{1}{\sigma^4} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{\gamma_t \delta_t} u_t^2 , 
\]

(4.8)

\[
\frac{\partial \text{LLF}}{\partial \text{vec}p(\Sigma_{t-1}^{1})^{-1}} = -\frac{1}{2v} \sum_{t=2k+2}^{T} \frac{1}{\delta_t} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' \otimes \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' R_{\Sigma_{t-1}} + 
+ \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{\gamma_t \delta_t} u_t^2 \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' \otimes \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' R_{\Sigma_{t-1}} , 
\]

(4.9)

where

\[
\gamma_t = 1 + \frac{1}{v} \left( y_t - \beta_0 - \beta' Y_{t-1}^{t+k} \right)' \left\{ \sigma^2 \left[ 1 + \frac{1}{v} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' \left( \Sigma_{t-1}^{1} \right)^{-1} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right) \right] \right\}^{-1} \left( Y_{t-1} - \beta_0 - \beta' Y_{t-1}^{t+k} \right) , 
\]

\[
\delta_t = 1 + \frac{1}{v} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right)' \left( \Sigma_{t-1}^{1} \right)^{-1} \left( Y_{t-1}^{1} - \mathbf{1}_{t-1} \mu \right) , 
\]

\[
u_t = y_t - \beta_0 - \beta' Y_{t-1}^{t-k} ,
\]

and \( R_{\Sigma_{t-1}} \) is a selector matrix which transforms \( \text{vec}p(\Sigma_{t-1}^{1})^{-1} \) into \( \text{vec}(\Sigma_{t-1}^{1})^{-1} \).\(^{10}\) Subscript \( T \) has been used on the left-hand side of (4.9) to emphasize the fact that functionally independent

---

\(^9\) Substitution of \( \mu (1 - \beta' \mathbf{1}_k) \) for \( \beta_0 \) was made in this calculation.

\(^{10}\) The vec (short for vec pattern) operator extracts functionally independent elements of a patterned matrix and arranges them in a column vector. The vec and vec operators are related via \( \text{vec}p X = P_X \text{vec}X \) and \( \text{vec}X = R_X \text{vec}p X \), where \( P_X \) and \( R_X \) are selector matrices. If the expression in question involves more than one patterned matrix, it may be unclear which pattern is relevant for selector matrices \( P \) and \( R \). To avoid confusion a subscript was added to indicate the relevant pattern. For further details see Henderson and Searle (1979).
elements in \((\Sigma_{t-1}^{-1})^{-1}\) are the same for all \(t>2k+1\). This fact will be explained in more detail in section 4.2.

As in the static case, equations (4.5.) - (4.9) are non-linear in unknown parameters and cannot be set equal to zero and solved analytically for their MLEs. To get an idea about the form of the implied MLEs, Spanos (1990), dealing with a similar case, solves the FOC for a fixed value of \(\gamma_t\) and \(\delta_t\). The resulting MLEs are of GLS type. He claims for them the property of consistency.

Since the MLEs cannot be derived analytically, in order to maximize (4.5), the system (4.6) - (4.9) has to be set equal to zero and solved iteratively using a numerical procedure. This gives rise to some additional difficulties. First, any numerical algorithm is likely to run into problems with the positive definiteness of \(\left(\hat{\Sigma}^{-1}_{t-1}\right)^{-1}\) and the positivity of \(\hat{\sigma}^2\). This issue has already been encountered in Chapter 2 and the same remedy can be applied in the present case.

The second problem has to do with the dimensionality of \(\left(\Sigma^{-1}_{t-1}\right)^{-1}\). In order to solve the first order conditions in a reasonable amount of time, \(\left(\Sigma_{t-1}^{-1}\right)^{-1}\) must fit into the computer’s memory. This limits the maximum allowable sample size to about 500 observations, using a hardware/software combination currently regarded as standard.\(^{11}\) This requirement is extremely restrictive given, for example, that financial data series are frequently much larger than 500 observations. A solution to this problem is described in the next section.

4.2. Operational Form of the Model.

Some of the problems outlined at the end of section 4.1 were already encountered (and solved) in Chapter 2. In particular, positivity of \(\hat{\sigma}^2\) was ensured by taking the relevant derivative with respect to \(\sigma\) rather than \(\sigma^2\). Later the same type of reasoning was applied to \(\Sigma^{-1}\), which was factorized into a product of two symmetric matrices.

\(^{11}\) GAUSS 386i running on a Pentium-133-MHz-based computer with 32 MB of RAM.
At first glance, the estimation of \( \left( \Sigma_{t-1}^1 \right)^{-1} \) poses a bigger problem than it did in Chapter 2, because of its dimensionality and the number of parameters involved. Even though the positive definiteness can be assured through the decomposition

\[
\left( \Sigma_{t-1}^1 \right)^{-1} = A_{t-1} A_{t-1}^\prime,
\]

the \( A_{t-1} \) matrices still have the same dimensions as \( \left( \Sigma_{t-1}^1 \right)^{-1} \). However, a closer look at \( \left( \Sigma_{t-1}^1 \right)^{-1} \) reveals a particular structure, which can be utilized to reduce the problem to manageable proportions. Under the assumed conditions (Markovness of order \( k \) in the mean and exponentially decaying covariances) \( \left( \Sigma_{t-1}^1 \right)^{-1} \) has the following structure (for clarity of illustration shown for \( k=3 \) and \( t-1=9 \))\(^{12}\)

\[
\left( \Sigma_{9}^1 \right)^{-1} = \begin{bmatrix}
  s_{11} & s_{21} & s_{31} & q_3 \\
  s_{21} & s_{22} & s_{21} & q_2 & q_3 \\
  s_{31} & s_{32} & s_{33} & q_1 & q_2 & q_3 \\
  q_3 & q_2 & q_1 & q_0 & q_1 & q_2 & q_3 \\
  q_3 & q_2 & q_1 & q_0 & q_1 & q_2 & q_3 \\
  q_3 & q_2 & q_1 & s_{33} & s_{32} & s_{31} \\
  q_3 & q_2 & s_{32} & s_{22} & s_{21} \\
  q_3 & s_{31} & s_{21} & s_{11}
\end{bmatrix},
\]

where the missing entries denote zeros, and the subscript on \( q \) refers to the difference \( |i-j| \), where \( i \) and \( j \) denote the row and column numbers respectively. As can be seen, the matrix has a symmetric band running along the main diagonal. The width of the band is equal to \( 2k+1 \), where \( k \) is the number of lags in the Markov mean of the stochastic process. The two areas in the North West and South East corners are both \( k \times k \) matrices which are mirror images of each other. They contain different elements than the rest of the matrix, but each one is symmetric along its main

\(^{12}\) See Spanos (1990) and references therein.
diagonal. Therefore, the total number of distinct elements in \((\Sigma_{t-1}^1)^{-1}\) equals \(\frac{(k+1)(k+2)}{2}\). The structure of \((\Sigma_{t-1}^1)^{-1}\) remains the same for all \(t>2k+1\). This explains why the decomposition in (4.1) stopped at \(t=2k+1\), and why \(\text{vecp}(\Sigma_{t-1}^1)^{-1}\) is identical for all \(t>2k+1\).

Another aspect of \((\Sigma_{t-1}^1)^{-1}\) deserving special attention are the implications of its structure for \(\sigma^2\). The best insight into this matter is provided by the following simple example.

Consider a stochastic process

\[
y_t = \beta y_{t-1} + u_t; \quad u_t \sim \text{St}(0, \omega^2, \nu); \quad t = 1, \ldots, T.
\]

The scale matrix for \(\Psi_t^1\) takes the form

\[
\Sigma_t^1 = \frac{\omega^2}{1 - \beta^2} \begin{bmatrix}
1 & \beta & \beta^2 & \cdots & \beta^{t-2} & \beta^{t-1} \\
\beta & 1 & \beta & \cdots & \beta^{t-3} & \beta^{t-2} \\
\beta^2 & \beta & 1 & \cdots & \beta^{t-4} & \beta^{t-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\beta^{t-2} & \beta^{t-3} & \beta^{t-4} & \cdots & 1 & \beta \\
\beta^{t-1} & \beta^{t-2} & \beta^{t-3} & \cdots & \beta & 1
\end{bmatrix}
\]

The inverse of \(\Sigma_t^1\) is given by

\[
(\Sigma_{t-1}^1)^{-1} = \frac{1}{\omega^2} \begin{bmatrix}
1 & -\beta \\
-\beta & 1 + \beta^2 & -\beta \\
-\beta & 1 + \beta^2 & -\beta & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\beta & 1 + \beta^2 & -\beta & \cdots & 1 \\
-\beta & 1 + \beta^2 & -\beta & \cdots & \beta \\
-\beta & 1 + \beta^2 & -\beta & \cdots & \beta & 1
\end{bmatrix}
\]

\[13 \frac{k(k+1)}{2}\] in the corner matrix and \(k+1\) in the rest of the diagonal band.

\[14\] Increasing the number of lags complicates the explanation without adding to the understanding.
Now, consider the implications of the above facts for
\[ \sigma_t^2 = \sigma_{t,t} - \sigma_{t-1,t} \left( \Sigma_{t-1}^{-1} \right)^{-1} \sigma_{t-1,t}. \]

For clarity of exposition let us first concentrate on the expression \( \left( \Sigma_{t-1}^{-1} \right)^{-1} \sigma_{t-1,t} \), and later apply the findings in the analysis of the full-blown \( \sigma_t^2 \).

\[
\left( \Sigma_{t-1}^{-1} \right)^{-1} \sigma_{t-1,t} = \frac{1}{\omega^2} \begin{bmatrix}
1 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta \\
-\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta \\
-\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta \\
-\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta & 1 + \beta^2 & -\beta \\
\end{bmatrix} \begin{bmatrix}
\beta^{-1} \\
\beta^{-2} \\
\beta^{-3} \\
\omega^2 \\
\omega^2 \\
1 - \beta^2 \\
1 - \beta^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\beta \\
\end{bmatrix}.
\]

The end result is hardly surprising, given that \( \left( \Sigma_{t-1}^{-1} \right)^{-1} \sigma_{t-1,t} \) is the formula for the true regression coefficients. Notice how the elements in the diagonal band in \( \left( \Sigma_{t-1}^{-1} \right)^{-1} \) interact with \( \sigma_{t-1,t} \) to wipe out all entries, except for the last one. (If the Markov mean of the stochastic process contained \( k \) lags, the diagonal band would be \( 2k+1 \)-elements wide, and the last \( k \) entries in the vector of regression coefficients would be different from zero. This result explains why the expression \( \beta Y_{t-1}^{t-k} \) in (4.3) contains only \( k \) lags of \( Y \).

Applying the above findings to \( \sigma_t^2 \) yields
\[
\sigma_t^2 = \frac{\omega^2}{1 - \beta^2} \left( 1 - \frac{\omega^2}{1 - \beta^2} \begin{bmatrix}
\beta^{-1} \\
\beta^{-2} \\
\beta^{-3} \\
\omega^2 \\
\omega^2 \\
1 - \beta^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\omega^2 \\
\end{bmatrix} = \frac{\omega^2}{1 - \beta^2} (1 - \beta^2) = \omega^2,
\]

61
which does not vary with $t$. This explains the claim made in section 4.1, that $\sigma^2$ is identical for all $t > 2k+1$.\(^{15}\)

After tying the loose ends from section 4.1 let us return to the main argument.

The Cholesky factorization of $(\Sigma_{t-1}^1)^{-1}$, again depicted for $k=3$ and $t-1=9$, takes the form

$\Sigma^1 = \frac{\omega^2}{1 - \beta_1^2 - \beta_2^2 - \beta_3^2 - \beta_4^2 + \beta_5^2}$

$= \frac{\omega^2}{(\Sigma)} \begin{bmatrix}
\sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{t-2} & \sigma_{t-1} \\
\sigma_1 & \sigma_0 & \sigma_1 & \cdots & \sigma_{t-3} & \sigma_{t-2} \\
\sigma_2 & \sigma_1 & \sigma_0 & \cdots & \sigma_{t-4} & \sigma_{t-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\sigma_{t-2} & \sigma_{t-3} & \sigma_{t-4} & \cdots & \sigma_0 & \sigma_1 \\
\sigma_{t-1} & \sigma_{t-2} & \sigma_{t-3} & \cdots & \sigma_{t-3} & \sigma_{t-2} \\
\end{bmatrix}$

and

$\left(\Sigma_{t-1}^1\right)^{-1} = \frac{1}{\omega^2} \begin{bmatrix}
1 & -\beta_1 & -\beta_2 \\
-\beta_1 & 1 + \beta_1^2 & \beta_2 - \beta_1 - \beta_2 \\
-\beta_2 & \beta_2 - \beta_1 & 1 + \beta_2^2 & \beta_3 - \beta_2 - \beta_1 - \beta_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\beta_t & \beta_2 - \beta_1 & 1 + \beta_t^2 & \beta_{t-1} - \beta_t - \beta_2 \\
\end{bmatrix}$

producing the same effect.

---

\(^{15}\) It can be verified that the same type of relationships hold for $k$ greater than 1. For example in case of 2 lags,
where the subscript convention is the same as the one for \( (\Sigma_{t-1}^{-1}) \). Notice that \( A_{t-1} \) also has a banded nature, but the band is \( k+1 \)-elements wide, and extends only in one direction from the main diagonal. Further, it should be noticed that \( A_{t-1} \) contains the same set of elements repeating itself over and over (with the exception of the lower right-hand-side (RHS) corner). Finally, notice that we can set the lower RHS corner equal to a \( k \times k \) identity matrix and still fully recover all elements of \( (\Sigma_{t-1}^{-1}) \). This is because the calculation

\[
(\Sigma_{t-1}^{-1})^{-*} = A_{t-1}^* A_{t-1}^* 
\]

(where \( A_{t-1}^* \) denotes the simplified \( A_{t-1} \) matrix) still produces the proper form of \( (\Sigma_{t-1}^{-1})^{-1} \), with the exception of a \( k \times k \) submatrix in the lower RHS corner. Since both corners are mirror images of each other, the original matrix can be reconstructed by rearranging the elements from the upper LHS corner.\(^{16}\)

\(^{16}\) Intuitively such a simplification should only lead to a slight reduction in efficiency because \( A_{t-1}^* \) uses less information than is available. Some idea whether the loss of efficiency is significant can be gained from the ability of the algorithm to converge.
At this stage the model is already operational for sample sizes less than 500. Applying the Cholesky factorization tackles the problem of positive definiteness and the structure of 
\[(\Sigma_{t-1}^1)^{-1}\] and \(A_{t-1}^*\) guarantees that the number of parameters is not unduly large. However, the problem of dimensionality still remains unresolved. The model involves matrices of the same dimension as \((\Sigma_{t-1}^1)^{-1}\), and, for samples containing more than 100 observations, the iteration process will be painfully slow.

To tackle that last obstacle a closer look has to be taken at the quadratic form

\[
\left(\mathbf{Y}_{t-1}^1 - \mathbf{1}_{t-1}\mu\right)' \mathbf{A}_{t-1}' \mathbf{A}_{t-1} \left(\mathbf{Y}_{t-1}^1 - \mathbf{1}_{t-1}\mu\right),
\]

where the star on \(A\) has been dropped since it causes no ambiguity. For simplicity restricting the attention to elements of \(\mathbf{Y}_{t-1}^1\) only, the RHS of the quadratic form looks like\(^{17}\)

\[
\begin{bmatrix}
    a_0 & a_1 & \ldots & a_k \\
    a_0 & a_1 & \ldots & a_k \\
    \vdots & \vdots & \ddots & \vdots \\
    a_0 & a_1 & \ldots & a_k \\
    1 & & & \ldots \\
    & & & \vdots \\
    & & & 1
\end{bmatrix}
\begin{bmatrix}
    \mathbf{Y}_1 \\
    \mathbf{Y}_2 \\
    \vdots \\
    \mathbf{Y}_{t-k-1} \\
    \mathbf{Y}_{t-k} \\
    \vdots \\
    \mathbf{Y}_{t-1}
\end{bmatrix}
\]

The essence of this system can be captured by

\(^{17}\) The argument carries over to the LHS when the matrices are primed.
Using this insight, and the partitioned matrix notation, the quadratic form can be rewritten as

\[
(Y_{t+1} - 1_{t+1} \mu)' A_{t-1} A_{t-1} (Y_{t+1} - 1_{t+1} \mu) = \left[a' (Y_{t+1} - 1_{t+1} \mu) \right]' \left[X_{t-1}' 0 \right] \left[X_{t-1} 0 \right] \left[a 0 \right] = a' X_{t-1}' X_{t-1} a + (Y_{t+1} - 1_{t+1} \mu)' (Y_{t+1} - 1_{t+1} \mu),
\]

where \( a = (a_0, a_1, \ldots, a_k)' \) and \( X_{t-1} = \begin{bmatrix} Y_{t-k-1}' & Y_{t-k}' & \cdots & Y_{t-1}' \end{bmatrix} - 1_{(t-k-1) \times (k+1)} \mu \).

Notice that \( X_{t-1} \) contains columns of lagged Y’s, which are arranged backwards, when compared to the traditional way of writing it. Notice also, that, in the new formulation, the order of X’s and a’s gets reversed, when compared to the order of A’s and Y’s in the original quadratic form.

Putting all the pieces together, the log-likelihood function can be written as

\[
\text{LLF} = \ln \left( \frac{\Gamma(\frac{V+T}{2})}{\left(2\pi\right)^{\frac{T}{2}} \Gamma(\frac{V+2}{2})} \right) - \frac{T - (2k + 1)}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=2k+2}^{T} \ln \left( 1 + \frac{1}{V} [a' X_{t-i}' X_{t-i} a + (Y_{t-i} - 1_{t-i} \mu)' (Y_{t-i} - 1_{t-i} \mu)] \right) - \frac{V + 1}{2} \ln \left( Y_{t+1} - \beta_0 - \beta Y_{t+1}' \right) \left( \sigma^2 \left( 1 + \frac{1}{V} [a' X_{t-i}' X_{t-i} a + (Y_{t-i} - 1_{t-i} \mu)' (Y_{t-i} - 1_{t-i} \mu)] \right) \right) \left( Y_{t+1} - \beta_0 - \beta Y_{t+1}' \right),
\]

where \( \beta_0 = \mu (1 - \beta' 1_k) \).
The first order conditions for this formulation are\(^\text{18}\)

\[
\frac{\partial \text{LLF}}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{T_i \delta_i} u_t \left( Y_{i-1}^{t-k} - 1_k \mu \right)',
\]

\[
\frac{\partial \text{LLF}}{\partial \mu} = \frac{1}{v} \sum_{t=2k+2}^{T} \frac{1}{T_i \delta_i} \left[ a' X_{i-1}^{t-k} (1_{(t-k)(k+1)}) a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' 1_k \right] + \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{T_i \delta_i} \left( 1 - \beta' 1_k \right) u_t \left( a' X_{i-1}^{t-k} (1_{(t-k)(k+1)}) a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' 1_k \right) \left( a' X_{i-1}^{t-k} (1_{(t-k)(k+1)}) a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' 1_k \right) \left( a' X_{i-1}^{t-k} (1_{(t-k)(k+1)}) a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' 1_k \right) \left( a' X_{i-1}^{t-k} (1_{(t-k)(k+1)}) a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' 1_k \right),
\]

\[
\frac{\partial \text{LLF}}{\partial \sigma} = -\frac{T - (2k + 1)}{\sigma^3} + \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{T_i \delta_i} u_t^2,
\]

\[
\frac{\partial \text{LLF}}{\partial a_0} = -\frac{1}{v} \sum_{t=2k+2}^{T} \frac{1}{T_i \delta_i} \left[ a_{0} (X_{i-1} X_{i-1})_{11} + a_{i,k} (X_{i-1} X_{i-1})_{21} \right] + \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{T_i \delta_i} u_t^2 \left[ a_{0} (X_{i-1} X_{i-1})_{11} + a_{i,k} (X_{i-1} X_{i-1})_{21} \right],
\]

\[
\frac{\partial \text{LLF}}{\partial a_{i,k}} = -\frac{1}{v} \sum_{t=2k+2}^{T} \frac{1}{T_i \delta_i} \left[ a_{0} (X_{i-1} X_{i-1})_{12} + a_{i,k} (X_{i-1} X_{i-1})_{22} \right] + \frac{1}{\sigma^2} \sum_{t=2k+2}^{T} \frac{v + t}{v} \frac{1}{T_i \delta_i} u_t^2 \left[ a_{0} (X_{i-1} X_{i-1})_{12} + a_{i,k} (X_{i-1} X_{i-1})_{22} \right],
\]

where

\[
\gamma_i = 1 + \frac{1}{\sigma^2} \left( Y_i - \beta_0 - \beta' Y_{i-1}^{t-k} \right) \left( \sigma^2 \left( 1 + \frac{1}{v} \left( a' X_{i-1}^{t-k} a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' \left( Y_{i-1}^{t-k} - 1_k \mu \right) \right) \right) \right)^{-1} \left( Y_i - \beta_0 - \beta' Y_{i-1}^{t-k} \right),
\]

\[
\delta_i = 1 + \frac{1}{\sigma^2} \left( a' X_{i-1}^{t-k} a + \left( Y_{i-1}^{t-k} - 1_k \mu \right)' \left( Y_{i-1}^{t-k} - 1_k \mu \right) \right),
\]

\[
a = (a_0, a_{i,k}')' \quad \text{with} \quad a_{i,k} = (a_1, \ldots, a_k)', \quad \text{and} \quad \left( X_{i-1} X_{i-1} \right) \text{has been partitioned conformably with vector} \ a, \quad \text{and the subscripts denote the partitions in an obvious notation.}
\]

\(^{18}\)The substitution of \( \mu(1 - \beta' 1_k) \) for \( \beta_0 \) has been made in this calculation. Additionally, it should be kept in mind that the formula for \( X_{i-1} \) includes \( \mu \).
Notice that by using this system of equations we will get ML estimates (and standard errors for) \( \mathbf{a} \), rather than the elements of \( \left( \hat{\Sigma}_{t-1}^1 \right)^{-1} \), which are the estimates of the parameters of interest. Invoking the invariance property of MLEs, the Maximum Likelihood estimate of \( \left( \Sigma_{t-1}^1 \right)^{-1} \) can be calculated as

\[
\left( \hat{\Sigma}_{t-1}^1 \right)^{-1} = \hat{\mathbf{A}}_{t-1} \hat{\mathbf{A}}_{t-1}',
\]

where \( \hat{\mathbf{A}}_{t-1} \) denotes an estimate of the simplified \( \mathbf{A} \) matrix (i.e. having a \( k \times k \) identity matrix in the lower RHS corner).

Standard errors for the elements of \( \left( \hat{\Sigma}_{t-1}^1 \right)^{-1} \) can be derived using the \( \delta \)-method. The equation for the variance-covariance matrix of the distinct elements of \( \left( \hat{\Sigma}_{t-1}^1 \right)^{-1} \) is given by\(^{19}\)

\[
\mathrm{vec} \left[ \mathrm{vecp} \left( \Sigma_{t-1}^1 \right)^{-1} \right] = \left[ \mathbf{P}_{\Sigma} \left( \mathbf{I}_m^m + \mathbf{K}_{m,m} \right) \left( \mathbf{I}_m \otimes \hat{\mathbf{A}}' \right) \mathbf{R}_A \right] \hat{\Delta} \left[ \mathbf{P}_{\Sigma} \left( \mathbf{I}_m^m + \mathbf{K}_{m,m} \right) \left( \mathbf{I}_m \otimes \hat{\mathbf{A}}' \right) \mathbf{R}_A \right]' ,
\]

where

\[ m = k+1 \ (\text{number of distinct elements in } \mathbf{A}), \]
\[ \mathbf{K}_{m,m} = \text{commutation matrix}, \]
\[ \mathbf{P}_{\Sigma} = \text{selector matrix transforming } \mathrm{vec} \left( \Sigma_{t-1}^1 \right)^{-1} \text{ into } \mathrm{vecp} \left( \Sigma_{t-1}^1 \right)^{-1}, \]
\[ \mathbf{R}_A = \text{selector matrix transforming } \mathrm{vecpA} \text{ into } \mathrm{vecA}, \]
\[ \hat{\Delta} = \text{estimated variance-covariance matrix for the elements of vector } \mathbf{a}. \]

---

\(^{19}\) First order \( \delta \)-method.

\(^{20}\) See Magnus and Neudecker (1988).
4.3. Alternative Forms of the Log-Likelihood Function.

The model can be simplified by estimating the joint distribution directly, possibly rewriting it, so as it involves explicitly all of the parameters of interest. Using the form of multivariate-t density given in Johnson and Kotz (1972), the joint density for all \( T \) observations can be written as

\[
D(Y_T^T) = \frac{\Gamma\left(\frac{v + T}{2}\right)}{(\pi v)^{\frac{T}{2}} \Gamma\left(\frac{v}{2}\right)} \left|\Sigma_T^1\right|^{-\frac{1}{2}} \left[1 + \frac{1}{v}(Y_T^1 - I_T^1 \mu) \left(\Sigma_T^1\right)^{-1}(Y_T^1 - I_T^1 \mu)\right]^{-\frac{v+T}{2}}.
\]

This joint distribution can be estimated in two basic ways.\(^{21}\)

The first approach is analogous to the static case and involves a successive decomposition of the scale matrix. In this method only \( \beta, \mu, \) and \( \sigma^2 \) are estimated directly. The scale matrix is then derived using the invariance property of Maximum Likelihood estimators.

The second approach involves direct estimation of \( \mu \) and \( \left(\Sigma_T^1\right)^{-1} \). The remaining parameters are derived later, again relying on the invariance property of the MLEs.

4.3.1. Direct Estimation of \( \beta \).

Applying the same decomposition as in the case of the static model, the joint density can be rewritten as\(^{22}\)

\[
D(Y_T^i) = \frac{\Gamma\left(\frac{v + T}{2}\right)}{(\pi v)^{\frac{T}{2}} \Gamma\left(\frac{v}{2}\right)} \left|\Sigma_T^1\right|^{-\frac{1}{2}} \left[1 + \frac{1}{v}(Y_T^1 - I_T^1 \mu) \left(\Sigma_T^1\right)^{-1}(Y_T^1 - I_T^1 \mu)\right]^{-\frac{v+T}{2}}.
\]

Repeating the operation \( T-k \) times results in

\(^{21}\) These are called “two basic ways” because they constitute the two extreme possibilities. A hybrid approach involving any combination of the two is also possible.

\[ D(y_i^t) = \frac{\Gamma\left(\frac{v+T}{2}\right)}{(\pi v)^{T/2}} \left(\sigma^2\right)^{-T/2} \left|\Sigma_k^i\right|^{1/2} \left[1 + \frac{1}{v}(y_k^i - 1_k \mu)^\prime \left(\Sigma_k^i\right)^{-1}(y_k^i - 1_k \mu) + \frac{1}{v \sigma^2} \sum_{t=k+1}^{T} (y_t - \beta_0 - \beta^\prime y_{t-1}^k)^2 \right]^{(v+T)/2}, \]

which gives the log-likelihood function

\[ \text{LLF} = \ln \left(\frac{\Gamma\left(\frac{v+T}{2}\right)}{(\pi v)^{T/2}} \left(\sigma^2\right)^{-T/2} \left|\Sigma_k^i\right|^{1/2} \right) - \frac{T-k}{2} \ln \sigma^2 + \frac{1}{2} \ln \left(\left|\Sigma_k^i\right|^{-1}\right) - \frac{v+T}{2} \ln \left[1 + \frac{1}{v}(y_k^i - 1_k \mu)^\prime \left(\Sigma_k^i\right)^{-1}(y_k^i - 1_k \mu) + \frac{1}{v \sigma^2} \sum_{t=k+1}^{T} (y_t - \beta_0 - \beta^\prime y_{t-1}^k)^2 \right]. \]  

This expression looks considerably simpler than the log-likelihood function implied by (4.4).

The first order conditions for (4.11) are given by

\[ \frac{\partial \text{LLF}}{\partial \beta} = \frac{v+T}{v} \frac{1}{\gamma \sigma^2} \sum_{t=k+1}^{T} u_t\left[y_{t-1}^k - 1_k \mu\right]^\prime, \]

\[ \frac{\partial \text{LLF}}{\partial \mu} = \frac{v+T}{v} \frac{1}{\gamma \sigma^2} \left[(y_k^i - 1_k \mu)^\prime \left(\Sigma_k^i\right)^{-1} 1_k + \frac{(1-\beta^\prime 1_k)}{\sigma^2} \sum_{t=k+1}^{T} u_t\right], \]

\[ \frac{\partial \text{LLF}}{\partial \sigma} = -\frac{T-k}{\sigma} + \frac{v+T}{v} \frac{1}{\gamma \sigma^3} \sum_{t=k+1}^{T} u_t^2, \]

where

\[ \gamma = 1 + \frac{1}{v}(y_k^i - 1_k \mu)^\prime \left(\Sigma_k^i\right)^{-1}(y_k^i - 1_k \mu) + \frac{1}{v \sigma^2} \sum_{t=k+1}^{T} (y_t - \beta_0 - \beta^\prime y_{t-1}^k)^2, \]

and the mystery of the missing equation for \(\left(\Sigma_k^i\right)^{-1}\) is explained below.

The above decomposition is equivalent to splitting the sample into two parts. The first part, whose size is equal to the number of rows of the inverted scale matrix in (4.11), is used to

---

23 Substitution of \(\mu(1-\beta^\prime 1_k)\) for \(\beta_0\) has been made in this calculation.
estimate the elements of the inverted scale matrix. The remaining part is used to estimate $\beta$. At this stage it should be noticed that the inverted scale matrix in (4.11) has $k$ rows. That means, that the first part of the sample has just $k$ observations. Since $(\Sigma_1^k)^{-1}$ involves $k$ distinct elements, the number of observations is clearly too small to provide a good estimate of the inverted scale matrix. However, the elements of $\Sigma_1^k$ are essentially just initial conditions, which, given the assumption of asymptotic independence, can be ignored. During the estimation process this is accomplished by setting $\Sigma_1^k$ equal to an identity matrix. Later, an indirect estimate of $\Sigma_1^k$ can be obtained utilizing the relationships

$$\beta = \Sigma_1^{-1}\sigma_{12},$$  \hspace{1cm} (4.12)

$$\sigma^2 = \sigma_{22} - \sigma_{21}\Sigma_1^{-1}\sigma_{12},$$  \hspace{1cm} (4.13)

where the elements on the RHS of (4.12) and (4.13) are defined as the partitions of

$$\Sigma_{k+1} = \begin{bmatrix}
\sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{k-1} & \sigma_k \\
\sigma_1 & \sigma_0 & \sigma_1 & \cdots & \sigma_{k-2} & \sigma_{k-1} \\
\sigma_2 & \sigma_1 & \sigma_0 & \cdots & \sigma_{k-3} & \sigma_{k-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\sigma_{k-1} & \sigma_{k-2} & \sigma_{k-3} & \cdots & \sigma_0 & \sigma_1 \\
\sigma_k & \sigma_{k-1} & \sigma_{k-2} & \cdots & \sigma_1 & \sigma_0
\end{bmatrix} = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}.$$

Once $\beta$ and $\sigma^2$ are known, one can solve (4.12) and (4.13) indirectly. The resulting estimation equations, for the cases involving 1 and 2 lags, are provided below:

- 1-lag case

$$\sigma_0 = \frac{\sigma^2}{1 - \beta^2},$$

$$\sigma_1 = \frac{\beta \sigma^2}{1 - \beta^2},$$

\footnote{It should be replaced with the proper matrix when the final value of the log-likelihood function is calculated.}
where \( \beta_i \) denotes the i-th element of the vector \( \beta \) counting from the bottom, and \( \sigma \) is the i-th off-diagonal element of the scale matrix (NOT its inverse). For higher order lags the same procedure should be followed, however, the solutions are more complicated and may require a symbolic mathematical program (like Mathematica).

4.3.2. **Direct estimation of \( (\Sigma_1^T)^{-1} \).**

At the other extreme of the range of simplified techniques is the direct estimation of \( (\Sigma_1^T)^{-1} \). Derivation of the relevant model again starts from the joint density, reproduced below for easy reference.

\[
D(Y_1^T) = \frac{\Gamma \left( \frac{v + T}{2} \right)}{(\pi v)^{\frac{T}{2}} \Gamma \left( \frac{v}{2} \right)} \left| (\Sigma_1^T)^{-1} \right|^{-\frac{1}{2}} \left[ 1 + \frac{1}{v} (Y_1^T - 1, \mu) (\Sigma_1^T)^{-1} (Y_1^T - 1, \mu) \right]^{-\frac{v + T}{2}}.
\]

One look at the expression inside the square brackets reveals the similarity to its counterpart in section 4.1. Thus, the techniques developed in section 4.2 are readily available to handle the difficulties involved in making that part operational.

---

25 The elements are counted from the bottom because in the vector \( Y \) the newer observations are added successively at the bottom, following the usual time-series convention.

26 \( \sigma_0 \) being the diagonal element.
One other term deserving special attention is the determinant of $\left(\Sigma_T^1\right)^{-1}$. In its current form, the dimensionality of $\left(\Sigma_T^1\right)^{-1}$ poses a problem. Applying the Cholesky factorization, the determinant can be transformed as follows

$$\left|\left(\Sigma_T^1\right)^{-1}\right| = |A' A| = |A||A| = |A|^2.$$  

At this stage it should be kept in mind that matrix $A$ possesses a particular structure. Its elements are arranged in a band extending in one direction from the main diagonal (see (4.10)). The band contains the same set of elements repeating itself over and over, with the exception of the lower RHS corner. Setting the lower RHS corner of $A$ equal to a $k \times k$ identity matrix reduces its determinant to a product of $T-k$ identical elements, namely $a_0$'s. Thus,

$$|A|^2 = a_0^{2(T-k)}.$$  

Applying the above findings, the log-likelihood function can be written as

$$\text{LLF} = \ln \left(\frac{\Gamma\left(\frac{\nu + T}{2}\right)}{(\pi \nu)^\frac{T}{2}}\right) + (T - k) \ln a_0 - \frac{\nu + T}{2} \ln \left\{ 1 + \frac{1}{\nu} \left[ a'X'Xa + \left( Y_{T}^{T-k+1} - 1_k \mu \right)' \left( Y_{T}^{T-k+1} - 1_k \mu \right) \right]\right\},$$

where the subscript in $X_T$ was dropped, because in the present case it causes no misunderstandings.

The first order conditions for this formulation are

$$\frac{\partial \text{LLF}}{\partial \mu} = \frac{\nu + T}{\nu} \gamma \left[ a'X'_{(T-k)\times(k+1)} a + \left( Y_{T}^{T-k+1} - 1_k \mu \right)' 1_k \right],$$

$$\frac{\partial \text{LLF}}{\partial a_0} = \frac{T - k}{a_0} - \frac{\nu + T}{\nu} \gamma \left[ a_0 (X'X)_{11} + a'_{1k} (X'X)_{21} \right],$$
\[
\frac{\partial \text{LLF}}{\partial \mathbf{a}_{1:k}} = -\frac{\mathbf{v} + \mathbf{T} \mathbf{I}}{\mathbf{v}} \gamma \left[ a_0 (\mathbf{X}^\prime \mathbf{X})_{12} + a_{1:k}^\prime (\mathbf{X}^\prime \mathbf{X})_{22} \right],
\]

where

\[
\gamma = 1 + \frac{1}{\mathbf{v}} \left[ \mathbf{a}^\prime \mathbf{X} \mathbf{X}^\prime \mathbf{a} + (\mathbf{Y}_T^{T-k+1} - \mathbf{I}_k \mathbf{\mu})^\prime (\mathbf{Y}_T^{T-k+1} - \mathbf{I}_k \mathbf{\mu}) \right],
\]

and \(a_0, a_{1:k}, \) and \(\mathbf{X}\) are defined analogously to the full-blown case.

When the estimation is carried out, the algorithm will provide standard errors for the vector \(\hat{\mathbf{a}}\). Standard errors for the elements of \(\left(\hat{\Sigma}_T^1\right)^{-1}\), which are the estimates of the parameters of interest, can again be derived using the \(\delta\)-method. The equation of the estimated variance-covariance matrix of the distinct elements of \(\left(\Sigma_T^1\right)^{-1}\) is identical to the full-blown case\(^{27}\)

\[
\mathbf{v} \mathbf{c} \left[ \text{vecp}(\mathbf{\Sigma}_T^1)^{-1} \right] = \left[ \mathbf{P}_{\Sigma} \left( \mathbf{I}_{m^2} + \mathbf{K}_{m,m} \right) \left( \mathbf{I}_{m} \otimes \hat{\mathbf{A}}^\prime \right) \mathbf{R} \right] \mathbf{\hat{\Delta}} \left[ \mathbf{P}_{\Sigma} \left( \mathbf{I}_{m^2} + \mathbf{K}_{m,m} \right) \left( \mathbf{I}_{m} \otimes \hat{\mathbf{A}}^\prime \right) \mathbf{R} \right]^\prime,
\]

where the notational convention is the same as in section 4.2.

When \(\left(\hat{\Sigma}_T^1\right)^{-1}\) is known, \(\hat{\mathbf{\beta}}\) can be calculated using the formula

\[
\hat{\mathbf{\beta}} = \left[ \left(\hat{\Sigma}_{2k+1}^1\right)^{-1} \hat{\sigma}_{2k+1,2k+2}^1 \right]_{k+2:2k+1},
\]

where \(k+2:2k+1\) denotes the bottom \(k\) elements of the relevant vector. The above calculation makes use of the fact that \(\left(\Sigma_{2k+1}^1\right)^{-1}\) is the smallest matrix with the same structure as \(\left(\Sigma_T^1\right)^{-1}\) (thus, it is known when \(\left(\Sigma_T^1\right)^{-1}\) is known). As was explained in section 4.2, the multiplication \(\left(\Sigma_{2k+1}^1\right)^{-1} \hat{\sigma}_{2k+1,2k+2}^1\) produces a \(2k+1\)-element vector with \(k+1\) zeros and the elements of \(\hat{\mathbf{\beta}}\) at the bottom. The estimate of \(\hat{\sigma}_{2k+1,2k+2}^1\) can be obtained by inverting \(\left(\hat{\Sigma}_{2k+2}^1\right)^{-1}\) (which, again, has the same structure as \(\left(\hat{\Sigma}_T^1\right)^{-1}\)).

\(^{27}\) First order \(\delta\)-method.
It can be verified that the first diagonal element in \( (\Sigma_T^{-1}) \) is equal to \( \frac{1}{\sigma^2} \), and the off-diagonal elements in the first row (or column) of \( (\Sigma_T^{-1}) \) are equal to the elements of \( \beta \) multiplied by \( \frac{1}{\sigma^2} \), with their sign reversed. Thus, when the inverted scale matrix is normalized, so that the first diagonal element equals 1, the off-diagonal elements in the first row are equal to the elements of \( \beta \) with the opposite sign. Furthermore, the off-diagonal elements in the first row of the Cholesky decomposition of this normalized matrix are also equal to the elements of \( \beta \) with their sign reversed. Thus, the asymptotic standard errors for the off-diagonal elements of \( A \), multiplied by the square root of the normalizing factor, can be used as the standard errors for the elements of \( \beta \).

4.3.3. The Hybrid Approach.

As was mentioned earlier, the models in sections 4.3.1 and 4.3.2 are located on the opposite ends of the range of possible treatments of the joint distribution. A number of intermediate approaches, falling somewhere between the two extremes, is possible as well. These models are lumped together in a category called the hybrid approach.

The hybrid model is based on the same principle as the model in section 4.3.1, but the decomposition of the scale matrix is not carried out till the end. Instead, it stops at some point \( t \), such that \( 2k+I < t < T \), with the midpoint being the natural choice. The log-likelihood function and the first order conditions are a straightforward combination of equations from sections 4.3.1 and 4.3.2, with a few trivial modifications. The model involves all of the parameters of interest and all relevant standard errors can be computed from the inverse of the final Hessian. The estimates are less precise than those from sections 4.3.1 and 4.3.2, because the sample is split into two equal parts. In principle the division point can be chosen anywhere between \( 2k+I \) and \( T \), enabling the estimation of many hybrid models based on different split points.

---

28 There is no special advantage in choosing the midpoint other than a natural appeal of splitting the sample into two equal parts. In principle the division point can be chosen anywhere between \( 2k+I \) and \( T \), enabling the estimation of many hybrid models based on different split points.

29 \( (\Sigma_T^{-1})^{-1} \) can be handled just like in section 4.3.2.
two halves. The first part is used to estimate \( (\Sigma_\tau^{-1}) \) and the second part is used to estimate \( \beta \).

4.4. The Case of the Markov Variance.

Consider the reduction

\[
D(Y^1_T) = D(Y^1_{T-1} | Y^1_{T-1}) \cdot D(Y^1_{T-2} | Y^1_{T-2}) \ldots D(Y^1_{k+1} | Y^1_k) \cdot D(Y^1_k) \\
= D(Y^1_k) \prod_{i=k+1}^{T} D(Y^1_i | Y^1_{i-1}).
\]

Assuming that all the past information can be summarized by the last \( k \) observations, this reduction can be written as

\[
D(Y^1_T) = D(Y^1_{T-k} \mid Y^1_{T-k-1}) \cdot D(Y^1_{T-k-2} \mid Y^1_{T-k-2}) \ldots D(Y^1_{k+1} \mid Y^1_k) \cdot D(Y^1_k) \\
= D(Y^1_k) \prod_{i=k+1}^{T} D(Y^1_{i-k} \mid Y^1_{i-k+1}),
\]

where the distribution of \( Y^1_T \) is no longer multivariate-t, but poly-t.\(^{30}\)

The “marginal” and conditional density functions take the form

\[
D(Y^1_k) = \frac{\Gamma(\frac{v+k}{2})}{\left(\pi \nu \right)^{\frac{k}{2}} \Gamma(\frac{v}{2})^{\frac{k}{2}}} \left| \Sigma_k \right|^{\frac{1}{2}} \left[ 1 + \frac{1}{v} \left( Y^1_k - \mu_k \right) \left( \Sigma_k \right)^{-1} \left( Y^1_k - \mu_k \right) \right]^{\frac{v+k}{2}},
\]

\[
D(Y^1_i \mid Y^1_{i-k}) = \frac{\Gamma(\frac{v+k+1}{2})}{\left(\pi \nu \right)^{\frac{k+1}{2}} \Gamma(\frac{v+k}{2})^{\frac{k+1}{2}}} \left| \sigma^2 \left[ 1 + \frac{1}{v} \left( Y^1_{i-k} - \mu_k \right) \left( \Sigma_k \right)^{-1} \left( Y^1_{i-k} - \mu_k \right) \right] \right|^{\frac{1}{2}} \times \left[ 1 + \frac{1}{v} \left( Y^1_i - \beta_0 - \beta^\prime Y^1_{i-k} \right) \right]^{\frac{1}{2}} \cdot \sigma^2 \left[ 1 + \frac{1}{v} \left( Y^1_{i-k} - \mu_k \right) \left( \Sigma_k \right)^{-1} \left( Y^1_{i-k} - \mu_k \right) \right]^{-1} \left( Y^1_i - \beta_0 - \beta^\prime Y^1_{i-k} \right) \left( Y^1_i - \beta_0 - \beta^\prime Y^1_{i-k} \right)^{\frac{v+k+1}{2}}.
\]

\(^{30}\) See Drèze (1977) and references therein.
Ignoring the initial conditions, the log-likelihood function can be written as

$$LLF = (T - k) \ln \left( \frac{v + k + 1}{(\pi v)^{\frac{v + k}{2}}} \right) - \frac{T - k}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=k+1}^{T} \ln \left[ 1 + \frac{1}{v} \left( \frac{1}{(Y_{i-1}^{t-k} - I_k \mu)'A'A(Y_{i-1}^{t-k} - I_k \mu)} \right) \right] -$$

$$- \frac{v + k + 1}{2} \sum_{i=k+1}^{T} \ln \left[ 1 + \frac{1}{v} \left( Y_{i-1} - \beta_0 - \beta Y_{i-1}^{t-k} \right)' \left( \sigma^2 \left[ 1 + \frac{1}{v} \left( Y_{i-1}^{t-k} - I_k \mu \right)'A'A(Y_{i-1}^{t-k} - I_k \mu) \right] \right)^{-1} \left( Y_{i-1} - \beta_0 - \beta Y_{i-1}^{t-k} \right) \right].$$

where \( (\Sigma_k)^{-1} \) has been replaced by \( A'A \) for obvious reasons, \( A \) being a symmetric matrix.

First order conditions are similar to those for the full-blown STAR model. 

$$\frac{\partial LLF}{\partial \beta} = \frac{v + k + 1}{v} \frac{1}{\sigma^2} \sum_{i=k+1}^{T} \frac{1}{\gamma_i} \frac{1}{\delta_i} u_i (Y_{i-1}^{t-k} - I_k \mu)' ,$$

$$\frac{\partial LLF}{\partial \mu} = \frac{1}{v} \sum_{i=k+1}^{T} \frac{1}{\gamma_i} \frac{1}{\delta_i} (Y_{i-1}^{t-k} - I_k \mu)' A'A_1_k +$$

$$+ \frac{v + k + 1}{v} \frac{1}{\sigma^2} \sum_{i=k+1}^{T} \frac{1}{\gamma_i} \frac{1}{\delta_i} \left[ (1 - \beta_0) - \frac{1}{v} \frac{1}{\delta_i} (Y_{i-1}^{t-k} - I_k \mu)' A'A_1_k \right] ,$$

$$\frac{\partial LLF}{\partial \sigma} = - \frac{T - k}{\sigma} + \frac{v + k + 1}{v} \frac{1}{\sigma^2} \sum_{i=k+1}^{T} \frac{1}{\gamma_i} \frac{1}{\delta_i} u_i^2 ,$$

$$\frac{\partial LLF}{\partial \text{vech}A} = - \frac{1}{v} \sum_{i=k+1}^{T} \frac{1}{\delta_i} \left[ (Y_{i-1}^{t-k} - I_k \mu)' \otimes \left( Y_{i-1}^{t-k} - I_k \mu \right)' A' \right] G +$$

$$+ \frac{v + k + 1}{v^2} \frac{1}{\sigma^2} \sum_{i=k+1}^{T} \frac{1}{\gamma_i} \frac{1}{\delta_i} u_i^2 \left[ (Y_{i-1}^{t-k} - I_k \mu)' \otimes \left( Y_{i-1}^{t-k} - I_k \mu \right)' A' \right] G ,$$

where \( G \) is a selector matrix transforming \( \text{vech}A \) into \( \text{vec}A \). The above FOC can be set equal to zero, and solved numerically for ML estimates of the parameters of interest. Thus, the model is already operational.

---

31 The initial conditions can be ignored because of finite memory assumption.
32 Substitution of \( \mu (1 - \beta_0 1_k) \) for \( \beta_0 \) has been made in this calculation.
4.5. The Adapted Static Model.

Consider again the RHS of (4.14), this time multiplying all the conditional densities by the part on which they are conditioned

\[
D(Y_{k+1}|Y^1_k) D(Y^1_k) = D(Y^1_{k+1}),
\]
\[
D(Y_{k+2}|Y^2_{k+1}) D(Y^2_{k+1}) = D(Y^2_{k+2}),
\]
\[
\vdots
\]
\[
D(Y_T|Y^{T-k}_{T-1}) D(Y^{T-k}_{T-1}) = D(Y^{T-k}_T),
\]

which can be compactly written as

\[
\prod_{t=k+1}^{T} D(Y_t|Y^{t-k}_{t-1}) D(Y^{t-k}_{t-1}) = \prod_{t=k+1}^{T} D(Y^{t-k}_t).
\]

The above form ignores all the information beyond lag \(k\). In this respect it is akin to the Markov formulation. However, it looks just like the decomposition in (2.1) performed in the derivation of the static model. The log-likelihood function and the first order conditions for this formulation are identical to those in Chapter 2. Therefore, this model can be estimated using the static algorithm. However, the value of the log-likelihood function will not be comparable to its counterpart from the full-blown STAR model. This occurs because the adapted static formulation is essentially a product of \(T-k\) joint distributions of \(k+l\) variables.
Chapter 5

Monte Carlo Examination of Dynamic Student $t$ Regression Models

In the previous chapter the scope of the analysis was broadened to include variables which exhibit correlation over time, and a number of dynamic Student $t$ models were derived. At this stage it is logical to ask, how well can these models describe the empirical data, and whether the findings from Chapter 3 generalize to the dynamic case. To investigate these issues the following Monte Carlo experiment was performed.

5.1. Design of the Monte Carlo Experiment.

The Monte Carlo experiment was performed with two objectives in mind. First, it was intended to illustrate the relative performance of different Dynamic Student $t$ Regression Models (DSTRMs) derived in the previous chapter. In this context the examination of the following issues was considered particularly important:

1. The ability of the MLEs to estimate the underlying parameters accurately.
2. The ability of the estimated standard errors to describe correctly the dispersion of the estimates.
3. The ability of the DSTRMs to model the conditional variance accurately.

The second objective of the Monte Carlo experiment was a comparison of the DSTRMs with the traditional Dynamic Linear Regression Model (DLRM). This comparison was motivated by the widespread practice of estimating the DLRM for data exhibiting dynamic heteroskedasticity. The experiment was designed to examine, how much improvement (if any) can be gained by utilizing all the information in the data. Since some researchers recognize the presence of heteroskedasticity in the data and rely on White’s standard errors (HCSEs) for testing, the ability of HCSEs to describe correctly the dispersion of the estimates was also examined.

The following models were selected to be studied in detail:
− the Student $t$ Autoregressive Model specified in terms of $\beta$’s (STAR-B) from section 4.3.1;
− the Student $t$ Autoregressive Model specified in terms of the scale matrix (STAR-S) from section 4.3.2;
− the Adapted Static Student $t$ Model (ASSTM) from section 4.5.

STAR-B and STAR-S were chosen because they are located on the opposite ends of the spectrum of possible approaches to the estimation of the STAR model (see section 4.3.3). ASSTM was chosen because it is akin to Markovness, in the sense that it ignores all information in the data located beyond the cutoff lag, and it can be estimated using the static algorithm. As was already mentioned, DLRM with both regular standard errors and HCSEs was also included for comparison.

The true structure of the data was given by

$$Y^t_i \sim St_t(1_i \mu, \Sigma^t_1, \nu), \quad t = 1, \ldots, T, \quad (5.1)$$

where $Y^t_i = (Y_1, Y_2, \ldots, Y_T)'$, $\mu = 0$, $\Sigma^t_1$ is such that

$$\Sigma^t_1 = \begin{bmatrix}
1 & -0.4 & -0.2 \\
-0.4 & 1.16 & -0.32 & -0.2 \\
-0.2 & -0.32 & 1.20 & -0.32 & -0.2 \\
 & & & & \\
& & & & \\
\end{bmatrix},$$

$$\left(\Sigma^t_1\right)^{-1} = \begin{bmatrix}
1 & -0.4 & -0.2 \\
-0.4 & 1.16 & -0.32 & -0.2 \\
-0.2 & -0.32 & 1.20 & -0.32 & -0.2 \\
 & & & & \\
& & & & \\
\end{bmatrix}^{-1}, \quad (5.2)$$

$^1$ The pattern in

$$\Sigma^t_1 = \frac{1}{0.576} \begin{bmatrix}
0.8 & 0.4 & 0.32 & 0.208 & 0.1472 & \cdots \\
0.4 & 0.8 & 0.4 & 0.32 & 0.208 & \cdots \\
0.32 & 0.4 & 0.8 & 0.4 & 0.32 & \cdots \\
0.208 & 0.32 & 0.4 & 0.8 & 0.4 & \cdots \\
0.1472 & 0.208 & 0.32 & 0.4 & 0.8 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}$$

is less intuitive than in $\left(\Sigma^t_1\right)^{-1}$. For an explanation of the structure of $\Sigma^t_1$ see section 4.2.
ν denotes the degrees of freedom, and the subscript/superscript convention for Σ is the same as in Chapter 4. (5.2) together with μ = 0 imply the following values for the remaining parameters:

\[ \beta = \begin{bmatrix} \beta_2 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}, \quad \sigma^2 = 1, \quad \beta_0 = 0, \]

and, for the ASSTM only,

\[ \Sigma^{-1} = \begin{bmatrix} 0.96 & -0.48 \\ -0.48 & 0.96 \end{bmatrix}. \]

In Chapter 2 it was claimed that the MLE is a Best Asymptotically Normal (BAN) estimator. Therefore, its performance in smaller and bigger samples is of interest. Additionally, when the DSTRM is estimated, the asymptotic standard errors for the parameter estimates are calculated from the inverse of the final Hessian. To gain insight into the accuracy of the inferences drawn using these asymptotic results, it is essential to examine how well the estimated standard errors approximate the true standard errors for different sample sizes. Sample sizes of 50, 100, 200, and 500 were chosen to study the above issues. The maximum number of observations was restricted to 500 because of hardware limitations (see section 5.2).

Another interesting question is how the performance of the DSTRMs varies with the degrees of freedom. The performance when the value of ν is low is most interesting, since many financial data series seem to possess this characteristic. Unfortunately, it also entails more difficulties in generating the data (see section 3.2). The generation is much easier for values of ν greater than 7, but then the distribution looks very much like normal. As a compromise the following cases were chosen to be studied: ν = 4, 6, and 8.

---

2 See section 4.1 and 4.2 for an explanation of the relationship between the primary parameters of the distribution and the values of the parameters of interest.

3 The elements in Σ^{-1} are different from the corresponding elements of (Σ^T)^{-1}, because the dimensions of Σ^{-1} are too small to exhibit the same general structure as (Σ^T)^{-1}. A matrix has to be at least 2k+1×2k+1 to have this structure. In the present case k=2, which means that the matrix would have to be at least 5×5. Σ^{-1} is only 2×2. It can be verified that σ^2 and β are the same as in the full-blown case. See section 4.2 for more details.


5 The similarity turns out to be only superficial, since normally distributed variables do not exhibit the second order dependence, as measured by the McLeod-Li test (see McLeod and Li (1983)).
For each combination of the sample size and $\nu$, 500 samples of data were generated. The number of repetitions was set at 500 to ensure comparability with the static case. Some preliminary runs were conducted using 1000 repetitions. The empirical distributions from these preliminary runs looked similar to the 500-repetition distributions.

For each sample the STAR-B, STAR-S, and ASSTM were estimated using Maximum Likelihood. For comparison purposes the DLRM was also estimated using OLS. The following information was archived for further analysis:

- Maximum Likelihood estimation:
  - estimates of the parameters;
  - asymptotic standard errors computed from the inverse of the final Hessian;
  - asymptotic standard errors for the elements of $\hat{\Sigma}_2^{-1}$ (ASSTM) and $(\hat{\Sigma}_T)^{-1}$ (STAR-S) calculated using the first order $\delta$-method;
  - elements of the product $\hat{\sigma}^2 \hat{\Sigma}_2^{-1}$ (ASSTM) and $\hat{\sigma}^2 (\hat{\Sigma}_T)^{-1}$ (STAR-B and STAR-S);
  - standard errors for the elements of $\hat{\sigma}^2 \hat{\Sigma}_2^{-1}$ (ASSTM) and $\hat{\sigma}^2 (\hat{\Sigma}_T)^{-1}$ (STAR-S) calculated using the first order $\delta$-method;
  - final value of the log-likelihood function;
  - elements of $\hat{A}_{22}$ (ASSTM) and $\hat{\Lambda}_T$ (STAR-S);

- OLS estimation:
  - estimates of the parameters (regression coefficients and the constant term);
  - standard errors of the estimated parameters;
  - HCSEs;
  - $\hat{\sigma}^2$.

The analysis of the results is presented in section 5.3.
5.2. Data Generation.

The data for the experiment were generated as follows. First, a raw series of Student \( t \) random numbers with mean 0, variance 1, and \( \nu \) degrees of freedom was generated. Next, the dependence structure was imposed on raw data through

\[
Y_t^\perp = \text{chol} \left( \frac{\nu}{\nu - 2} \Sigma_T^1 \right) J,
\]

where \( J \) is the series of raw Student \( t \) numbers, \( \Sigma_T^1 \) is the scale matrix from (5.1), and \( \text{chol}(\cdot) \) refers to the Cholesky factorization. The series of raw Student \( t \) numbers was generated using the procedure discussed in section 3.2.

The above method limits the maximum sample size to about 500 observations. When bigger samples are generated, the computer runs out of RAM, and starts using the hard drive as virtual memory. This slows the execution of the program considerably.\(^6\)

The above technique generates the joint Student \( t \) distribution directly. An alternative approach to data generation, which utilizes the conditional distribution, was also explored. In this approach the first \( 2k+1 \) raw Student \( t \) numbers are left unchanged.\(^7\) They constitute the initial conditions. Subsequent observations are generated recursively as

\[
Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t, \quad u_t = \sqrt{\omega_t^2} v_t,
\]

where

\[
\omega_t^2 = \frac{\nu}{\nu + (t-1) - 2} \sigma^2 \left[ 1 + \frac{1}{\nu} \left( Y_{t-1}^1 - \mu \right) \left( \Sigma_{t-1}^1 \right)^{-1} \left( Y_{t-1}^1 - \mu \right) \right],
\]

\(^6\) 500 observations is the maximum size that a computer equipped with 32 MB of RAM can handle without resorting to temporary storage on hard drive. The sample size can be increased to about 700 when the computer has 64 MB of RAM.

\(^7\) \( k \) being the number of lags in the Markov mean.
and \( \{v_t, t=1, \ldots, T\} \) is a series of IID Student \( t \) numbers with mean 0, variance 1, and \( v \) degrees of freedom. The series generated this way did not show second order dependence when tested with McLeod - Li test.\(^8\) Therefore, this technique was not used in the experiment.

5.3. Results.

Maximum Likelihood estimation was carried out in the same manner as in the static case (see section 3.3). The results are reported below. Following the same logic as in section 3.4, the attention is restricted to \( \beta, \beta_0, \sigma^2, \Sigma_{22}^{-1}, \) and \( \left(\Sigma_1^{-1}\right)^{-1} \). Since this time three separate models were estimated by Maximum Likelihood, the following convention is used to distinguish between the ML estimators: ML-A - the ML estimator based on ASSTM, ML-B - the ML estimator based on STAR-B, and ML-S - the ML estimator based on STAR-S. The OLS estimator is denoted as OLSE.

5.3.1. Estimates of \( \beta_1 \) and \( \beta_2 \).

Kernel density graphs of the empirical distribution of \( \hat{\beta}_1 \), estimated by all four methods, are shown in Figure 5.1 and Figure 5.2, for \( v=6 \) and various sample sizes.\(^9\) The graphs of the empirical distributions for \( v=4 \) and \( v=8 \) can be found in Appendix 4. Additionally, graphs of the distribution of the estimated standard errors for \( v=6 \), and sample sizes of 50 and 500, are shown in Figure 5.3. The dashed line in the pictures represents the contour of the normal density with the same mean and variance as the data whose kernel density is shown in the graph. Descriptive statistics for the empirical distributions of estimates are reported in Table 5.1 and Table 5.2. The tables also contain the means of the estimated standard errors. Descriptive statistics for the empirical distributions of the estimated standard errors can be found in Appendix 5.

Graphs in Figure 5.1 show that the empirical distribution of \( \hat{\beta}_1 \), estimated using the ML-A, is close to normal. The distribution appears symmetric, and its variance decreases as the

---

\(^8\) See McLeod and Li (1983).
\(^9\) Normal kernel density with bandwidth equal to \( 1.06\sigma T^{-0.2} \), where \( \sigma \) is the standard deviation, and \( T \) is the number of observations. See Silverman (1986).
sample size gets bigger. (Notice the change of scale in subsequent pictures). There are no major differences between the graphs of the empirical distribution of estimates for different values of \( \nu \). Descriptive statistics reported in Table 5.1 indicate possible problems with the mean of the distribution. For \( \nu = 4 \), the mean estimate exhibits erratic behavior and does not converge to the true \( \hat{\beta}_1 (=0.4) \). For \( \nu = 6 \), it diverges away from the true value. Finally, for \( \nu = 8 \), it moves towards the true value at first, but then it overshoots the mark. For the sample size of 500, the difference between the mean estimate and the true value of \( \beta_1 \) is significant at 5% level (\( t \)-statistics 13.48, 8.23, and 3.45 for \( \nu = 4, 6, \) and 8 respectively). The distribution of the estimates of \( \beta_2 \) exhibits the same problem for \( \nu = 4 \) (see Table 5.1). Thus \( \hat{\beta}_{ML-A} \) may be a biased and inconsistent estimator of \( \beta \). The graphs in Figure 5.3 indicate that the empirical distribution of the estimated standard errors is slightly skewed, but its mean correctly approximates the empirical standard deviation (see Table 5.1 and Appendix 5).

The graphs of the empirical distribution of \( \hat{\beta}_1 \), estimated by ML-B and ML-S, are similar to each other (see Figure 5.2). In both cases the mode of the distribution can be seen converging to the true value of \( \beta_1 \) and the variance gets smaller as the sample size increases. Again there are no major differences between the graphs showing the distribution for different values of \( \nu \). Descriptive statistics in Table 5.1 indicate that the means of the two distributions are always close to each other. In both cases the convergence of the mean to the true value of \( \beta_1 \) is evident, even though the difference between the mean estimate and the true \( \beta_1 \) is significant at 5% level, for \( \nu = 4 \) and the sample size of 500 (\( t \)-statistic = -2.85 for ML-B and -2.61 for ML-S). The convergence is also evident for \( \beta_2 \), despite similar problems. Thus, both \( \hat{\beta}_{ML-B} \) and \( \hat{\beta}_{ML-S} \) appear to be consistent estimators of \( \beta \).

---

10 The mean of the superimposed normal graph is equal to the mean of the data whose distribution is shown in the graph.

11 It is not significant for \( \nu = 6 \) (\( t \)-statistic = -1.18 for ML-B and -1.24 for ML-S) and \( \nu = 8 \) (\( t \)-statistic = -1.5 for ML-B and -1.34 for ML-S).
The distribution of the standard errors for both ML-B and ML-S is skewed (see Figure 5.3 and Appendix 5). Despite this fact, ML-B standard errors constitute a reasonably good approximation of the empirical standard deviation. In the case of ML-S, the mean of the estimated standard errors provides a good approximation of the empirical standard deviation for \( \beta_2 \), and, in smaller samples, for \( \beta_1 \). In bigger samples, both the mean and the mode of the empirical distribution of the estimated standard errors for \( (\hat{\beta}_1)_{ML-S} \) tend to overstate the true variability.

12 Standard errors for \( \hat{\beta}_{ML-B} \) were calculated by Maxlik automatically (from the inverse of the final Hessian). Standard errors for \( \hat{\beta}_{ML-S} \) were obtained using the method outlined in section 4.3.2. No adjustment was made for the normalizing factor (a little bit of outside knowledge was used in this case, i.e. that \( \sigma^2 = 1 \)).

13 The abnormally high skewness reported for the sample size of 50 (\( \nu = 6 \)) is largely due to the presence of a single extreme outlier.
Figure 5.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=6$. 
ML-B estimates of $\hat{\beta}_1$, $\nu=6$ (STAR-B)

ML-S estimates of $\hat{\beta}_1$, $\nu=6$ (STAR-S)

Figure 5.2. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=6$. 
Table 5.1. Selected descriptive statistics of the empirical distributions of ML estimates of $\beta_1$ and $\beta_2$.

<table>
<thead>
<tr>
<th>Parm. Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>ML-A (ASSTM)</th>
<th>ML-B (STAR-B)</th>
<th>ML-S (STAR-S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>ν=6</td>
<td>50</td>
<td>0.4080</td>
<td>0.1571</td>
<td>-0.0031</td>
</tr>
<tr>
<td></td>
<td>ν=6</td>
<td>100</td>
<td>0.4027</td>
<td>0.0960</td>
<td>0.0245</td>
</tr>
<tr>
<td></td>
<td>ν=6</td>
<td>200</td>
<td>0.4121</td>
<td>0.0665</td>
<td>0.1440</td>
</tr>
<tr>
<td></td>
<td>ν=6</td>
<td>500</td>
<td>0.4162</td>
<td>0.0440</td>
<td>0.0774</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>ν=4</td>
<td>50</td>
<td>0.4315</td>
<td>0.1608</td>
<td>0.0825</td>
</tr>
<tr>
<td></td>
<td>ν=4</td>
<td>100</td>
<td>0.4213</td>
<td>0.1606</td>
<td>0.1069</td>
</tr>
<tr>
<td></td>
<td>ν=4</td>
<td>200</td>
<td>0.4333</td>
<td>0.0712</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>ν=4</td>
<td>500</td>
<td>0.4293</td>
<td>0.0486</td>
<td>-0.0772</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>ν=8</td>
<td>50</td>
<td>0.4080</td>
<td>0.1571</td>
<td>-0.0031</td>
</tr>
<tr>
<td></td>
<td>ν=8</td>
<td>100</td>
<td>0.3972</td>
<td>0.0992</td>
<td>0.0718</td>
</tr>
<tr>
<td></td>
<td>ν=8</td>
<td>200</td>
<td>0.4121</td>
<td>0.0665</td>
<td>0.1440</td>
</tr>
<tr>
<td></td>
<td>ν=8</td>
<td>500</td>
<td>0.4162</td>
<td>0.0440</td>
<td>0.0774</td>
</tr>
</tbody>
</table>

Degrees of freedom.
** Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
*** Empirical standard deviation.
**** Mean estimated standard error.
Table 5.2. Selected descriptive statistics of the empirical distributions of OLS estimates of $\beta_1$ and $\beta_2$.

<table>
<thead>
<tr>
<th>Parm. Name</th>
<th>DF</th>
<th>Sample Size*</th>
<th>Mean Est.</th>
<th>Emp. SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Mean SE</th>
<th>Mean HCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>0.3512</td>
<td>0.1372</td>
<td>-0.1825</td>
<td>3.1801</td>
<td>0.1425</td>
<td>0.1425</td>
<td>0.1795</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.3732</td>
<td>0.0992</td>
<td>0.1391</td>
<td>3.1300</td>
<td>0.1391</td>
<td>0.0993</td>
<td>0.1302</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.3952</td>
<td>0.0709</td>
<td>0.0363</td>
<td>3.3185</td>
<td>0.0698</td>
<td>0.0698</td>
<td>0.0931</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.3940</td>
<td>0.0461</td>
<td>0.0773</td>
<td>3.1052</td>
<td>0.0439</td>
<td>0.0439</td>
<td>0.0600</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.3685</td>
<td>0.1502</td>
<td>-0.174</td>
<td>4.3764</td>
<td>0.1428</td>
<td>0.1428</td>
<td>0.1628</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.3721</td>
<td>0.0976</td>
<td>0.0924</td>
<td>2.9243</td>
<td>0.0994</td>
<td>0.0994</td>
<td>0.1188</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.3901</td>
<td>0.0687</td>
<td>0.0688</td>
<td>2.8558</td>
<td>0.0698</td>
<td>0.0698</td>
<td>0.0842</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.3974</td>
<td>0.0432</td>
<td>0.0279</td>
<td>2.9207</td>
<td>0.0440</td>
<td>0.0440</td>
<td>0.0538</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td></td>
<td>$\nu=6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.3510</td>
<td>0.1366</td>
<td>-0.2404</td>
<td>3.5872</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1479</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.3832</td>
<td>0.0978</td>
<td>0.0381</td>
<td>3.1047</td>
<td>0.0995</td>
<td>0.0995</td>
<td>0.1069</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.3886</td>
<td>0.0700</td>
<td>0.0581</td>
<td>2.9462</td>
<td>0.0698</td>
<td>0.0698</td>
<td>0.0764</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.3967</td>
<td>0.0432</td>
<td>0.0010</td>
<td>2.9977</td>
<td>0.0440</td>
<td>0.0440</td>
<td>0.0488</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.1516</td>
<td>0.1382</td>
<td>-0.2713</td>
<td>2.8790</td>
<td>0.1418</td>
<td>0.1418</td>
<td>0.1353</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1775</td>
<td>0.0999</td>
<td>-0.2785</td>
<td>2.9460</td>
<td>0.0991</td>
<td>0.0991</td>
<td>0.1024</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1872</td>
<td>0.0698</td>
<td>-0.0468</td>
<td>3.0942</td>
<td>0.0697</td>
<td>0.0697</td>
<td>0.0750</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.1944</td>
<td>0.0451</td>
<td>-0.2281</td>
<td>2.9415</td>
<td>0.0439</td>
<td>0.0439</td>
<td>0.0482</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td></td>
<td>$\nu=4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.1383</td>
<td>0.1386</td>
<td>-0.1846</td>
<td>3.0458</td>
<td>0.1417</td>
<td>0.1417</td>
<td>0.1335</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1696</td>
<td>0.0973</td>
<td>-0.2844</td>
<td>2.9042</td>
<td>0.0991</td>
<td>0.0991</td>
<td>0.1010</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1901</td>
<td>0.0683</td>
<td>-0.1082</td>
<td>2.9700</td>
<td>0.0697</td>
<td>0.0697</td>
<td>0.0729</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.1925</td>
<td>0.0451</td>
<td>0.0029</td>
<td>2.9112</td>
<td>0.0439</td>
<td>0.0439</td>
<td>0.0467</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.1424</td>
<td>0.1382</td>
<td>-0.2063</td>
<td>2.6868</td>
<td>0.1421</td>
<td>0.1421</td>
<td>0.1312</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1715</td>
<td>0.0930</td>
<td>-0.2199</td>
<td>3.0187</td>
<td>0.0991</td>
<td>0.0991</td>
<td>0.0982</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1847</td>
<td>0.0701</td>
<td>-0.1398</td>
<td>3.0736</td>
<td>0.0697</td>
<td>0.0697</td>
<td>0.0704</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.1921</td>
<td>0.0451</td>
<td>-0.0967</td>
<td>2.8154</td>
<td>0.0439</td>
<td>0.0439</td>
<td>0.0450</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

The graphs of the empirical distribution of the OLS estimates of $\beta_1$ appear close to normal (see Figure 5.1). The mode of the distribution converges to the true value, although some erratic behavior can be observed for $\nu=4$. The variance of the distribution decreases as the sample size gets bigger. Descriptive statistics in Table 5.2 show that the mean of the empirical distribution of estimates converges to the true value for both $\beta_1$ and $\beta_2$. Again, for the sample size of 500, the difference between the mean estimate and the true value is significant at 5% level for $\nu=4$ in case of $\beta_1$, and for all values of $\nu$ in case of $\beta_2$. Despite this fact, convergence is apparent by looking at the numbers in Table 5.2. Thus, the OLS estimator of $\beta$ appears to be consistent.

The empirical distribution of the regular OLS standard errors is skewed (see Figure 5.3 and Appendix 5), but, despite that, its mode and mean are close to the empirical standard deviation. The distribution of White’s standard errors for $\hat{\beta}_1$ appears normal, but for $\hat{\beta}_2$ it is skewed. In either case its mean tracks the empirical standard deviation well.
Figure 5.3. Kernel density estimates of the empirical distribution of standard errors for $\hat{\beta}_{1}, \nu=6$. 
A four-way comparison of the above models divides them into two groups. The first group consists of STAR-B, STAR-S, and regular DLRM. The MLEs implied by the two STAR models seem consistent, as does the OLSE. For all three estimators, the graphs of the empirical distribution of estimates show increasing degree of similarity as the sample size gets bigger. For the sample size of 500 the graphs are virtually identical. The mean estimates from all three models are very close, and so are the empirical standard deviations. The mean of the estimated standard errors is also quite similar, despite the varying degree of skewness of their distributions (with the exception of the standard errors of \((\hat{\beta}_1)_{\text{ML-S}}\), in bigger samples).

The second group consists of the ASSTM only. The ML estimator of regression coefficients implied by this formulation may be biased and inconsistent. Additionally, this model stands apart from the other three models, whose estimation seems to produce very similar results. This observation is confirmed by a look at Figure 5.4, which shows graphs of the first 100 estimates of \(\beta_1\) for all four models. Sample sizes of 50 and 500, and \(v=6\), were chosen for this illustration. The estimates obtained using ML-B, ML-S, and OLS tend to move together. The ML-A estimate preserves the general direction of the movement, but not its magnitude. Thus, only a general similarity exists between the two sets of patterns.
Figure 5.4. 100 estimates of $\beta_1$, $\nu=6$. 
Another interesting aspect of all four models considered in this chapter are the inferences about $\beta$ drawn on their basis. To gain insight into this matter, the relevant empirical power curves have to be examined.

The empirical power curves of the traditional $t$-test, used for testing hypotheses about $\beta_1$ in five different ways (the fifth one being the OLS with White’s standard errors), are shown in Figure 5.5, for sample sizes of 50 and 500 and $\nu=6$.\textsuperscript{14} The power curves for the $t$-test using the results of ML-A, ML-B, OLS, and OLS with White’s standard errors look fairly similar. The probability of a type-I error for all these tests is close to the nominal 5%. The probability of a type-II error is similar between the different tests, even though some of the curves exhibit slight asymmetry. The $t$-test using the results of ML-S, has less power in bigger samples than the other four $t$-tests. It is also the only one with a 0% probability of a type-I error. This phenomenon results from the overstatement of the actual variation of $(\hat{\beta}_{1\text{ ML-S}})$ in bigger samples, by the estimated standard errors. This does not occur in the case of $(\hat{\beta}_{2\text{ ML-S}})$, because the estimated standard errors describe the variability correctly (see Table 5.1).

\textsuperscript{14} These empirical power curves were obtained using the method described in section 3.4.
Figure 5.5. Empirical power curves for tests of hypotheses about $\beta_1$ at 5% significance level, $\nu=6$. 
5.3.2. Estimates of $\hat{\beta}_0$.

Kernel density estimates of the empirical distribution of $\hat{\beta}_0$ are shown in Figure 5.6 and Figure 5.7, for $\nu=6$ and various sample sizes. Graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4.

Examination of the kernel density graphs of the empirical distribution of $\hat{\beta}_0$, estimated by ML-A (see Figure 5.6), reveals that the distribution is close to normal, with some evidence of leptokurticity for smaller sample sizes. Its mode is always close to the true value of the constant term ($\beta_0=0$), and its variance goes down as the sample size increases. There is not much difference in the shape of the distribution for different values of $\nu$. Descriptive statistics, reported in Table 5.3, show that the mean of the empirical distribution is close to the true value of $\beta_0$. The hypothesis, that the mean estimate is not significantly different from true $\beta_0$, is not rejected at 5% significance level for any combination of $\nu$ and sample size. Thus, ML-A appears to be an unbiased and consistent estimator of $\beta_0$.

The graphs of the empirical distribution of $\hat{\beta}_0$, estimated by ML-B, show evidence of bimodality for bigger sample sizes (see Figure 5.7). This phenomenon seems to be more pronounced for lower values of $\nu$ (see Appendix 4). Despite the problem with the mode, the mean manages to stay on target (see superimposed normal graph, with the same mean as the actual estimates). It should be kept in mind, that the scale changes between the graphs, thus exaggerating the magnitude of the distortions. The variance of the empirical distribution decreases as the sample size gets bigger. Descriptive statistics reported in Table 5.3 confirm these findings. It should also be noticed, that higher values of $\nu$ result in lower empirical variance and a slightly more narrow range of $(\hat{\beta}_0)_{ML-B}$.

The graphs in Figure 5.7 indicate that the empirical distribution of $\hat{\beta}_0$, estimated by ML-S, is close to normal, with some evidence of leptokurticity for smaller sample sizes and $\nu=4$. Descriptive statistics in Table 5.3 show that its mean is almost exactly equal to true value of the constant term, and the variance decreases as the sample size gets bigger. The hypothesis, that the
mean estimate is not significantly different from the true $\beta_0$, is not rejected at 5% significance level for any combination of $\nu$ and sample size. Thus, $(\hat{\beta}_0)_{ML-S}$ seems to be an unbiased and consistent estimator of $\beta_0$.

The empirical distribution of OLS estimates of $\beta_0$ is close to normal for $\nu=8$, but for $\nu=4$ and $\nu=6$ shows evidence of leptokurticity (see Figure 5.6 and Appendix 4). Its mode and mean are close to the underlying true value and the variance decreases as the sample size gets bigger. The hypothesis, that the mean estimate is not significantly different from true $\beta_0$, tested with the usual $t$-test, cannot be rejected at 5% significance level for any combination of sample size and $\nu$ (see descriptive statistics in Table 5.4). This suggests that OLSE may be an unbiased and consistent estimator of $\beta_0$.

In conclusion, ML-A, ML-S, and OLSE seem unbiased and consistent estimators of $\beta_0$. The ML-B estimator of $\beta_0$ suffered from problems with bimodality in this experiment.
ML-A estimates of $\beta_0$

- 50 observations

OLS estimates of $\beta_0$

- 50 observations

- 100 observations

- 200 observations

- 500 observations

Figure 5.6. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$. 
Figure 5.7. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=6$. 

ML-B estimates of $\beta_0$

ML-S estimates of $\beta_0$
Table 5.3. Selected descriptive statistics of the empirical distribution of ML estimates of $\beta_0$.

<table>
<thead>
<tr>
<th>DF</th>
<th>Sample Size*</th>
<th>ML-A (ASSTM)</th>
<th>ML-B (STAR-B)</th>
<th>ML-S (STAR-S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>v=4</td>
<td>50</td>
<td>-0.0030</td>
<td>0.0493</td>
<td>-0.2304</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0031</td>
<td>0.0339</td>
<td>-0.0137</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0002</td>
<td>0.0223</td>
<td>0.1297</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0006</td>
<td>0.0157</td>
<td>-0.0247</td>
</tr>
<tr>
<td>v=6</td>
<td>50</td>
<td>0.0011</td>
<td>0.0485</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0007</td>
<td>0.0273</td>
<td>-0.2708</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0004</td>
<td>0.0176</td>
<td>0.0175</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0005</td>
<td>0.0110</td>
<td>0.0516</td>
</tr>
<tr>
<td>v=8</td>
<td>50</td>
<td>-0.0024</td>
<td>0.0413</td>
<td>0.1651</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0002</td>
<td>0.0238</td>
<td>-0.0836</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0001</td>
<td>0.0146</td>
<td>-0.0099</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0001</td>
<td>0.0083</td>
<td>0.1548</td>
</tr>
</tbody>
</table>

* Size of the Student t data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.

Table 5.4. Selected descriptive statistics of the empirical distribution of OLS estimates of $\beta_0$.

<table>
<thead>
<tr>
<th>DF</th>
<th>Sample Size*</th>
<th>Empirical Distribution</th>
<th>Estimated Standard Errors</th>
<th>HCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>v=4</td>
<td>50</td>
<td>-0.0005</td>
<td>0.0778</td>
<td>-0.1005</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0030</td>
<td>0.0374</td>
<td>-0.0910</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0007</td>
<td>0.0191</td>
<td>0.2695</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0003</td>
<td>0.0069</td>
<td>0.0940</td>
</tr>
<tr>
<td>v=6</td>
<td>50</td>
<td>-0.0001</td>
<td>0.0658</td>
<td>0.0430</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0008</td>
<td>0.0323</td>
<td>0.2361</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0012</td>
<td>0.0156</td>
<td>0.1628</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0001</td>
<td>0.0067</td>
<td>-0.0435</td>
</tr>
<tr>
<td>v=8</td>
<td>50</td>
<td>-0.0031</td>
<td>0.0623</td>
<td>-0.0103</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0017</td>
<td>0.0310</td>
<td>0.0144</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0009</td>
<td>0.0148</td>
<td>-0.0204</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0001</td>
<td>0.0063</td>
<td>0.0511</td>
</tr>
</tbody>
</table>

* Size of the Student t data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
5.3.3. Estimates of $\sigma^2$.

Kernel density graphs of the empirical distribution of $\hat{\sigma}^2$, for $\nu=6$ and various sample sizes, are shown in Figure 5.8 and Figure 5.9. Graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4. Descriptive statistics for the distribution of $\hat{\sigma}^2$ are reported in Table 5.5.

The graphs in Figure 5.8 indicate that the empirical distribution of $\hat{\sigma}^2$, estimated using the ML-A, is skewed to the left. The distribution collapses, as the sample size gets bigger, but its mode does not converge to the underlying true value ($\sigma^2=1$). In fact, the mode shifts away from the true $\sigma^2$, as the sample size increases. Descriptive statistics in Table 5.5 reveal the existence of a similar problem for the mean. As the sample size increases, the mean of the empirical distribution shifts to the right, away from the true $\sigma^2$. The usual $t$-statistics for the hypothesis, that the mean estimate is not significantly different from the underlying true value, in the sample of 500, are 78.9, 24.42, and 10.22 for $\nu=4, 6, \text{and } 8$ respectively. These $t$-statistics lie so far beyond the 5% critical value (±1.96), that, despite the skewed nature of the distribution of $\hat{\sigma}^2_{\text{ML-A}}$, it can be safely concluded, that the hypothesis is rejected. Together the above findings suggest that $\hat{\sigma}^2_{\text{ML-A}}$ may be an inconsistent estimator of $\sigma^2$.

The graphs in Figure 5.9 show that empirical distribution of $\hat{\sigma}^2$, estimated using ML-B and ML-S, is strongly skewed to the left. Additionally, the shape of the ML-B distribution is severely distorted for $\nu=4$ (see Appendix 4). The mode of both distributions is far off from the true value and does not get any closer as the sample size increases. On the other hand, descriptive statistics in Table 5.5 show that the mean of the ML-B distribution does get closer to the true value of 1, and the empirical standard deviation decreases slightly for bigger samples. This means that $\hat{\sigma}^2_{\text{ML-B}}$ might still be a consistent estimator of $\sigma^2$, but, for any reasonable sample size, it is likely to be way off from the true value. For ML-S there is no discernible pattern in the behavior of the mean of its empirical distribution, and the mean appears to be significantly different from the true value. This suggests biasedness and inconsistency of $\hat{\sigma}^2_{\text{ML-S}}$.
The graphs in Figure 5.8 show that the empirical distribution of $\hat{\sigma}^2$ estimated by OLS is leptokurtic and skewed to the left. As the sample size increases, the mode of the distribution converges to $\frac{v}{v-2}\sigma^2_{ML}$, as expected.\textsuperscript{15} Descriptive statistics in Table 5.5 indicate that the mean converges to the same value.

In conclusion, all of the above estimators of $\sigma^2$ appear either biased, or inconsistent, or are likely to be far off from the true value. Among the Maximum Likelihood estimators, the estimates provided by ML-A were somewhat closer to the true $\sigma^2$ than the estimates obtained using the other two MLEs.

\textsuperscript{15} This equals 2, 1.5, and 1.33 for $v=4$, 6, and 8 respectively.
ML-A estimates of $\sigma^2, \nu=6$ (ASSTM)

OLS estimates of $\sigma^2, \nu=6$ (DLRM)

Figure 5.8. Kernel density estimates of the empirical distribution of $\sigma^2, \nu=6$. 
ML-B estimates of $\sigma^2$, $\nu=6$ (STAR-B)

ML-S estimates of $\sigma^2$, $\nu=6$ (STAR-S)

Figure 5.9. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=6$. 
Table 5.5. Selected descriptive statistics of the empirical distributions of ML and OLS estimates of $\sigma^2$.

<table>
<thead>
<tr>
<th>DF</th>
<th>Sample Size$^*$</th>
<th>ML-A (ASSTM)</th>
<th>ML-B (STAR-B)</th>
<th>ML-S (STAR-S)</th>
<th>OLS (DLRM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>v=4</td>
<td>50</td>
<td>1.0171</td>
<td>0.1414</td>
<td>-0.3999</td>
<td>2.8689</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.1094</td>
<td>0.0980</td>
<td>-0.3625</td>
<td>3.2312</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.1694</td>
<td>0.0699</td>
<td>-0.0632</td>
<td>2.9087</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.2104</td>
<td>0.0477</td>
<td>-0.0494</td>
<td>2.7933</td>
</tr>
<tr>
<td>v=6</td>
<td>50</td>
<td>0.9685</td>
<td>0.1114</td>
<td>-0.4221</td>
<td>2.8361</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.0158</td>
<td>0.0707</td>
<td>-0.3425</td>
<td>2.9453</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.0484</td>
<td>0.0477</td>
<td>-0.3539</td>
<td>3.0104</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0727</td>
<td>0.0281</td>
<td>-0.0791</td>
<td>2.8827</td>
</tr>
<tr>
<td>v=8</td>
<td>50</td>
<td>1.0005</td>
<td>0.0870</td>
<td>-0.4928</td>
<td>3.1706</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.0273</td>
<td>0.0577</td>
<td>-0.5028</td>
<td>3.3158</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.0428</td>
<td>0.0353</td>
<td>-0.3303</td>
<td>3.1156</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0527</td>
<td>0.0234</td>
<td>-0.4233</td>
<td>3.0299</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
5.3.4. Estimates of $\Sigma_{22}^{-1}$ and $(\Sigma_{T}^{1})^{-1}$.

The empirical distribution of the estimates of two elements of $\Sigma_{22}^{-1}$ (inverted scale matrix in the ASSTM) is shown in Figure 5.10 for $\nu=6$. The graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4. The elements chosen for this illustration are $(\Sigma_{22}^{-1})_{2,2} = 0.96$ - the second diagonal element, and $(\Sigma_{22}^{-1})_{2,1} = -0.48$ - the off-diagonal element. The second diagonal element was chosen for reasons which will become apparent when the distribution of $\hat{\sigma}^{2} \hat{\Sigma}_{22}^{-1}$ is discussed. Descriptive statistics for the empirical distributions of estimates of both of these elements are reported in Table 5.6.

The graphs in Figure 5.10 show that the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$ is skewed to the right. As the sample size increases, the mode of the distribution shifts left. It should be observed that the mode shifts in the opposite direction than it did for $\hat{\sigma}^{2}_{ML-A}$. This is reminiscent of the static case, where the product $\hat{\sigma}^{2} \hat{\Sigma}_{22}^{-1}$ was found to be well-behaved (see section 3.4). In a moment the same possibility will be explored in the present case.

Descriptive statistics in Table 5.6 confirm the above observations. The mean of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$ initially moves towards the true value, but then it overshoots it. The usual $t$-statistics for the hypothesis that the mean estimate is equal to the true $(\Sigma_{22}^{-1})_{2,2}$, in the sample of 500, are -104.9, -82.5, and -93.1 for $\nu=4$, 6, and 8 respectively. These $t$-statistics lie so far beyond the 5% critical value ($\pm 1.96$), that, despite the skewed nature of the distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$, it can be safely concluded that the hypothesis is rejected.

Graphs of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,1}$ show that it is close to normal. Descriptive statistics in Table 5.6 confirm this observation, and show that the mean estimate exhibits erratic behavior (converges for $\nu=8$, diverges for $\nu=4$).
Taken together the above findings suggest that $(\hat{\Sigma}_{22}^{-1})_{\text{ML-A}}$ may be a biased and inconsistent estimator of $(\Sigma_{22}^{-1})_{\text{ASM}}$.

Table 5.6. Selected descriptive statistics of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{2,2}$ and $(\hat{\Sigma}_{22}^{-1})_{2,1}$ (ML-A).

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>DF</th>
<th>Sample Size¹</th>
<th>Mean Estimate</th>
<th>Empirical SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\hat{\Sigma}<em>{22}^{-1})</em>{2,2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu=4$</td>
<td>50</td>
<td>1.0840</td>
<td>0.1626</td>
<td>1.4574</td>
<td>7.8512</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9520</td>
<td>0.0957</td>
<td>1.0163</td>
<td>5.2063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.8909</td>
<td>0.0633</td>
<td>0.5075</td>
<td>3.2427</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.8491</td>
<td>0.0411</td>
<td>0.3096</td>
<td>3.0951</td>
<td></td>
</tr>
<tr>
<td>$\nu=6$</td>
<td>50</td>
<td>1.0600</td>
<td>0.1235</td>
<td>1.1812</td>
<td>5.6248</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9969</td>
<td>0.0716</td>
<td>0.3183</td>
<td>3.0050</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9581</td>
<td>0.0505</td>
<td>0.5274</td>
<td>4.0598</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9310</td>
<td>0.0301</td>
<td>0.0447</td>
<td>2.8955</td>
<td></td>
</tr>
<tr>
<td>$\nu=8$</td>
<td>50</td>
<td>0.9880</td>
<td>0.1011</td>
<td>0.7959</td>
<td>4.1858</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9586</td>
<td>0.0599</td>
<td>0.5565</td>
<td>3.2057</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9414</td>
<td>0.0396</td>
<td>0.0957</td>
<td>2.8782</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9317</td>
<td>0.0265</td>
<td>0.3843</td>
<td>3.1116</td>
<td></td>
</tr>
<tr>
<td>$(\hat{\Sigma}<em>{22}^{-1})</em>{2,1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu=4$</td>
<td>50</td>
<td>-0.4979</td>
<td>0.1578</td>
<td>-0.2734</td>
<td>3.4215</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.4543</td>
<td>0.0998</td>
<td>-0.1836</td>
<td>3.3615</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.4414</td>
<td>0.0694</td>
<td>-0.1617</td>
<td>3.2840</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.4245</td>
<td>0.0452</td>
<td>0.1474</td>
<td>3.1949</td>
<td></td>
</tr>
<tr>
<td>$\nu=6$</td>
<td>50</td>
<td>-0.4787</td>
<td>0.1651</td>
<td>-0.1018</td>
<td>4.9025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.4648</td>
<td>0.1031</td>
<td>-0.0470</td>
<td>3.2815</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.4677</td>
<td>0.0685</td>
<td>-0.0273</td>
<td>3.0212</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.4626</td>
<td>0.0444</td>
<td>-0.0973</td>
<td>3.1472</td>
<td></td>
</tr>
<tr>
<td>$\nu=8$</td>
<td>50</td>
<td>-0.4217</td>
<td>0.1456</td>
<td>0.1326</td>
<td>3.9584</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.4545</td>
<td>0.1014</td>
<td>-0.0120</td>
<td>3.2021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.4544</td>
<td>0.0703</td>
<td>-0.0945</td>
<td>3.1486</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.4614</td>
<td>0.0423</td>
<td>0.0191</td>
<td>3.0108</td>
<td></td>
</tr>
</tbody>
</table>

¹ Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
ML-A estimates of $\left(\Sigma_{22}^{-1}\right)_{2,1}$, $\nu=6$ (ASSTM)

Figure 5.10. Kernel density estimates of the empirical distribution of $\left(\Sigma_{22}^{-1}\right)_{2,1}$, $\nu=6$. 

ML-A estimates of $\left(\Sigma_{22}^{-1}\right)_{2,1}$, $\nu=6$ (ASSTM)
Graphs of the empirical distribution of estimates of two elements of $(\Sigma_T^{-1})$, estimated by ML-B and ML-S, are shown in Figure 5.11 and Figure 5.12 for $\nu=6$. Graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4. The elements chosen for the illustration are $(\Sigma_T^{-1})_{,i} = 1.2$ - the center element inside the diagonal band, and $(\Sigma_T^{-1})_{,i-2} = -0.2$ - the outermost element in the band $(2 < i < T-1)$. Descriptive statistics for the empirical distribution of the estimates of these elements can be found in Table 5.7.

The graphs of the empirical distribution of $(\hat{\Sigma}_T^{-1})_{,i,ML-B}$ and $(\hat{\Sigma}_T^{-1})_{,i,ML-S}$ look similar. They are both strongly skewed to the right and their modes do not show signs of convergence to the underlying true value. Descriptive statistics in Table 5.7 indicate that there is no discernible pattern in the behavior of the mean estimate.

The graphs of the empirical distribution of $(\hat{\Sigma}_T^{-1})_{,i-2,ML-B}$ and $(\hat{\Sigma}_T^{-1})_{,i-2,ML-S}$ are strongly skewed to the left. Descriptive statistics in Table 5.7 indicate, that, in this case, the mean estimate does shift towards the underlying true value as the sample size increases.

Taken together, the above observations suggest that both $(\hat{\Sigma}_T^{-1})_{ML-B}$ and $(\hat{\Sigma}_T^{-1})_{ML-S}$ may be biased and inconsistent estimators of $(\Sigma_T^{-1})$. 
Table 5.7. Selected descriptive statistics of the empirical distribution of \((\hat{\Sigma}_T^{i})_i\) and \((\hat{\Sigma}_T^{i})_{i-1}\) (ML-B, ML-S).

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>(\hat{\Sigma}_T^{i})_i)</th>
<th>(\hat{\Sigma}<em>T^{i})</em>{i-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>STAR-B</td>
<td>STAR-S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean Est.</td>
<td>Emp. SD</td>
</tr>
<tr>
<td>(\nu=4)</td>
<td>50</td>
<td>0.9759</td>
<td>1.2714</td>
<td>3.4555</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9575</td>
<td>1.1300</td>
<td>3.7529</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9385</td>
<td>0.8626</td>
<td>4.1861</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9848</td>
<td>1.0794</td>
<td>6.0510</td>
</tr>
<tr>
<td>(\nu=6)</td>
<td>50</td>
<td>1.0682</td>
<td>0.8342</td>
<td>3.0469</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9981</td>
<td>0.7068</td>
<td>5.2869</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9955</td>
<td>0.6853</td>
<td>5.5032</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0063</td>
<td>0.6466</td>
<td>4.2136</td>
</tr>
<tr>
<td>(\nu=8)</td>
<td>50</td>
<td>1.0252</td>
<td>0.4351</td>
<td>2.6026</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.0528</td>
<td>0.4904</td>
<td>3.4612</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.0406</td>
<td>0.3934</td>
<td>2.5830</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0409</td>
<td>0.4040</td>
<td>3.4584</td>
</tr>
</tbody>
</table>

* Size of the Student t data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
Figure 5.11. Kernel density estimates of the empirical distribution of \((\hat{\Sigma}_1^{-1})_{i,1}^i\) and \((\hat{\Sigma}_1^{-1})_{i,2}^i\), \(\nu=6\).
ML-S estimates of $(\Sigma_j^{-1})_{1,1}, \nu=6$ (STAR-S)  

ML-S estimates of $(\Sigma_j^{-1})_{1,2}, \nu=6$ (STAR-S)

Figure 5.12. Kernel density estimates of the empirical distribution of $(\Sigma_j^{-1})_{1,1}$ and $(\Sigma_j^{-1})_{1,2}, \nu=6$.  

50 observations

100 observations

200 observations

500 observations
5.3.5. The Product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ and $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})^{-1}$.

Figure 5.13 shows the empirical distributions of $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,2}$ and $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,1}$ for $\nu=6$. The graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4. The true values of these expressions are the elements of $\sigma^2 \Sigma_{22}^{-1}$, which equals

$$
1 \times \begin{bmatrix}
0.96 & -0.48 \\
-0.48 & 0.96
\end{bmatrix} =
\begin{bmatrix}
0.96 & -0.48 \\
-0.48 & 0.96
\end{bmatrix}.
$$

The second diagonal element was chosen for this illustration, because its true value (0.96) is different from $(\Sigma_T^{-1})_{1,1}^r=1$, $(\Sigma_T^{-1})_{2,2}^r = 1.16$, and $(\Sigma_T^{-1})_{i,i}^r=1.2$ ($2<i<T-1$). This property can provide a clue, whether the empirical distribution of estimates is concentrated around the true value for the ASSTM formulation, or around the true value of any of the elements in the full-blown model. Descriptive statistics for the empirical distributions of $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,2}$ and $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,1}$ are reported in Table 5.8.

The graphs in Figure 5.13 show that the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,2}$ exhibits a slight positive skewness and leptokurticity in smaller samples. As the sample size increases, the skewness changes from positive to negative, and the leptokurticity is no longer present. At the same time, the mode of the distribution does not converge to the underlying true value. On the other hand, descriptive statistics in Table 5.8 show that the mean estimate does move towards the true $\sigma^2 (\Sigma_{22}^{-1})_{2,2}$ as the sample size increases. However, the hypothesis, that the difference between the mean estimate and the true $\sigma^2 (\Sigma_{22}^{-1})_{2,2}$ is not significant, is rejected at 5% level, for all samples of 500 ($t$-statistics 50.87, 37.11, and 20.83 for $\nu=4, 6,$ and 8 respectively).

The mode of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,1}$ exhibits erratic behavior, although, for bigger samples, it does get closer to the true value (see Figure 5.13). Descriptive statistics in Table 5.8 reveal, that, for $\nu=4$, the mean of the empirical distribution moves away from the true

---

16 A product of estimates, NOT an estimate of the product.
$\sigma^2 (\Sigma^{-1}_{22})_{2,1}$ as the sample size increases. For $\nu=6$ and $\nu=8$, it initially moves towards the true value, but then it overshoots it. The difference between the mean estimate and the true value is significant at 5% level in all samples of 500 ($t$-statistics -14.39, -7.74, and -2.78 for $\nu=4$, 6, and 8 respectively).

Taken together, the above findings suggest, that $\hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{\text{ML-A}}$ may be not as well behaved as its counterpart in the static case.

Table 5.8. Selected descriptive statistics of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{2,2}$ and $\hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{2,1}$ (ML-A).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Sample Size</th>
<th>Mean Estimate</th>
<th>Empirical SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\sigma}^2 (\hat{\Sigma}^{-1}<em>{22})</em>{2,2}$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>1.0884</td>
<td>0.1341</td>
<td>0.0599</td>
<td>3.2162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.0496</td>
<td>0.0759</td>
<td>0.1025</td>
<td>4.1886</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.0386</td>
<td>0.0495</td>
<td>-0.1403</td>
<td>3.2096</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.0262</td>
<td>0.0291</td>
<td>-0.1058</td>
<td>3.2301</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>1.0188</td>
<td>0.1089</td>
<td>0.2483</td>
<td>3.8125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.0095</td>
<td>0.0620</td>
<td>-0.0290</td>
<td>2.8455</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.0028</td>
<td>0.0381</td>
<td>-0.2341</td>
<td>2.9978</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.9980</td>
<td>0.0229</td>
<td>-0.1996</td>
<td>2.8786</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>0.9844</td>
<td>0.0962</td>
<td>0.0714</td>
<td>3.5298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.9828</td>
<td>0.0545</td>
<td>-0.2307</td>
<td>2.8932</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.9809</td>
<td>0.0348</td>
<td>-0.5217</td>
<td>3.2695</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.9804</td>
<td>0.0259</td>
<td>-0.3346</td>
<td>3.3235</td>
</tr>
<tr>
<td>$\hat{\sigma}^2 (\hat{\Sigma}^{-1}<em>{22})</em>{2,1}$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>-0.5058</td>
<td>0.1642</td>
<td>0.0562</td>
<td>2.9399</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.5023</td>
<td>0.1097</td>
<td>0.0955</td>
<td>3.0170</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.5147</td>
<td>0.0760</td>
<td>0.1565</td>
<td>3.0281</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.5131</td>
<td>0.0514</td>
<td>0.3421</td>
<td>3.2087</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>-0.4650</td>
<td>0.1607</td>
<td>0.2332</td>
<td>3.3024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.4719</td>
<td>0.1061</td>
<td>0.2268</td>
<td>3.0076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.4899</td>
<td>0.0711</td>
<td>0.1250</td>
<td>2.8355</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.4960</td>
<td>0.0462</td>
<td>0.0289</td>
<td>3.0203</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>-0.4243</td>
<td>0.1502</td>
<td>0.1794</td>
<td>3.5215</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>-0.4672</td>
<td>0.1058</td>
<td>0.2008</td>
<td>3.1768</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>-0.4737</td>
<td>0.0735</td>
<td>0.1391</td>
<td>2.9023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>-0.4856</td>
<td>0.0450</td>
<td>0.0964</td>
<td>3.0933</td>
</tr>
</tbody>
</table>

Size of the Student $t$ data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
Figure 5.13. Kernel density estimates of the empirical distribution of $\sigma^2 (\hat{\Sigma}^{-1})_{2,2}$ and $\sigma^2 (\hat{\Sigma}^{-1})_{2,3}$ (ML-A), $\nu=6$. 

ML-A estimates of $\sigma^2 (\Sigma^{-1})_{2,2}$, $\nu=6$ (ASSTM)

ML-A estimates of $\sigma^2 (\Sigma^{-1})_{2,3}$, $\nu=6$ (ASSTM)
The graphs of the empirical distributions of $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{i,j}$ and $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{i,i-2}$, $2 < i < T - 1$, estimated by ML-B and ML-S, are shown in Figure 5.14 and Figure 5.15. The graphs for $\nu=4$ and $\nu=8$ can be found in Appendix 4. Descriptive statistics for the empirical distribution of the estimates of both elements are reported in Table 5.9.

Both sets of graphs of $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{i,j}$ exhibit similar behavior. For smaller samples the distribution is skewed to the right, then, as the sample size increases, it tends towards normality. Its mode quickly converges to the true value. A look at Table 5.9 reveals that the mean is always very close to the true value and the standard deviation decreases as the sample size gets bigger. For the sample size of 500, the hypothesis, that the mean estimate is not significantly different from the true value, cannot be rejected at 5% significance level for any value of $\nu$ ($t$-statistics -1.92, -0.59, -1.13 (ML-B), and -1.64, -0.59, -0.90 (ML-S) for $\nu=4$, 6, and 8 respectively; 5% critical value ±1.96). The above findings suggest that $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{i,j}$ might be a consistent estimator of $\sigma^2 (\Sigma_1^{-1})_{i,j}$.

The graphs of the empirical distributions of $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{i,i-2}$, estimated using ML-B and ML-S, exhibit some rightward skewness. However, both distributions pass the Bera-Jarque test for normality for all samples of 500.\textsuperscript{17} In both cases the mode converges to the true value and the variance goes down as the sample size increases. Descriptive statistics in Table 5.9 show that the mean also converges to the true value as the sample size gets bigger. Even though the hypothesis, that the mean estimate is equal to the true value, is rejected for the sample size of 500 for all $\nu$’s,\textsuperscript{18} convergence is apparent by looking at the numbers in Table 5.9.

Taken together, the above findings suggest that both $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{\text{ML-B}}$ and $\hat{\sigma}^2 (\hat{\Sigma}_1^{-1})_{\text{ML-S}}$ might be consistent estimators of $\sigma^2 (\Sigma_1^{-1})^{-1}$.

\textsuperscript{17} Test statistics 3.71, 0.15, 2.14 (ML-B), and 3.66, 1.28, 1.74 (ML-S) for $\nu=4$, 6, and 8 respectively. 5% critical value 5.99. See Bera and Jarque (1982).

\textsuperscript{18} $t$-statistics 2.77, 3.55, 3.86 (ML-B), and 2.73, 3.50, 3.85 (ML-S) for $\nu=4$, 6, and 8 respectively. 5% critical value ±1.96.
As was the case with the static model, the above findings help to salvage the usefulness of the STAR models as a viable way of modeling the conditional variance. The argument underlying this claim was outlined in section 3.4. Therefore, let us only state here, that the predicted conditional variance should be off by a fairly constant amount, and it should be able to mimic the path followed by the true conditional variance reasonably well.

Table 5.9. Selected descriptive statistics of the empirical distribution of \( \hat{\sigma}^2 (\hat{\Sigma}_T)^{-1} \) and \( \hat{\sigma}^2 (\hat{\Sigma}_T)^{-1} \).

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>ML-B (STAR-B)</th>
<th>ML-S (STAR-S)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean Est.</td>
<td>Emp. SD</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 (\hat{\Sigma}_T)^{-1} )</td>
<td>v=4</td>
<td>50</td>
<td>1.1980</td>
<td>0.1056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.1939</td>
<td>0.0700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.2024</td>
<td>0.0487</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.1973</td>
<td>0.0313</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 (\hat{\Sigma}_T)^{-1} )</td>
<td>v=6</td>
<td>50</td>
<td>1.2067</td>
<td>0.1201</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.1879</td>
<td>0.0688</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.1983</td>
<td>0.0474</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.1992</td>
<td>0.0305</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 (\hat{\Sigma}_T)^{-1} )</td>
<td>v=8</td>
<td>50</td>
<td>1.1888</td>
<td>0.1004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.1965</td>
<td>0.0725</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.1956</td>
<td>0.0494</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.1985</td>
<td>0.0298</td>
</tr>
</tbody>
</table>

Size of the Student \( t \) data set used in estimation. For each size of the data set, 500 estimates were used to obtain the numbers reported in the table.
Figure 5.14. Kernel density estimates of the empirical distribution of $\sigma^2 (\Sigma_i^{-1})$, $\nu$=6 (STAR-B).
Figure 5.15. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_T^{-1})_{ii}$ and $\hat{\sigma}^2 (\hat{\Sigma}_T^{-1})_{ii+2}$ (ML-S), $\nu=6$. 

ML-S estimates of $\sigma^2 (\Sigma_T^{-1})_{ii}, \nu=6$ (STAR-S)

ML-S estimates of $\sigma^2 (\Sigma_T^{-1})_{ii+2}, \nu=6$ (STAR-B)
5.3.6. Modeling the Conditional Variance.

To examine which model is able to approximate the behavior of the conditional variance most closely, the following experiment was performed. A series of 200 observations of \( Y_t \) was generated using the procedure described in section 5.2. The true structure of the series was the same as in (5.1). All three Student \( t \) models were estimated for these data, and their predictions of the conditional variance were plotted. The DLRM was also included for comparison. The results are presented in Figure 5.16.

The pictures in Figure 5.16 show, that, after an initial period of volatility, the conditional variance stabilizes. This smoothing effect occurs, because the weight given to each new observation decreases, as new observations are added to the series. STAR-B and STAR-S are able to follow the pattern of the actual conditional variance reasonably well. It should be noticed,
that, in this particular experiment, they erred on different sides. The STAR-B overestimated and the STAR-S underestimated the true conditional variance. The DLRM by construction assumes that the conditional variance is homoskedastic, and tries to come up with a single prediction that best characterizes the data. It is only logical, that the conditional variance predicted by the DLRM seems to be the same as the long-term actual variance.

The conditional variance predicted by the ASSTM shows a totally different pattern than the true conditional variance. A fairly large amount of variability suggests that the ASSTM predictions might reflect a short-term variance. To gain some insight into this possibility, the sample variance was estimated using the OLS window method. The size of the window was set to 4 observations. Figure 5.17 shows both the conditional variance predicted by ASSTM and the conditional variance estimated using OLS window. The two variances seem to have a similar pattern.

![Figure 5.17. Window OLS and ASSTM estimates of the conditional variance.](image)

5.4. Conclusions.

In this chapter, an effort was undertaken to evaluate the performance of several Dynamic Student *t* Regression Models. The questions of interest included the usefulness of the MLEs implied by these DSTRMs for estimation and testing, the models’ ability to capture the changes in the conditional variance, and their relative performance when compared to the traditional DLRM.
The experiment seemed to confirm that ML-B, ML-S, and OLSE are consistent estimators of regression coefficients. It also revealed some problems with ML-A, which may be a biased and inconsistent estimator of regression coefficients.

The empirical power curves for the $t$-test based on ML-A, ML-B, OLS, and OLS with White’s standard errors looked very similar. The actual size of the $t$-test for all these estimators was close to the nominal 5%. The ML-S $t$-test had less power in bigger samples than the other four $t$-tests. It also had an actual size of 0% when the nominal size was set at 5%.

The ML-A, ML-B, and ML-S estimates of $\sigma^2$, and inverted scale matrix, all exhibited characteristics pointing towards biasedness and/or inconsistency. The same was true about the product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ estimated by ML-A. However, the product $\hat{\sigma}^2 (\hat{\Sigma}_T^{-1})^{-1}$, estimated by ML-B and ML-S, was found to be reasonably well behaved. This last finding helped justify the usefulness of the DSTRMs in modeling of the conditional variance.

STAR-B and STAR-S proved capable of modeling the heteroskedastic conditional variance accurately. Some differences between these two models manifested themselves, as the STAR-B overpredicted and the STAR-S underpredicted the true variance. However, predictions from both models were able to mimic the pattern of the actual conditional variance very well. On the other hand, the DLRM by construction is incapable of modeling the heteroskedastic conditional variance. Its prediction turned out to be identical to the long-run level of the true conditional variance. The ASSTM was not able to predict the true conditional variance accurately, either. Instead, its predictions seem to model a short term variance.
Chapter 6

Conclusions

In this dissertation an attempt was undertaken to evaluate the performance of a number of Static and Dynamic Student $t$ Regression Models. The questions of interest included:

1. The ability of the implied MLEs to estimate the underlying parameters accurately.
2. The ability of the estimated standard errors to describe correctly the dispersion of the estimates.
3. The ability of the Student $t$ models to model the conditional variance accurately.

The models were also compared with the benchmark of the traditional formulations estimated by OLS. This was done to gain an idea about the potential improvement resulting from utilizing all the information in the data, and to gain insight into the consequences of ignoring/overlooking non-normality. The evaluation was based on a series of Monte Carlo experiments.

The experiments involving static Student $t$ data seem to confirm that both MLE and OLSE are consistent estimators of regression coefficients. The OLS estimator was found to have a lower empirical variance than the ML estimator, despite the fact that its distribution was skewed. Significant differences between ML and OLS were uncovered regarding testing. The ML $t$-test with the nominal size of 5% was found to have an actual size of 0%. In the same experiment, the OLS $t$-test showed much more power than the ML $t$-test, but it also had a rejection rate of 15.6%–23% for the true hypothesis, depending on the sample size. The OLS $t$-test using White’s standard errors was found to have even less power than the ML $t$-test. The choice among the above three methods should depend on the particular set of circumstances.

The experiments produced evidence suggesting that the ML estimators of $\sigma^2$ and $\Sigma_{22}^{-1}$ may be biased and inconsistent. However, in this experiment, the ML estimators of these two parameters tended to err in different directions. As a consequence, the product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ was found to be well behaved, because the mistakes canceled each other out. This fact enabled the SSTRM to approximate the pattern of the true conditional variance very closely. On the other hand, the NLRM by construction is incapable of modeling the changes in the conditional variance.
The above findings were largely invariant to changes in the value of the degrees of freedom parameter.

The performance of a number of the DSTRMs was examined using dynamic Student $t$ data. The models selected for detailed evaluation included STAR-B, STAR-S, and ASSTM. The models were compared with each other and with the benchmark of the DLRM. The comparisons included usefulness of their implied ML estimators (ML-B, ML-S, and ML-A) for estimation and testing and the ability to predict the conditional variance correctly.

The experiment seemed to confirm that the ML-B, ML-S and OLSE are consistent estimators of regression coefficients. The estimates obtained using these three estimators were found to be close to each other and their empirical distributions showed increasing degree of similarity as the sample size got bigger. On the other hand, the experiments suggested that the ML-A may be a biased and inconsistent estimator of regression coefficients.

Surprisingly little difference in power was uncovered between different $t$-tests. The empirical power curves for the $t$-tests based on the results of ML-A, ML-B, OLS, and OLS with White’s standard errors looked very similar and the actual size of these tests was close to the nominal 5%. The ML-S $t$-test was found to have less power in bigger samples than the other four $t$-tests.

The empirical distributions of the estimates of $\sigma^2$ and inverted scale matrix (which had different dimensions for ASSTM and the STAR models) suffered from the same problems as their counterparts in the static case. However, in the experiment involving the STAR models, the product of these two estimates was again relatively well behaved. This enabled both STAR models to approximate the pattern of the actual conditional variance very closely. The two STAR models differed somewhat in their predictions of the conditional variance. The STAR-B tended to overestimate and the STAR-S tended to underestimate the actual variance. The prediction from the DLRM was found to be the same as the level at which the true heteroskedastic variance stabilizes in the long run.
On the other hand, in the experiment involving ASSTM, the product $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$ did not show clear signs of convergence to the underlying true value. Additionally, the conditional variance predicted by ASSTM exhibited a pattern completely different from the true conditional variance. A comparison with the OLS window estimates suggests that the conditional variance predicted by ASSTM may be approximating a short-term variation. The reliability of this approximation remains unknown, due to erratic behavior of $\hat{\sigma}^2 \hat{\Sigma}_{22}^{-1}$.

Comparing the results of the experiments involving the static models with the results of the experiments involving the dynamic models, a striking difference in the performance of OLSE can be observed. In the static case, the empirical distribution of the OLS estimates of regression coefficients was skewed, and the OLS $t$-test showed much more power than either the ML $t$-test or the OLS $t$-test using White’s standard errors. In the dynamic case, the empirical distribution of the OLS estimates of regression coefficients was close to normal. Additionally, the empirical power curves for the $t$-test, using both regular and White’s standard errors, were almost identical to the empirical power curve of the ML-B $t$-test. One possible explanation of this difference between the static and the dynamic case may be the presence of collinearity (in the static case the columns of $X$ were orthogonal to each other). This suggests, that an examination of the impact of collinearity on the performance of OLSE, and MLE implied by the SSTRM, might constitute an interesting extension of this study.

Another interesting extension might involve a Monte Carlo experiment performed on the proper Markov formulation from section 4.4. Of particular interest in this context is the question whether the results obtained by estimation of the Markov model would be different from those obtained by estimation of the ASSTM.

Finally, another question of interest involves the performance of the full-blown STAR model from section 4.2. A Monte Carlo experiment involving this model could not be performed because of insufficient speed of current computers. However, the fast pace of progress in the computer industry should render this experiment feasible very soon.
References


Appendix 1

Derivation of the Conditional Student t Density

Consider a vector \((Y_t, X_t)\)', \(Y_t: 1\times 1, X_t: k\times 1\), jointly distributed as

\[
(Y_t, X_t) \sim \text{St}_{k+1}(\mu, \Sigma, \nu) = \text{St}_{k+1}\left[\begin{pmatrix} \mu_1 \\ \sigma_{12} \\ \sigma_{21} \\ \Sigma_{22} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}, \nu\right].
\]

The conditional density \(D(Y_t|X_t)\) is defined as

\[
D(Y_t|X_t) = \frac{D(Y_t, X_t)}{D(X_t)}.
\]

Using the density formula given in Johnson and Kotz (1972) this can be written as

\[
D(Y_t|X_t) = \frac{\Gamma\left(\frac{v+k+1}{2}\right)}{(\pi v)^{k+1/2} \Gamma\left(\frac{v}{2}\right)} \left|\Sigma\right|^{1/2} \left[1 + \frac{1}{v} (Y_t - \mu_1, X_t - \mu_2) \Sigma^{-1} (Y_t - \mu_1, X_t - \mu_2)\right]^{-\frac{v+k+1}{2}}.
\]

Applying the results stated in Searle (1982) p.258 and p.260 this can be rewritten as

\[
D(Y_t|X_t) = \frac{\Gamma\left(\frac{v+k+1}{2}\right)}{(\pi v)^{k+1/2} \Gamma\left(\frac{v}{2}\right)} \left|\Sigma_{22}\right|^{1/2} \left|\sigma_{11} - \sigma_{12} \Sigma_{22}^{-1} \sigma_{21}\right|^{1/2} \times
\]

\[
\left[1 + \frac{1}{v} (X_t - \mu_2) \Sigma_{22}^{-1} (X_t - \mu_2) + \frac{1}{v\sigma^2} (Y_t - \beta_0 - \beta' X_t)^2\right]^{\frac{v+k+1}{2}} \times
\]

\[
\left[1 + \frac{1}{v} (X_t - \mu_2) \Sigma_{22}^{-1} (X_t - \mu_2)\right]^{\frac{v+k+1}{2}},
\]

128
where

\[ \sigma^2 = \sigma_{11} - \sigma_{12} \Sigma^{-1}_{22} \sigma_{21}. \]

This is equivalent to

\[
D(Y|\mathbf{X}_t) = \frac{\Gamma\left(\frac{\nu + k + 1}{2}\right)}{(\pi \nu)^{\frac{1}{2}} \Gamma\left(\frac{\nu + k}{2}\right)} (\sigma^2)^{\nu + k + 1} \times \frac{\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right]^{\nu + k + 1}}{\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right]^{(\frac{\nu + k}{2})}}.
\]

Simplifying, and grouping the elements together, we get

\[
D(Y|\mathbf{X}_t) = \frac{\Gamma\left(\frac{\nu + k + 1}{2}\right)}{(\pi \nu)^{\frac{1}{2}} \Gamma\left(\frac{\nu + k}{2}\right)} \left[\sigma^2\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right]\right]^{\nu + k + 1} \times \frac{\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right]^{\frac{\nu + k}{2}}}{\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right]^{\nu + k + 1}},
\]

(A1.1)

which is a \( t \) density\(^1\) with:

- mean = \( \beta_0 + \mathbf{\beta}'\mathbf{X}_t \);
- scale matrix = \( \sigma^2\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right] \) (here reduced to a scalar);
- \( \nu + k \) degrees of freedom.

This can be written compactly as

\[
D(Y_t|\mathbf{X}_t) \sim \text{St}\left(\beta_0 + \mathbf{\beta}'\mathbf{X}_t, \sigma^2\left[1 + \frac{1}{\nu} (\mathbf{X}_t - \mathbf{\mu}_2)' \Sigma^{-1}_{22} (\mathbf{X}_t - \mathbf{\mu}_2)\right], \nu + k\right).
\]

\(^1\) For a proof that (A1.1) constitutes a proper density function see Zellner (1971).
Zellner (1971) arrives at a more general version of this expression considering the
distribution of a partitioned vector. His original formulation uses the partitions of the inverted
scale matrix. Re-expressed in terms of the underlying primary parameters Zellner’s equation takes
the form

\[
D(y|x) = \frac{\Gamma \left( \frac{v + m + k}{2} \right)}{(\pi v)^{\frac{m}{2}}} \Gamma \left( \frac{v + k}{2} \right) \left[ 1 + \frac{1}{v} (x - \mu_2) \Sigma_{22}^{-1} (x - \mu_2) \right] \left( \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)^{-\frac{1}{2}} \times \\
\times \left[ 1 + \frac{1}{v} (y - \beta_0 - B' x) \right] \left[ 1 + \frac{1}{v} (x - \mu_2) \Sigma_{22}^{-1} (x - \mu_2) \right]^{-1} \left( y - \beta_0 - B' x \right)^{-\frac{v + m + k}{2}},
\]

where \( m \) is the number of variables in vector \( y \) and \( k \) in vector \( x \). This constitutes a multivariate
generalization of (A1.1).
Appendix 2

Delta Method

The δ-method constitutes a convenient way of approximating the distribution of an arbitrary function of a random variable, when the distribution of that variable is known. The method is extremely useful when the estimable model cannot be expressed directly in terms of the parameters of interest. In this case the estimates and their approximate standard errors still can be recovered using the procedure outlined below.

To develop a methodology general enough to accommodate a number of cases let us assume that we are dealing with patterned matrices and a generic elliptical distribution.¹

Suppose

\[
\sqrt{n}\text{vecp}_\Theta (X - \Theta) \sim EC(\Theta, \Omega, \phi), \tag{A2.1}
\]

where \(n\) is the number of observations, \(\Theta\) is the matrix of true parameters, \(X\) is an estimate of matrix \(\Theta\), the subscript on \text{vecp} denotes the relevant pattern for the \text{vecp} operator, and EC denotes an elliptically-contoured distribution with the mean vector \(\Theta\), scale matrix \(\Omega\), and parameters vector \(\phi\). Further, suppose that the theoretical parameter of interest is not \(\Theta\), but some function of it, say \(g(\Theta)\).

Using the first order Taylor series expansion \(\text{vecp}_{g(\Theta)}(X)\) can be approximated as

\[
\text{vecp}_{g(\Theta)}(X) \cong \text{vecp}_{g(\Theta)}(\Theta) + \frac{1}{1!} \frac{\partial\text{vecp}_{g(\Theta)}(\Theta)}{\partial \text{vecp}_\Theta} \text{vecp}_\Theta (X - \Theta).
\]

This is equivalent to

¹ When the matrix is symmetric \text{vecp} is replaced by \text{vech} and when the matrix contains no pattern whatsoever plain \text{vec} is sufficient. See Chapter 2 and Chapter 4 for definitions of \text{vech} and \text{vecp}. 

131
\[
\sqrt{n}\text{vec}_{\theta}(g(X) - g(\Theta)) \
\equiv \frac{\partial \text{vec}_{\theta} g(\Theta)}{\partial \text{vec}_{\theta} \Theta} \sqrt{n} \text{vec}_{\theta}(X - \Theta),
\]

where the distribution of \(\sqrt{n} \text{vec}_{\theta}(X - \Theta)\) is given in (A2.1). Using the fact that\(^2\)

\[Z \sim \text{EC}(\mu, \Omega, \psi) \rightarrow AZ \sim \text{EC}(A\mu, A\Omega A', \psi),\]

where \(A\) and \(Z\) are arbitrary matrices of conformable dimensions, we find that

\[
\sqrt{n} \text{vec}_{\theta}(g(X) - g(\Theta)) \sim \text{EC}
\left(0, \left[\frac{\partial \text{vec}_{\theta} g(\Theta)}{\partial \text{vec}_{\theta} \Theta} \right] \Omega \left[\frac{\partial \text{vec}_{\theta} g(\Theta)}{\partial \text{vec}_{\theta} \Theta} \right]', \phi\right).
\]

\(^2\) See theorem (2.16) in Fang, Kotz, and Ng (1990) p.43.
Appendix 3

Graphs of the Empirical Distributions of Estimates from the Static Model

Maximum Likelihood estimates of $\hat{\beta}_1, \nu=4$.

OLS estimates of $\hat{\beta}_1, \nu=4$.

Figure A3.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, \nu=4$. 
Maximum Likelihood estimates of $\beta_0$, $\nu=4$. OLS estimates of $\beta_0$, $\nu=4$.

Figure A3.1. 2. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=4$. 

50 observations

100 observations

200 observations

500 observations
Maximum Likelihood estimates of $\sigma^2$, $\nu=4$.

OLS estimates of $\sigma^2$, $\nu=4$.

Figure A3.1.3. Kernel density estimates of the empirical distribution of $\sigma^2$, $\nu=4$. 

50 observations

100 observations

200 observations

500 observations

50 observations

100 observations

200 observations

500 observations
Maximum Likelihood estimates of $(\Sigma_{22}^{-1})_{11}, \nu=4$.

Figure A3.1.4. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_{22}^{-1})_{11}$ and $(\hat{\Sigma}_{22}^{-1})_{21}, \nu=4$. 

Maximum Likelihood estimates of $(\Sigma_{22}^{-1})_{21}, \nu=4$.
\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{11} - \text{Maximum Likelihood, } \nu=4.

Figure A3.1.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{11}$ and $\hat{\sigma}^2(\hat{\Sigma}_{22}^{-1})_{21}$, $\nu=4$.  

50 observations

100 observations

200 observations

500 observations

50 observations

100 observations

200 observations

500 observations
Maximum Likelihood estimates of $\hat{\beta}_1, v=8$. 

OLS estimates of $\hat{\beta}_1, v=8$. 

Figure A3.2.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, v=8$. 
Maximum Likelihood estimates of $\beta_0$, $\nu=8$.

OLS estimates of $\beta_0$, $\nu=8$.

Figure A3.2. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=8$. 
Maximum Likelihood estimates of $\sigma^2$, $\nu=8$.

OLS estimates of $\sigma^2$, $\nu=8$.

Figure A3.2.3. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=8$.  

142
Maximum Likelihood estimates of \((\Sigma_{22}^{-1})_{11}, \nu=8\).

Figure A3.2.4. Kernel density estimates of the empirical distribution of \((\hat{\Sigma}_{22}^{-1})_{11}\) and \((\hat{\Sigma}_{22}^{-1})_{21}\) (ML), \(\nu=8\).
$\hat{\sigma}^2 (\Sigma^{-1}_{22})_{11}$ - Maximum Likelihood, $\nu = 8.$

$\hat{\sigma}^2 (\Sigma^{-1}_{22})_{21}$ - Maximum Likelihood, $\nu = 8.$

Figure A3.2.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\Sigma^{-1}_{22})_{11}$ and $\hat{\sigma}^2 (\Sigma^{-1}_{22})_{21}$ (ML), $\nu = 8.$
Appendix 4

Graphs of the Empirical Distributions of Estimates from the Dynamic Models

ML-A estimates of $\hat{\beta}_1, \nu=4$, (ASSTM).

OLS estimates of $\hat{\beta}_1, \nu=4$ (DLRM).

Figure A4.1.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, \nu=4$. 
ML-B estimates of $\hat{\beta}_1$, $\nu=4$ (STAR-B).

ML-S estimates of $\hat{\beta}_1$, $\nu=4$ (STAR-S).

Figure A4.1. 2. Kernel density estimates of the empirical distribution of $\hat{\beta}_1$, $\nu=4$. 
ML-A estimates of $\hat{\beta}_0$, $\nu=4$ (ASSTM).

OLS estimates of $\hat{\beta}_0$, $\nu=4$ (DLRM).

Figure A4.1. 3. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=4$. 
Figure A4.1.4. Kernel density estimates of the empirical distribution of $\hat{\beta}_0, \nu=4$. 

ML-B estimates of $\beta_0, \nu=4$ (STAR-B). 

ML-S estimates of $\beta_0, \nu=4$ (STAR-S).
ML-A estimates of $\sigma^2, \nu=4$ (ASSTM).

OLS estimates of $\sigma^2, \nu=4$ (DLRM).

Figure A4.1.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2, \nu=4$. 
ML-B estimates of $\sigma^2, \nu=4$ (STAR-B).

ML-S estimates of $\sigma^2, \nu=4$ (STAR-S).

Figure A4.1.6. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2, \nu=4$. 
Figure A4.1. 7. Kernel density estimates of the empirical distribution of \((\hat{\Sigma}_{22}^{-1})_{2,2}\), and \((\hat{\Sigma}_{22}^{-1})_{2,1}\), \(\nu=4\).
ML-B estimates of $(\Sigma_{ii}^{-1})_{ii}, \nu=4$ (STAR-B).

ML-B estimates of $(\Sigma_{ii}^{-1})_{ii-2}, \nu=4$ (STAR-B).

Figure A4.1. 8. Kernel density estimates of the empirical distribution of $(\hat{\Sigma}_T^{-1})_{ii}$ and $(\hat{\Sigma}_T^{-1})_{ii-2}, \nu=4$. 

50 observations

100 observations

200 observations

500 observations
ML-S estimates of \( (\Sigma_{22}^{-1})_{i,i} \), \( \nu=4 \) (STAR-S).

ML-S estimates of \( (\Sigma_{22}^{-1})_{i,i-2} \), \( \nu=4 \) (STAR-S).

Figure A4.1. 9. Kernel density estimates of the empirical distribution of \( (\hat{\Sigma}^{-1}_T)_{i,i} \) and \( (\hat{\Sigma}^{-1}_T)_{i,i-2} \), \( \nu=4 \).
Figure A4.1. 10. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,2}$ and $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{2,2}^2$, $\nu=4$. 
Figure A4.1. 11. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{ij}^{-1})_{ijkl} \cdot \text{ML-B, } \nu=4 \text{ (STAR-B)}$.

$\hat{\sigma}^2 (\hat{\Sigma}_{ij}^{-1})_{ijkl} - \text{ML-B, } \nu=4 \text{ (STAR-B)}$.
Figure A4.1. 12. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{11}^{-1})_{i,i}$ and $\hat{\sigma}^2 (\hat{\Sigma}_{11}^{-1})_{i,i-2}$, $\nu=4$. 
ML-A estimates of $\hat{\beta}_1, \nu=8$ (ASSTM).

OLS estimates of $\hat{\beta}_1, \nu=8$ (DLRM).

Figure A4.2.1. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, \nu=8$.  

50 observations

100 observations

200 observations

500 observations

50 observations

100 observations

200 observations

500 observations
ML-B estimates of $\hat{\beta}_1, \nu=8$ (STAR-B).

ML-S estimates of $\hat{\beta}_1, \nu=8$ (STAR-S).

Figure A4.2. 2. Kernel density estimates of the empirical distribution of $\hat{\beta}_1, \nu=8$. 

50 observations

100 observations

200 observations

500 observations

200 observations

500 observations

100 observations

50 observations
ML-A estimates of $\beta_0$, $\nu=8$ (ASSTM).

OLS estimates of $\beta_0$, $\nu=8$ (DLRM).

Figure A4.2.3. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=8$. 
ML-B estimates of $\hat{\beta}_0$, $\nu=8$ (STAR-B).

ML-S estimates of $\hat{\beta}_0$, $\nu=8$ (STAR-S).

Figure A4.2.4. Kernel density estimates of the empirical distribution of $\hat{\beta}_0$, $\nu=8$. 
ML-A estimates of $\sigma^2, \nu=8$ (ASSTM).

OLS estimates of $\sigma^2, \nu=8$ (DLRM).

Figure A4.2.5. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2, \nu=8$. 
ML-B estimates of $\sigma^2, \nu=8$ (STAR-B).

ML-S estimates of $\sigma^2, \nu=8$ (STAR-S).

50 observations

100 observations

200 observations

500 observations

Figure A4.2. 6. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2$, $\nu=8$. 
ML-A estimates of $(\Sigma_{22}^{-1})_{2,2}, \nu=8$ (ASSTM).

ML-A estimates of $(\Sigma_{22}^{-1})_{2,1}, \nu=8$ (ASSTM).

Figure A4.2. 7. Kernel density estimates of the empirical distribution of $(\widehat{\Sigma}_{22}^{-1})_{2,2}$ and $(\widehat{\Sigma}_{22}^{-1})_{2,1}, \nu=8$. 

50 observations

100 observations

200 observations

500 observations
ML-B estimates of \((\Sigma_{22}^{-1})_{ii}, \nu=8\) (STAR-B).

ML-B estimates of \((\Sigma_{22}^{-1})_{i,i-2}, \nu=8\) (STAR-B).

Figure A4.2. 8. Kernel density estimates of the empirical distribution of \((\Sigma_{11}^{-1})_{ii}\) and \((\Sigma_{11}^{-1})_{i,i-2}, \nu=8\).
ML-S estimates of \((\Sigma_2^{-1})_{ii}, \nu=8\) (STAR-S).

50 observations

100 observations

200 observations

500 observations

ML-S estimates of \((\Sigma_2^{-1})_{i,i-2}, \nu=8\) (STAR-S).

50 observations

100 observations

200 observations

500 observations

Figure A4.2. 9. Kernel density estimates of the empirical distribution of \((\hat{\Sigma}_1^{-1})_{ii}\) and \((\hat{\Sigma}_1^{-1})_{i,i-2}, \nu=8\).
\[ \hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{2,2} \text{ - ML-A, } \nu=8 \text{ (ASSTM)}. \]

Figure A4.2.10. Kernel density estimates of the empirical distribution of \( \hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{2,2} \) and \( \hat{\sigma}^2 (\hat{\Sigma}^{-1}_{22})_{2,1}, \nu=8. \)
Figure A4.11. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 \left( \hat{\Sigma}_{i}^{-1} \right)_{i,i}$ and $\hat{\sigma}^2 \left( \hat{\Sigma}_{i}^{-1} \right)_{i,i-2}$, $\nu=8$. 

For 50, 100, 200, and 500 observations.
$\hat{\sigma}^2 (\hat{\Sigma}_{11}^{-1})_{11} - \text{ML-S}, \nu=8$ (STAR-S).

$\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{22} - \text{ML-S}, \nu=8$ (STAR-S).

50 observations

100 observations

200 observations

500 observations

Figure A4.12. Kernel density estimates of the empirical distribution of $\hat{\sigma}^2 (\hat{\Sigma}_{11}^{-1})_{11}$ and $\hat{\sigma}^2 (\hat{\Sigma}_{22}^{-1})_{22}$, $\nu=8$. 

170
Appendix 5

Descriptive Statistics for the Empirical Distributions

of the Estimated Standard Errors
Table A5.1. Selected descriptive statistics for the estimated standard errors of $\beta_1$ and $\beta_2$ - Maximum Likelihood.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>ML-A (ASSTM)</th>
<th>ML-B (STAR-B)</th>
<th>ML-S (STAR-S)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean SE</td>
<td>SD</td>
<td>Sample</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>0.1665</td>
<td>0.0096</td>
<td>-0.2560</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1150</td>
<td>0.0044</td>
<td>-0.2749</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0807</td>
<td>0.0023</td>
<td>-0.9111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0508</td>
<td>0.0009</td>
<td>-0.0147</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>0.1582</td>
<td>0.0077</td>
<td>-0.0710</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1109</td>
<td>0.0037</td>
<td>-0.3587</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0778</td>
<td>0.0018</td>
<td>-0.1976</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0490</td>
<td>0.0007</td>
<td>-0.2348</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>0.1530</td>
<td>0.0068</td>
<td>-0.1598</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1076</td>
<td>0.0031</td>
<td>-0.3216</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0759</td>
<td>0.0016</td>
<td>-0.1796</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0479</td>
<td>0.0006</td>
<td>-0.2030</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$\nu=4$</td>
<td>50</td>
<td>0.1637</td>
<td>0.0105</td>
<td>0.0409</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1140</td>
<td>0.0044</td>
<td>-0.0520</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0802</td>
<td>0.0020</td>
<td>-0.3970</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0504</td>
<td>0.0007</td>
<td>-0.1102</td>
</tr>
<tr>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>0.1557</td>
<td>0.0088</td>
<td>0.0228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1097</td>
<td>0.0036</td>
<td>-0.1014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0774</td>
<td>0.0016</td>
<td>-0.1659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0488</td>
<td>0.0006</td>
<td>-0.2220</td>
</tr>
<tr>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>0.1508</td>
<td>0.0077</td>
<td>-0.1978</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1069</td>
<td>0.0032</td>
<td>-0.2825</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.0755</td>
<td>0.0015</td>
<td>-0.4863</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0477</td>
<td>0.0006</td>
<td>-0.3266</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set 500 estimates were used to obtain the numbers reported in the table.
Table A5.2. Selected descriptive statistics for the estimated standard errors of $\beta_1$ and $\beta_2$ - OLS and HCSE.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>DF</th>
<th>Sample Size</th>
<th>OLS Standard Errors</th>
<th>HCSEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean SE</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td></td>
<td>$\nu=4$</td>
<td>50</td>
<td>0.1425</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0698</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>0.1428</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0698</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0440</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>0.1429</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0698</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0440</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td></td>
<td>$\nu=4$</td>
<td>50</td>
<td>0.1418</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=6$</td>
<td>50</td>
<td>0.1417</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu=8$</td>
<td>50</td>
<td>0.1421</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.0991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.0697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.0439</td>
</tr>
</tbody>
</table>

* Size of the Student $t$ data set used in estimation. For each size of the data set 500 estimates were used to obtain the numbers reported in the table.
Vita

Remigiusz Paczkowski was born on April 19, 1964 in Kamienna Gora, Poland. In October 1985 he enrolled at the Nicholas Copernicus University in Torun, Poland. He graduated in June 1990 with the degree of Master of Business Administration. That same year he enrolled in the Department of Economics at Virginia Polytechnic Institute and State University. In May 1993 he received the Master of Arts degree in Economics. He continued his education at the Department of Agricultural and Applied Economics at Virginia Tech, where he received the Doctor of Philosophy degree in December 1997.