Modelling Nonlinear Vector Economic Time Series
Modelling Nonlinear Vector Economic Time Series

By Yukai Yang

A dissertation submitted to
Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Management
To my family
“Everything is related with every other thing, and this relation involves the emergence of a relational quality. The qualities cannot be known a priori, though a good number of them can be deduced from certain fundamental characteristics.”

---Jaina philosophy

The Jaina Philosophy of Non-Absolutism by S. Mookerjee, q.v. Mahalanobis(1957)
Preface

This PhD thesis was written in the period from January 2009 to December 2011 during my studies at the Center for Research in Econometric Analysis of Time Series (CREATES), School of Economics and Management, Aarhus University. I am grateful to CREATEES, funded by the Danish National Research Foundation, for providing excellent research facilities and a stimulating environment.

A number of people have contributed to the making of this thesis. First of all, I would like to thank my thesis advisor Timo Teräsvirta for the excellent support and guidance. I really appreciate his patience, encouragement and inspiration to me. I have benefited greatly from his constructive comments and suggestions during our discussions. I will cherish the time period in which I collaborated with him. I would also like to thank Niels Haldrup for his encouragement and stimulation. I am also indebted to a number of people who have given many valuable comments and suggestions, especially Søren Johansen and Christian M. Dahl.

I would like to thank my colleagues. Special thanks go to my friends Niels Husted Dall-Hansen, Anders Bredahl Kock and Laurent Callot for inviting me to play American football in the campus. I would also like to thank my office mate Kenneth Dencker Petersen with whom I have shared a lot of non-academic discussions.

I would like to thank my wife Jingying for her love and encouragement. I would also like to thank my parents, my sisters and all my family members for understanding and support through these years.

Yukai Yang, Aarhus, December 2011
Updated preface

The pre-defence took place on 17 February 2012 in Aarhus. I would like to express my greatest gratitude to the members of the assessment committee: Dick van Dijk, Rickard Sandberg and Henning Bunzel, for their careful reading of the dissertation and their useful comments and suggestions. Most of the suggestions have been incorporated into the present version of the dissertation. Thank you very much!

*Yukai Yang, Louvain-La-Neuve, May 2012*
## Contents

Preface i  
Summary ix  
List of tables xiii  
List of figures xv  

1 Vector smooth transition autoregressive models 1  
\hspace{1cm} 1.1 Introduction 2  
\hspace{1cm} 1.2 The statistical framework 4  
\hspace{1cm} \hspace{1cm} 1.2.1 The vector logistic STAR model 4  
\hspace{1cm} \hspace{1cm} 1.2.2 Restrictions on linear parameters 7  
\hspace{1cm} 1.3 Modelling strategy 8  
\hspace{1cm} 1.4 Specification of the Logistic Vector STAR model 9  
\hspace{1cm} 1.5 Estimation of parameters 11  
\hspace{1cm} \hspace{1cm} 1.5.1 Nonlinear least squares estimation 11  
\hspace{1cm} \hspace{1cm} 1.5.2 Maximum likelihood estimation 13  
\hspace{1cm} \hspace{1cm} 1.5.3 Asymptotic properties of the estimators 14  
\hspace{1cm} \hspace{1cm} 1.5.4 Grid with a zoom 15  
\hspace{1cm} 1.6 Evaluation 16  
\hspace{1cm} \hspace{1cm} 1.6.1 Constancy of the error covariance matrix 16  
\hspace{1cm} \hspace{1cm} 1.6.2 Stability of the system 18  
\hspace{1cm} 1.7 Heteroskedasticity-robust tests 18  
\hspace{1cm} 1.8 Empirical applications 19  
\hspace{1cm} \hspace{1cm} 1.8.1 Gasoline price and consumption 19  
\hspace{1cm} \hspace{1cm} 1.8.2 Icelandic river flow 22  
\hspace{1cm} 1.9 Concluding remarks 28  
Appendix A: The two equivalent forms of the maximum likelihood estimator 29  
Appendix B: Existence and consistency of the NLS and ML estimators 30
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables</td>
<td>32</td>
</tr>
<tr>
<td>Figures</td>
<td>37</td>
</tr>
<tr>
<td><strong>2 Linearity and misspecification tests</strong></td>
<td>59</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>60</td>
</tr>
<tr>
<td>2.2 The vector logistic smooth transition model</td>
<td>61</td>
</tr>
<tr>
<td>2.3 Joint linearity tests</td>
<td>62</td>
</tr>
<tr>
<td>2.3.1 Testing linearity against VLSTAR with single transition variable</td>
<td>62</td>
</tr>
<tr>
<td>2.3.2 Testing linearity against VLSTAR with different transition variables</td>
<td>65</td>
</tr>
<tr>
<td>2.3.3 Higher-order linearity tests</td>
<td>66</td>
</tr>
<tr>
<td>2.4 Linearity test statistic with improved size</td>
<td>68</td>
</tr>
<tr>
<td>2.5 Evaluation tests</td>
<td>70</td>
</tr>
<tr>
<td>2.5.1 Serial correlation in the error process</td>
<td>70</td>
</tr>
<tr>
<td>2.5.2 Additive nonlinearity</td>
<td>73</td>
</tr>
<tr>
<td>2.5.3 Parameter constancy</td>
<td>75</td>
</tr>
<tr>
<td>2.5.4 Curse of dimensionality</td>
<td>77</td>
</tr>
<tr>
<td>2.6 Simulation study</td>
<td>78</td>
</tr>
<tr>
<td>2.6.1 P-value plot and p-value discrepancy plot</td>
<td>78</td>
</tr>
<tr>
<td>2.6.2 Size experiments of linearity tests and misspecification tests of the evaluation</td>
<td>79</td>
</tr>
<tr>
<td>2.7 Concluding remarks</td>
<td>82</td>
</tr>
<tr>
<td>Appendix A: Proof of Theorem 2.3.1</td>
<td>83</td>
</tr>
<tr>
<td>Appendix B: LM test statistic against an additive component</td>
<td>83</td>
</tr>
<tr>
<td>Appendix C: Proof of Theorem 2.4.1</td>
<td>86</td>
</tr>
<tr>
<td>Appendix D: The first-order partial derivatives of $\Psi'Bx_t$</td>
<td>88</td>
</tr>
<tr>
<td>Figures</td>
<td>89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3 Testing constancy of the error covariance matrix</strong></td>
<td>101</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>102</td>
</tr>
<tr>
<td>3.2 The model</td>
<td>103</td>
</tr>
<tr>
<td>3.3 LM test statistic</td>
<td>104</td>
</tr>
<tr>
<td>3.4 Specifications for heteroskedastic residuals</td>
<td>107</td>
</tr>
<tr>
<td>3.5 Finite sample properties of the test</td>
<td>108</td>
</tr>
<tr>
<td>3.5.1 Size experiments</td>
<td>109</td>
</tr>
<tr>
<td>3.5.2 Power experiments</td>
<td>110</td>
</tr>
<tr>
<td>3.6 Concluding remarks</td>
<td>112</td>
</tr>
<tr>
<td>Appendix A: Proof for Theorem 3.3.1</td>
<td>114</td>
</tr>
<tr>
<td>Figures</td>
<td>116</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Bibliography</td>
<td>149</td>
</tr>
<tr>
<td><strong>R code</strong></td>
<td>153</td>
</tr>
<tr>
<td>Specification</td>
<td>153</td>
</tr>
<tr>
<td>Evaluation</td>
<td>167</td>
</tr>
<tr>
<td>Testing constancy of the error covariance matrix</td>
<td>176</td>
</tr>
</tbody>
</table>
Summary

This thesis is concerned with nonlinear vector economic time series modelling. It is well known that relationships between many economic variables are nonlinear, and that nonlinear models abound in economic theory. There are markets that do not clear because some variables, often prices, are sticky, which yields nonlinearity between quantity and price. Central banks may set bounds for the exchange rate, which implies that the relationship between the exchange rate and the underlying that determines its value is nonlinear. In labour markets, many economic theories about the behaviour of firms employing workers suggest asymmetric fluctuations in employment on the macroeconomic level. A number of these economic phenomena and theories have given rise to nonlinear econometric models. Economic theory may offer guidance as to the functional form of the relationship that the researcher wants to use for characterizing dynamic behaviour of a number of economic variables. However, this is not always the case, and many nonlinear relationships have to be found using the data.

The thesis consists of three self-contained chapters, one-single-authored, and two written with a co-author. In chapter 1, together with Timo Teräsvirta, we concentrate on the Vector Smooth Transition Regression model. Smooth transition autoregressive models have become popular and are frequently applied to economic time series data. Teräsvirta (1994) devised a strategy for building univariate STAR models, consisting of specification, estimation and evaluation of the model. Camacho (2004) proposed a modelling strategy for a bivariate STAR model and applied it to joint modelling of the US GDP and a leading indicator by the Conference Board.

In this thesis, Camacho’s model is generalized into a multivariate STAR model and extended in various ways. A reliable modelling strategy for determining whether a given relationship is nonlinear, what the nonlinearity looks like, and whether it is adequately described by a particular parametric model, is needed. We devise a modelling strategy consisting of specification, including testing linearity, estimation and evaluation of these models.

I consider the case where each equation can have its own transition variable controlling the nonlinear behavior. Linearity testing is then carried out equation by equation and, as
in Camacho (2004), the classical tests for univariate time series models can be used. But then, I also consider the interesting special case in which the model only contains a single transition variable. This means that the existing univariate linearity tests are generalized into a multivariate joint tests.

Another extension is that I allow for multiple regimes in my STAR model. Modelling proceeds from specific to general. First I test linearity. If it is rejected, I estimate a standard multivariate STAR model with a single transition. I evaluate the model using misspecification tests. If the results suggest adding another transition, I do that, estimate the extended model and evaluate it as before. The fact that my model can have more than one transition is another extension to the bivariate model introduced by Camacho (2004).

Maximum likelihood estimation of the parameters of the model is discussed, and the selection of starting-values for nonlinear estimation receives attention.

Checking the stability of the system is a necessary but sometimes neglected step in the evaluation of the STAR model in empirical applications, when weak stationarity is assumed. I have devised a computational method for checking stability of the estimated vector STAR model, which generalizes the method used in the univariate case.

Two applications demonstrate vector smooth transition modelling in practice. The first one is based on monthly U.S. gasoline price and consumption time series. The issue is possible asymmetry in the response of consumers to a price change. This has been considered in many articles. Using a vector STAR model, I find that the price-consumption relationship is nonlinear and illustrate the behaviour of the VLSTAR model using generalized impulse response functions that can be applied to studying dynamic properties of these models. This method has not been used in previous studies of the problem. The second example is based on the Icelandic river flow data that Tsay (1998) analyzed using the multivariate threshold model. I re-analyzed the same data time series using my vector smooth transition model. The model successfully captures the nonsynchronization in switching of regimes. My results show that the regime-switching of the two river flows, which are controlled by the temperature, behaves differently not only in smoothness but in turning-points as well. These are new findings which cannot be found using other modelling procedures.

Chapter 2, together with Timo Teräsvirta, is concerned with the linearity and misspecification tests in vector smooth transition autoregressive models. First, I develop Lagrange-multiplier type test statistics for the null hypothesis of a linear VAR model against the alternative hypothesis of a vector smooth transition model. These new tests are likely to suffer from the problem that the number of observations available does not suffice for asymptotic inference, which can lead to significant size distortion of the tests.
Summary

This means that they are not reliable in small and moderate samples. For this reason, small-sample properties of the tests have to be carefully studied. We suggest and investigate two test statistics with improved size properties: Wilks’s Λ and Rao’s F-statistic. My results show that they alleviate the problem of size distortion in standard test statistics.

There are many ways for an estimated vector STAR model to be misspecified. Finding out whether the model satisfies the assumptions under which it was estimated should be an integral part of a normal modelling procedure. We consider three misspecification tests for possible model extensions: the test of no serial correlation, the test of no additive nonlinearity and the parameter constancy test. They are either Lagrange multiplier or Lagrange multiplier type tests.

We generalize the univariate misspecification tests in Eitrheim and Teräsvirta (1996) to multivariate joint tests. As already mentioned, small-sample properties of the tests should be carefully studied due to the dimension of the models. It turns out in my simulations that Wilks’s Λ and Rao’s statistic efficiently remedy the size distortion problem present in standard LM and LM-type tests.

In Chapter 3, the nonlinearity in error covariance matrix is investigated. I consider multivariate (vector) time series models in which the error covariance matrix may be time-varying. A recent statistic for testing the constancy of the error covariance matrix can be found in Eklund and Teräsvirta (2007). They derive a family of test statistics against various types of misspecifications making the use of the constant conditional correlation framework of Bollerslev (1990). The motivation of this chapter is to develop a new multivariate heteroskedasticity test as an alternative to the one proposed in Eklund and Teräsvirta (2007).

Tests of constancy of the error covariance matrix against the alternative that the covariance matrix changes over time are developed based on the spectral decomposition of the error covariance matrix. The idea with this decomposition is to obtain tests against parsimoniously parameterised alternatives such that the resulting tests would be powerful against many kinds of departure from parameter constancy. A new family of Lagrange-multiplier type tests which allow for various types of misspecifications under this decomposition is developed.

As in Eklund and Teräsvirta (2007), three types of alternatives to constancy are considered in this chapter. The first one may be viewed as a multivariate generalization of the heteroskedasticity test of White (1980), and the second one generalizes the test against autoregressive conditional heteroskedasticity of Engle (1982). The third variant of the test generalizes the univariate constant variance test of Medeiros and Veiga (2003), in which it is assumed that under the alternative hypothesis the variance changes smoothly over time. It can be seen that the joint constancy test for error covariance matrix against
multivariate heteroskedasticity is very easy to implement and use. It has satisfactory size and power properties even in high-dimensional vector models. Furthermore, the test is still robust when the vector model is misspecified.
### List of Tables

1. **P-values of linearity tests for the VAR(2) model against the VLSTAR model in Section 1.8.1.** The upper part is the linearity tests in the gasoline price change equation, and the lower part is the linearity tests in the gasoline consumption change equation.  
3" represents the parsimonious test of order three.  
*tran. var.* stands for transition variables, WLK for Wilks’s statistics, RAO for Rao’s statistics and WB for wild bootstrapping. The sample size of the wild bootstrapping is 5000. Note that all of the p-values are multiplied by 100 and hence read percent. We denote \( \tau = t/T \), where \( t = 1, \ldots, T \).

2. **P-values of the evaluation tests of the estimated model (1.8.2).** GP stands for gasoline price change equation, and GC stands for gasoline consumption change equation. The tests of no additive nonlinearity are based on the third-order Taylor expansion. WLK stands for Wilks’s statistics, RAO for Rao’s statistics and WB for wild bootstrapping. The sample size of the wild bootstrapping is 1000. Notice that p-values are all multiplied by 100. We denote \( \tau = t/T \), where \( t = 1, \ldots, T \).

3. **P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using ARCH specification.** LM stands for Lagrange-multiplier test. Notice that p-values are all multiplied by 100.

4. **P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using smooth transition (ST) specification.** LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.

5. **P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using White specification.** LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.
List of Tables

1.6 p-values of misspecification tests for the VAR(15) model against additive nonlinearities.
   Jokul stands for the Jökulsá flow equation, and Vatns for the Vatnsdalsá flow equation.
   Only the wild bootstrap results from the linearity tests of order 3 are reported. t.v. stands for transition variables. y_{1,t-d} stands for the Jökulsá flow at time t – d,
   y_{2,t-d} stands for the Vatnsdalsá flow at time t – d, x_{t-d} for the precipitation at time t – d and z_{t-d} for the temperature at time t – d. The sample size of the wild bootstrapping is 5000. Note that all of the p-values are multiplied by 100 and hence read percent. 35

1.7 P-values of the joint serial correlation tests of the estimated model (1.8.4) and (1.8.5).
   WB stands for the wild bootstrapping. The sample size of the wild bootstrapping is 1000. Notice that p-values are all multiplied by 100. 35

1.8 P-values of the partial and joint parameter constancy tests of the estimated model (1.8.4) and (1.8.5). WB stands for the wild bootstrapping. The sample size of the wild bootstrapping is 1000. Notice that p-values are all multiplied by 100. 36

1.9 P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using ARCH specification. LM stands for Lagrange-multiplier test. Notice that p-values are all multiplied by 100. 36

1.10 P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using smooth transition (ST) specification. LM stands for Lagrange-multiplier test. τ = t/T. Notice that p-values are all multiplied by 100. 36

1.11 P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using White specification. LM stands for Lagrange-multiplier test. τ = t/T. Notice that p-values are all multiplied by 100. 36

2.1 The empirical size of the bootstrapping test. 67
List of Figures

1.1 Grid with a zoom .......................................................... 37

1.2 Top panel: monthly growth rate of gasoline price (log difference); Bottom panel: monthly
growth rate of gasoline consumption (log difference). The sample period is from Feb.
1973 to Dec. 1998, and 311 observations in all ....................................... 38

1.3 Autocorrelation functions (ACF) and partial autocorrelation function (Partial ACF) of
the residuals of the estimated model (1.8.2) for the monthly gasoline price change (GP)
and the monthly gasoline consumption change (GC) .................................. 39

1.4 Transition functions for the gasoline price change equation and the consumption change
equation: (Top) the first transition driven by \( y_{1,t-4} \) in the price equation, (Middle) the
second transition driven by \( y_{1,t-5} \) in the price equation, and (Bottom) the transition
driven by \( y_{1,t-1} \) in the consumption equation. Each point represents an observation. . 40

1.5 Regime-switching through time for the gasoline price change equation and the con-
sumption change equation: (Top) the first regime-switching driven by \( y_{1,t-4} \) in the price
equation, (Middle) the second regime-switching driven by \( y_{1,t-5} \) in the price equation,
and (Bottom) the regime-switching driven by \( y_{1,t-1} \) in the consumption equation. . 41

1.6 Stability of the two processes. The upper plot is the price change equation which
converges to about -0.302, and the lower plot is the consumption change equation which
converges to about 0.218. We use all the history in the data set as the initial values of
the DGP. .......................................................... 42

1.7 Generalized impulse response functions of the VLSTAR model (1.8.2) represented with
50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots:
gasoline price percent change to positive price shocks (left); gasoline price percent change
to negative price shocks (right). Bottom two boxplots: gasoline consumption
percent change to positive price shocks (left); gasoline consumption percent change to
negative price shocks (right). .................................................. 43
1.8 Generalized impulse response functions of the VLSTAR model (1.8.2) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: gasoline price percent change to positive consumption shocks (left); gasoline price percent change to negative consumption shocks (right). Bottom two boxplots: gasoline consumption percent change to positive consumption shocks (left); gasoline consumption percent change to negative consumption shocks (right).

1.9 River flow, precipitation and temperature series.

1.10 The regime-switching of the Jökulsá flow driven by the temperature.

1.11 The regime-switching of the Vatnsdalsá flow driven by the temperature.

1.12 The regime-switching of the Jökulsá flow driven by the precipitation.

1.13 The regime-switching of the Vatnsdalsá flow driven by the precipitation.

1.14 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Jökulsá shock below freezing (left) and above freezing (right).

1.15 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Jökulsá shock below freezing (left) and above freezing (right).

1.16 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Vatnsdalsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Vatnsdalsá shock below freezing (left) and above freezing (right).

1.17 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Vatnsdalsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Vatnsdalsá shock below freezing (left) and above freezing (right).

1.18 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Jökulsá shock below freezing (left) and above freezing (right).
List of Figures

1.19 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnshalsà to positive Jökulsà shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnshalsà to negative Jökulsà shock below freezing (left) and above freezing (right). ........................................ 55

1.20 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsà to positive Vatnshalsà shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsà to negative Vatnshalsà shock below freezing (left) and above freezing (right). ........................................ 56

1.21 Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnshalsà to positive Vatnshalsà shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnshalsà to negative Vatnshalsà shock below freezing (left) and above freezing (right). ........................................ 57

2.1 Size discrepancy plot for Design 1 ........................................ 89
2.2 Size discrepancy plot for Design 2 ........................................ 89
2.3 Size discrepancy plot for Design 3 ........................................ 90
2.4 Size discrepancy plot for Design 4 ........................................ 90
2.5 Size discrepancy plot for Design 5 ........................................ 91
2.6 Size discrepancy plot for Design 6 ........................................ 91
2.7 Size discrepancy plot for test of no serial correlation .................. 92
2.8 Size discrepancy plot for test of no additive nonlinearity ............. 93
2.9 Size discrepancy plot for test of parameter constancy ................. 94
2.10 The size discrepancy plot for the LM test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ........................................ 95
2.11 The size discrepancy plot for the rescaled F test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ........................................ 95
2.12 The size discrepancy plot for Wilks’s test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ........................................ 96
2.13 The size discrepancy plot for Rao’s test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ........................................ 96
2.14 The size discrepancy plot for the LM test: \( p = 2, k = 1 \) and \( \rho = 0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 97

2.15 The size discrepancy plot for the rescaled F test: \( p = 2, k = 1 \) and \( \rho = 0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 97

2.16 The size discrepancy plot for Wilks’s test: \( p = 2, k = 1 \) and \( \rho = 0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 98

2.17 The size discrepancy plot for Rao’s test: \( p = 2, k = 1 \) and \( \rho = 0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 98

2.18 The size discrepancy plot for the LM test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 99

2.19 The size discrepancy plot for the rescaled F test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 99

2.20 The size discrepancy plot for Wilks’s test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 100

2.21 The size discrepancy plot for Rao’s test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom). ............................................................................. 100

3.1 The size discrepancy plot: Gaussian errors \( p = 2, T = 100 \) and \( N = 10000 \). LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. ................................................................. 116

3.2 The size discrepancy plot: Gaussian errors \( p = 2, T = 500 \) and \( N = 10000 \). LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. ................................................................. 116

3.3 The size discrepancy plot: Gaussian errors \( p = 5, T = 100 \) and \( N = 10000 \). LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. ................................................................. 117
List of Figures

3.4 The size discrepancy plot: Gaussian errors $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. .......................................................... 117

3.5 The size discrepancy plot: $t(5)$ errors $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. .......................................................... 118

3.6 The size discrepancy plot: $t(5)$ errors $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. .......................................................... 118

3.7 The size discrepancy plot: $t(5)$ errors $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. .......................................................... 119

3.8 The size discrepancy plot: $t(5)$ errors $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region. .......................................................... 119

3.9 The size discrepancy plot of LM test against White specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. .......................................................... 120

3.10 The size discrepancy plot of LM test against ARCH specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. .......................................................... 121

3.11 The size discrepancy plot of LM test against smooth transition specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. .......................................................... 122

3.12 The size discrepancy plot of LM test against White specification: $t(5)$ errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. .......................................................... 123

3.13 The size discrepancy plot of LM test against ARCH specification: $t(5)$ errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. .......................................................... 124
3.14 The size discrepancy plot of LM test against smooth transition specification: \(t(5)\) errors 
\(p = 2, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: 
\(\rho = 0.9, 0, -0.9, \ldots \) .................................................... 125

3.15 The size discrepancy plot of LM test against White specification: Gaussian errors \(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: \(\rho = 0.9, 0\). 126

3.16 The size discrepancy plot of LM test against ARCH specification: Gaussian errors \(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: \(\rho = 0.9, 0\). 127

3.17 The size discrepancy plot of LM test against smooth transition specification: Gaussian errors \(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: \(\rho = 0.9, 0\). 128

3.18 The size discrepancy plot of LM test against White specification: \(t(5)\) errors \(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: \(\rho = 0.9, 0\). 129

3.19 The size discrepancy plot of LM test against ARCH specification: \(t(5)\) errors \(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: \(\rho = 0.9, 0\). 130

3.20 The size discrepancy plot of LM test against smooth transition specification: \(t(5)\) errors 
\(p = 3, T = 25, 50, 100, 250, 500, 1000\) from 1 to 6 and \(N = 10000\). From top to bottom: 
\(\rho = 0.9, 0, \ldots \) .................................................... 131

3.21 The power plot: \(\Sigma_t = P\Lambda_t P'\) with threshold change at \(T/2\), \(p = 2, T = 100\) and 
\(N = 10000\). LM test against smooth transition specification (solid), LM test against 
ARCH specification (dashed), LM test against White specification (dotted) and test in 

3.22 The power plot: \(\Sigma_t = P\Lambda_t P'\) with threshold change at \(T/2\), \(p = 2, T = 500\) and 
\(N = 10000\). LM test against smooth transition specification (solid), LM test against 
ARCH specification (dashed), LM test against White specification (dotted) and test in 

3.23 The power plot: \(\Sigma_t = P\Lambda_t P'\) with threshold change at \(T/2\), \(p = 5, T = 100\) and 
\(N = 10000\). LM test against smooth transition specification (solid), LM test against 
ARCH specification (dashed), LM test against White specification (dotted) and test in 

3.24 The power plot: \(\Sigma_t = P\Lambda_t P'\) with threshold change at \(T/2\), \(p = 5, T = 500\) and 
\(N = 10000\). LM test against smooth transition specification (solid), LM test against 
ARCH specification (dashed), LM test against White specification (dotted) and test in 

3.25 The power plot: \(\Sigma_t = P_t A P_t'\) with threshold change at \(T/2\), \(p = 2, T = 100\) and 
\(N = 10000\). LM test against smooth transition specification (solid), LM test against 
ARCH specification (dashed), LM test against White specification (dotted) and test in 
3.26 The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 134

3.27 The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 135

3.28 The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 135

3.29 The power plot: $\Sigma_t = PA_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 136

3.30 The power plot: $\Sigma_t = PA_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 136

3.31 The power plot: $\Sigma_t = PA_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 137

3.32 The power plot: $\Sigma_t = PA_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 137

3.33 The power plot: $\Sigma_t = P_t \Delta P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 138

3.34 The power plot: $\Sigma_t = P_t \Delta P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed). 138
3.35 The power plot: $\Sigma_t = P_t \Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.36 The power plot: $\Sigma_t = P_t \Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.37 The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.38 The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.39 The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.40 The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.41 The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.42 The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

3.43 The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).
Chapter 1

Specification, Estimation and Evaluation of Vector Smooth Transition Autoregressive Models with Applications
Chapter 1. Vector smooth transition autoregressive models

Specification, Estimation and Evaluation of Vector Smooth Transition Autoregressive Models with Applications

abstract

This work concentrates on the nonlinear vector model called the Vector Smooth Transition Regression model. The bivariate single-transition vector smooth transition regression model of Camacho (2004) is generalised to a multivariate and multitransition one. A modelling strategy consisting of specification, including testing linearity, estimation and evaluation of these models is constructed. Maximum likelihood estimation of the parameters of the model is discussed, and the selection of starting-values for nonlinear estimation receives attention. Evaluation by misspecification tests is carried out using tests derived in a companion paper. The use of the modelling strategy is illustrated by two applications. In the first one, the dynamic relationship between the US gasoline price and consumption is studied and possible asymmetries in it considered. The second application consists of modelling two well known Icelandic riverflow series, previously considered by many hydrologists and time series analysts.

1.1 Introduction

Regime-switching nonlinear models with an observable switch or transition variable have a rather long history. Quandt (1958, 1960) considered a model in which the coefficients of a linear model change at a certain value of an observable stochastic variable. This means that there is at some point an abrupt switch from one regression line to another. Bacon and Watts (1971) generalized this feature such that instead of a switch, the transition from one line to the other is smooth, hence the term 'smooth transition'. The univariate dynamic counterparts of these models appeared in the time series analysis a few years later. Tong (1978) and Tong and Lim (1980) introduced the threshold autoregressive (TAR) model, whereas Chan and Tong (1986) were the first to consider a smooth transition variant of it. Tsay (1989) constructed a systematic model building strategy for threshold autoregressive models, for an alternative approach, see Strikholm and Teräsvirta (2006) and Teräsvirta, Tjøstheim and Granger (2010, Section 16.4). Teräsvirta (1994) developed a similar strategy for smooth transition autoregressive (STAR) models. The latter work was completed by Eitrheim and Teräsvirta (1996) who derived misspecification tests for STAR models. A coherent modelling strategy for smooth transition regression (STR) models, including misspecification tests, appeared in Teräsvirta (1998). For

\footnote{This chapter is joint work with Timo Teräsvirta.}
Chapter 1. Vector smooth transition autoregressive models

a recent review, see Teräsvirta, Tjøstheim and Granger (2010, Chapter 3). For a thorough
treatment of univariate TAR models, see Tong (1990).

These models are single-equation models. The first nonlinear vector model with an observ-
able switch variable was the vector threshold autoregressive model (VTAR) that Tsay (1998)
introduced. The same threshold variable controlled the switch in each equation, and the thresh-
old parameter was also the same. Anderson and Vahid (1998) discussed testing the linear vector
autoregressive (VAR) model against a vector smooth transition model. Rothman, van Dijk and
Franses (2001) introduced a Vector Logistic Smooth Transition Error Correction (VLSTEC)
model to investigate the Granger-causality hypothesis between money, output, inflation and in-
terest rates. In their model, a single transition function controlled the transition in all equations.
Camacho (2004) considered a bivariate logistic smooth transition model with exogenous regres-
sors, in which each equation could have its own the transition variable. He devised a modelling
strategy, consisting of specification, estimation and evaluation stages, for building such models.

In this paper, the previous work is generalized in various ways. As in Rothman et al. (2001),
our Vector Logistic Smooth Transition Regression (VLSTR) model contains several equations.
It can also have more than one transition. Exogenous variables are allowed as in Camacho
(2004). Seasonality is introduced using seasonal dummies or trigonometric functions. The
VLSTR system can either have the same transition function for all equations or the transition
variable may vary from one equation to the next. In the latter case, linear equations are allowed
as well.

Following Teräsvirta (1994) and Camacho (2004), a complete modelling strategy is con-
structed for building VLSTR models. Linearity and misspecification tests when a single tran-
sition variable is controlling the transition in the whole system need special consideration and
are discussed in a companion paper, see Chapter 2. In addition to the misspecification tests
for the conditional mean, testing constancy of the error covariance matrix is considered as well.
Maximum likelihood estimation and the problem of finding initial values are discussed in detail.
Consistency and asymptotic normality of the maximum likelihood estimators is proved under
standard regularity conditions. Dynamic properties of estimated models are investigated using
generalized impulse response functions, see Koop, Pesaran and Potter (1996). How the mod-
elling strategy works is illustrated by applications to the relationship between gasoline price and
consumption in the US and modelling the daily flow of two Icelandic rivers. Tsay (1998) fitted
a bivariate VTAR model with exogenous variables, temperature and precipitation, to the same
pair of time series.

The plan of the paper is as follows. The VLSTAR model is introduced in Section 1.2 and the
modelling procedure in Section 1.3. The specification of the model is the topic of Section 1.4.
Parameter estimation by nonlinear least squares and maximum likelihood is considered in Sec-
tion 1.5 and model evaluation by misspecification tests in Section 1.6. Section 1.8 contains two
empirical applications. Final remarks can be found in Section 1.9. The relevant proofs are in
the Appendix.
Chapter 1. Vector smooth transition autoregressive models

1.2  The statistical framework

The linear Vector AutoRegressor model with $k$ lags (VAR($k$)) is defined as follows:

$$y_t = A_1' y_{t-1} + A_2' y_{t-2} + \ldots + A_k' y_{t-k} + \Phi' d_t + \varepsilon_t$$

where $F = (A_1', \ldots, A_k', \Phi')'$, is a $(kp + q) \times p$ matrix, $x_t = (y'_{t-1}, \ldots, y'_{t-k}, d_t')'$ is a $(kp + q) \times 1$ vector, $y_t$ is a $p \times 1$ column vector, and each $A_i$, $i = 1, \ldots, k$, is a $p \times p$ matrix. Furthermore, $d_t$ is a $q \times 1$ vector consisting of deterministic components such as intercepts, trends and seasonal dummies, and exogenous variables. Finally, $\Phi$ is a $q \times p$ matrix containing the coefficients of the elements of $d_t$. The $p \times 1$ error vector $\varepsilon_t$ is assumed i.i.d. $\mathcal{N}(0, \Omega)$, where $\Omega$ is positive definite.

Throughout the paper, matrices will be denoted by boldface capital letters, and vectors by lowercase boldface ones.

1.2.1  The vector logistic STAR model

We generalize (1.2.1) into the Vector Logistic Smooth Transition AutoRegressive (VLSTAR) model. This model has the following representation:

$$y_t = \left\{ \sum_{i=1}^{m} (G_i^{-1} - G_i^0)F_i' \right\} x_t + \varepsilon_t$$

(1.2.2)

where $F_i$ is a $(kp + q) \times p$ matrix: $F_i = (A_{i1}', \ldots, A_{ik}', \Phi_i')'$, $G_i^0$ is a diagonal matrix of transition functions:

$$G_i^0 = \text{diag} \{ g(s_{1i1} | \gamma_{i1}, c_{i1}), \ldots, g(s_{pi1} | \gamma_{ip}, c_{ip}) \}$$

(1.2.3)

for $i = 1, \ldots, m - 1$, and $G_i^0 = I_p$. $G_i^m = 0$. The diagonal elements of $G_i^0$ in (1.2.3) are logistic functions of their transition variables:

$$g(s_{ijt} | \gamma_{ij}, c_{ij}) = (1 + \exp \{ -\gamma_{ij} (s_{ijt} - c_{ij}) \})^{-1}, \quad \gamma_{ij} > 0$$

(1.2.4)

for $i = 1, \ldots, m - 1$ and $j = 1, \ldots, p$.

The function (1.2.4) is a continuous (for $\gamma_{ij} < \infty$), monotonically increasing sigmoid function of its argument $s_{ijt}$ and bounded between zero and one. We assume that the transition variable $s_{ijt}$ is a weakly stationary random variable, but it can also be a time trend: $s_{ijt} = t/T$, where $T$ is the number of observations; see for example He et al. (2009). Furthermore, $\gamma_{ij}$ is the slope parameter, or the smoothness parameter, determining the shape of the function, or the smoothness of the transition. The parameter $c_{ij}$ is a location parameter determining the midpoint of the transition. When $\gamma_{ij} = 0$, the corresponding equation becomes linear, and when $\gamma_{ij} \to \infty$, the transition in that equation becomes abrupt. In that case, when also $s_{11t} = \ldots = s_{pt1}$, $c_{j1} = \ldots = c_{jp}$ and $m = 2$, the resulting model is the multivariate TAR model of Tsay (1998).
As an example, consider the case where \( m = 2 \). Then (1.2.2) becomes

\[
y_t = \{(I_p - G_1^t)F_1' + G_1^tF_2'\}x_t + \varepsilon_t
\]  

(1.2.5)

so there is a single parameter shift in the model. In this case, each location parameter \( c_{1j} \), \( j = 1, ..., p \), has a straightforward interpretation. It represents the inflection point in which the transition function has value \( 1/2 \), i.e., one is halfway through the transition from \( F_1' \) to \( F_2' \) in the sense that in (1.2.5) the changing parameter matrix \( F_t' = (1/2)(F_1' + F_2') \). When \( p = 1 \), (1.2.5) is the univariate logistic STAR model of Teräsvirta (1994).

This type of regime-switching can be convenient for modelling, for example, structural breaks when the transition variable is the normalized time \( t/T \), or business cycle asymmetry where the regimes represent expansions and recessions, see Teräsvirta and Anderson (1992) and Skalin and Teräsvirta (2002) for empirical examples. We shall see in the following that the model is flexible in the sense that it can be easily extended or simplified.

The VLSTAR model defined in (1.2.2) has \( p \) different transition functions for each \( i = 1, ..., m - 1 \), and each one can have its own transition variable. If \( \gamma_{1i} = \gamma_{2i} = ... = \gamma_{pi} = \gamma_i \), \( c_{1i} = c_{2i} = ... = c_{pi} = c_i \), and \( s_{1it} = s_{2it} = ... = s_{pit} = s_{it} \), then the \( i \)th transition matrix is \( G_i^t = g(s_{it} | \gamma_i, c_i)I_p \). This may sometimes be a reasonable special case in applications and does reduce the size (complexity) of the model.

For the general case (1.2.2), (1.2.3) and (1.2.4), one may wonder whether the model is identified. And if not, under which conditions the model is identified. Consider the case when the transition variable \( s_{ijt} = s_{jt} \) is identical for equation \( j \). It is easy to find that the likelihood function is invariant for all permutations of the location parameters \( \left(c_{1j}, ..., c_{mj}\right) \). This is similar to the so-called ”label switching problem” discussed in mixture models by Redner and Walker (1984), see also Diebolt and Robert (1994) and Richardson and Green (1997). A solution to this problem is to restrict the location parameters \( c_{ij} \) in equation \( j \) to be in monotonically increasing order during the estimation.

This restriction should also be imposed when only two transition variables are identical in certain equation. The situation becomes even more complex when the transition variables differ across equations and transitions. One should be aware of the fact that for a rather complex model with high dimension and many transitions, the identification cannot be ensured in small sample size. For large sample case, Appendix B gives the proof that, if for each equation we restrict the location parameters of the same transition variables to be in monotonically increasing or decreasing order, the average residual sum of squares function and the average log-likelihood function have a unique optimizer in the limit \( T \to \infty \).

Therefore, our suggestions to avoid the unidentification problem are simply 1. in each equation, restrict the location parameters of the same transition variables to be in increasing order, 2. reduce the model as simple as possible, and 3. more samples are preferred.

The VLSTAR model can be reparameterised as follows:

\[
y_t = (B_1' + G_1^mB_2' + ... + G_1^{m-1}B_m')x_t + \varepsilon_t = \Psi_t'b'x_t + \varepsilon_t
\]  

(1.2.6)
where $\Psi_t = \begin{pmatrix} I_p, G_1^t, ..., G_m^t \end{pmatrix}'$ is a $mp \times p$ full rank matrix. Furthermore, $B = (B_1, B_2, ..., B_m)$ is a $(kp + q) \times mp$ matrix, where $B_1 = F_1$, and $B_i = F_i - F_{i-1}$, $i = 2, ..., m$.

The representation in (1.2.6) is equivalent to the one in (1.2.2). We see that the representation in (1.2.2) describes the transition through different extreme regimes $F_i$, $i = 1, ..., m$, while the reparametrized form in (1.2.6) is much easier for specification, estimation and evaluation, and hence will be employed hereafter. The special case (1.2.5) where $m = 2$ has the form

$$y_t = (B_1' + G_1^t B_2')x_t + \varepsilon_t$$

Camacho (2004) considered this representation for modelling bivariate time series.

Moreover, the number of regimes or transitions in every equation is not necessarily restricted to be the same. Suppose that equation $j$ has $m_j$ regimes. Let $m = \max(m_1, ..., m_p)$ and formulate the models (1.2.2) and (1.2.6). In (1.2.2), restrict all the $j^{th}$ columns in $F_i$ to be the same, for all $i \geq m_j$. And equivalently, in (1.2.6), restrict all the $j^{th}$ columns in $B_i$ to be zeros, for all $i > m_j$. In the following subsection, we will discuss the details about how to make such restrictions.

As $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \Omega)$, the model may be represented (up to a marginal distribution) by the multivariate conditional density function

$$f(y_t | F_{t-1}) = (2\pi)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (y_t - \Psi_t'B'x_t)' \Omega^{-1} (y_t - \Psi_t'B'x_t) \right\}$$

where $F_{t-1}$ is the conditioning information up to $t - 1$. The conditional log-likelihood function has the form

$$\log L(\theta) = -(Tp/2) \log 2\pi - (T/2) \log |\Omega|$$

$$-\frac{1}{2} \sum_{t=1}^{T} (y_t - \Psi_t'B'x_t)' \Omega^{-1} (y_t - \Psi_t'B'x_t)$$

(1.2.8)

The set of parameters to be estimated is $\theta = \{B, \Omega, \Gamma, C\}$, where $\Gamma = [\gamma_{ij}]$ and $C = [c_{ij}]$ contain the parameters in the transition functions.

Hereafter, we assume in addition that:

**Assumption 1.2.1.** The log-likelihood function $\log L(\theta)$ is second-order differentiable with respect to the parameters $\theta$.

**Assumption 1.2.2.** Both the first-order derivative and second-order derivative of the log-likelihood function $\log L$ with respect to the parameters $\theta \in \Theta$ are $L_T$-integrable.

**Assumption 1.2.3.** We assume that the following limiting information exists

$$\lim_{T \to \infty} T^{-1}E_T(i_T(\theta)) = i(\theta) < \infty$$

(1.2.9)

where $i_T(\theta) = -\partial^2 \log L/\partial \theta^2$.

These are the regularity conditions introduced by Feigin (1976). We will employ these assumptions in Chapter 2. Furthermore, we assume that:
Assumption 1.2.4. The sequence of $y_t$ in (1.2.2), $t = 1, ..., T$ is weakly stationary through time.

We see that under Assumption 1.2.4, the nonlinear cointegration model with error correction form is also allowed. The stability of the system after estimation will be discussed in Section 1.6.2.

We shall design a modelling strategy for VLSTAR models. It consists of three stages: specification, estimation and evaluation, and can be viewed as a multivariate generalisation of the one discussed in Teräsvirta (1998), van Dijk et al. (2002) or Teräsvirta, Tjostheim and Granger (2010, Chapter 16). We shall begin the modelling considerations by discussing specification issues relevant to the VLSTAR model. Estimation and evaluation will be considered thereafter.

1.2.2 Restrictions on linear parameters

In this section, we consider possible restrictions on the column space of the linear parameter $B$, for example, multiple switches between the two extreme regimes, and linearity of certain equations of the system.

In certain applications it is more appropriate to specify the transition function such that the extreme regimes associated with small and large absolute values of $s_t - c$ are identical. This can be achieved in a 2-regime model by using the exponential function

$$g(s_{jt} | \gamma_j, c_j) = 1 - \exp\{-\gamma_j (s_{jt} - c_j)^2\}, \gamma_j > 0$$

for applications see Michael et al. (1997), Sarantis (1999), and Peel et al. (2001), or the second-order logistic function

$$g(s_{jt} | \gamma_j, c_j) = (1 + \exp\{-\gamma_j (s_{jt} - c_j)(s_{jt} - c_j)\})^{-1}, \gamma_j > 0$$

as proposed in Jansen and Teräsvirta (1996). More generally, multiple switches between the two extreme regimes can be described by the general $n$th-order logistic transition function.

In the framework of the vector logistic STAR model, assuming multiple switches between the two extreme regimes can also be achieved by imposing restrictions on the parameter matrix $B$.

As an example, consider the 3-regime multivariate model

$$y_t = (B_1' + G_1^2 B_2' + G_2^2 B_3') x_t + \varepsilon_t = \Psi' B' x_t + \varepsilon_t$$

The assumption of multiple switches between two extreme regimes implies $B_2 + B_3 = 0$. This restriction could be written as $BR = 0$ where $R = (0_p, I_p, -I_p)'$. Equivalently, we have $B = \bar{B}H'$, where

$$\bar{B} = (\bar{B}_1, \bar{B}_2), \quad H = \begin{pmatrix} I_p & 0 \\ 0 & I_p \\ 0 & -I_p \end{pmatrix}.$$
Matrix $\bar{B}$ is a $(kp + q) \times 2p$ new linear parameter matrix, and $H$ is known and satisfies $H \perp R$.

Multiple switches between the two extreme regimes in the $j^{th}$ equation of the system imply that the sum of the $j^{th}$ column in $B_2$ and the $j^{th}$ column in $B_3$ equals a null vector. This is achieved by setting $r = (0, ..., 0, 1, 0, ..., 0, -1, 0, ...)'$ in $Br = 0$, where the $(p + j)^{th}$ and the $(2p + j)^{th}$ elements equal 1 and -1, respectively, and the rest equal zero. We thus have the reparameterization $B = \bar{B}H'$ where $H$ has dimension $mp \times (mp - 1)$ and $H \perp r$. More generally, if we have $n$ restrictions $r_1, r_2, ..., r_n$, these could be combined to form $R = (r_1, r_2, ..., r_n)$. Constructing the corresponding matrix $H$ is straightforward. This yields

$$y_t = \Phi'_t H\bar{B}'x_t + \varepsilon_t \quad (1.2.10)$$

In the previous subsection, we have shown that our model allows for the case when the number of regimes or transitions is not necessarily to be the same for each equation. That is, in (1.2.6), restrict all the $j^{th}$ columns in $B_i$ to be zeros, for all $i > m_j$. Making restrictions that the $j^{th}$ column in $B_i$ is equal to 0, where $i > m_j$, can be achieved by setting the corresponding $r = (0, ..., 0, 1, 0, ..., 0, 1, 0, ...)'$ in $Br = 0$, where the $(ip+j)^{th}$ elements equal 1 for all $m_j < i < m$, and the rest equal zero. Thus, by constructing the corresponding matrix $H$, we can also write the formula (1.2.10). Moreover, notice that if the $j^{th}$ columns in $B_i$, for all $i = 2, ..., m$, are null vectors, the $j^{th}$ equation becomes linear. In this case, the corresponding parameters $\gamma_{ij}$ and $c_{ij}$ are unidentified nuisance parameters. In order to avoid this problem, a practical solution is to simply set $g(s_{jt} | \gamma_{ij}, c_{ij}) \equiv 1$ in (1.2.3).

### 1.3 Modelling strategy

Modelling stationary vector time series with VLSTAR models is carried out in a systematic fashion. The modelling strategy may be divided into three stages: specification, estimation, and evaluation. Corresponding procedures for single-equation models exist and have been successfully applied; see for example Box and Jenkins (1970) for ARIMA models, Tsay (1989) and Teräsvirta, Tjøstheim and Granger (2010, Section 16.4) for threshold autoregressive and switching regression models, and Teräsvirta (1998) or Teräsvirta, Tjøstheim and Granger (2010, Section 16.3) for smooth transition regression models. Camacho (2004) designed a modelling strategy for bivariate STAR models, and we generalise it to our family of VLSTAR models. We also consider the special case in which all equations are assumed to have the same transition variable. Tsay (1998) makes this assumption for his vector threshold autoregressive (VTAR) model. In that model, even the whole transition function, i.e., both the threshold variable and the threshold, is the same for all equations of the model.

Estimating a linear VAR model is the first stage in specifying of the VLSTAR model. This involves selecting the lag length for the VAR model. Specification consists of testing the linear VAR model against VLSTAR one and, if linearity is rejected, determining the structure of the VLSTAR model. This implies selecting the transition variable(s) and determining the lag structure of the model. The latter means reducing the size of the model by imposing appropriate
Chapter 1. Vector smooth transition autoregressive models

parameter restrictions. The way linearity is tested depends on the assumptions made about the system. If it is assumed that the VLSTAR model only has a single transition variable, that is, it is the same for all equations, a joint test involving the whole model can be applied. If this assumption is not made, testing and transition variable selection may be carried out equation by equation as in Camacho (2004), see Luukkonen, Saikkonen and Teräsvirta (1988) and Teräsvirta (1994, 1998).

Reducing the size of the model already involves parameter estimation. The parameters of the VLSTAR model are estimated using nonlinear least squares. Numerical aspects of this estimation problem will be considered later. In the linear VAR case, necessary and sufficient conditions for the least-squares estimators to be consistent and asymptotically normal exist, but this is not the case for nonlinear VLSTAR model. Stability of the model is a necessary condition, but explicit conditions for stability of the general VLSTAR model do not seem to exist.

Evaluation of the model is done by checking (a necessary condition for) stability numerically and subjecting the estimated model to misspecification tests. These include testing the null hypothesis of no error autocorrelation, the null of no additive nonlinearity, and testing parameter constancy. Furthermore, constancy of the covariance matrix is tested as well. We now consider these three stages of model building and begin with specification.

1.4 Specification of the Logistic Vector STAR model

As already mentioned, specification involves testing linearity against the VLSTAR model and selecting the transition functions. Linearity testing requires a test of \( m = 1 \) against \( m = 2 \), where \( m \) is the number of ‘extreme states’ and \( m - 1 \) the number of transitions in the system. Before the estimation, we need to determine \( m \). Testing \( m = 1 \) against \( m = 2 \) is a part of the specification stage, while testing \( m = 2 \) against \( m > 2 \) belongs to the evaluation stage.

Since there may not be economic theory available to determine this parameter, its value has to be determined from the data. There is a statistical argument in favour of building the model ‘from specific to general’: if \( m \) is chosen too large, the model will contain unidentified nuisance parameters. This invalidates the asymptotic inference as the parameters of the model cannot be consistently estimated. This is a well known problem, first formulated and discussed by Davies (1977, 1987), see also Watson and Engle (1985), and, later, in the univariate STAR context, by Saikkonen and Luukkonen (1988), Luukkonen et al. (1988) and Teräsvirta (1994). A solution based on constructing the empirical null distribution of the test statistic can be found in Hansen (1996), see also Teräsvirta et al. (2010, Chapter 5) for discussion.

The choice of the transition variables for the VLSTAR model can in some cases be based on economic theory implications. More often, however, economic theory may suggest many potential transition variables. For example, the theory may not be explicit about which lag of a given variable to choose. In the univariate case, a common way is to conduct a linearity test for each potential transition variable and choose the one which produces the strongest rejection
measured in the $p$-value.

In the multivariate case we can conduct linearity tests equation by equation as in in Luukkonen et al. (1988) and Camacho (2004). For each equation, one may choose the transition variable that produces the smallest $p$-value in the test as in the univariate case. If linearity is not rejected for any transition variable in the set, the corresponding equation is assumed to be linear. However, in the multivariate case, if there are several very significant transition variable candidates for some equations, the question is which combination of transition variables to choose for the VLSTAR model. A joint test of linearity against a VLSTAR model in which different equations would have different (predetermined) transition variables would be useful.

For better understanding, consider the $p$-dimensional 2-regime logistic VSTAR model in (1.2.6):

\[ y_t = B_1' x_t + G_1 B_2' x_t + \varepsilon_t \]  

(1.4.1)

where the sequence $\{\varepsilon_t\}$ is \(i.i.d. \mathcal{N}(0, \Omega)\). The null hypothesis of linearity can be written as $H_0 : \gamma_j = 0, j = 1, \ldots, p$, and the alternative hypothesis is $H_1 : \text{at least one } \gamma_j > 0, j = 1, \ldots, p$. The basic idea of the joint test is to replace the transition function by first-order Taylor expansion. Such tests are developed in Chapter 2.

If it can be assumed \textit{a priori} that potential nonlinearity in the vector system is controlled by a single transition variable, this has to be taken into account in testing linearity. Economic theories or in some cases common sense may suggest this special case. A good example is the investigation of the evolution of different prices of an asset in different markets controlled by the difference between the prices, see Tsay (1998). In that paper, the price difference is the sole transition variable. A joint linearity test against VLSTAR with a single transition variable is therefore advisable, as it makes it possible to control the (asymptotic) size of the test for the whole system. The joint test with a single transition variable is just a special case of the the joint test with different transition variables, see Chapter 2.

Our suggestion for specifying a VLSTAR model with a single transition variable is to conduct the joint linearity tests with each potential transition variable. To do this, we use the joint tests developed in Chapter 2. If none of these tests rejects linearity, one can retain the linear VAR model. On the other hand, if the null is rejected for at least one transition variable, we choose the one producing the smallest $p$-value. Furthermore, we test each equation separately using the selected transition variable in order to find out whether some equations are in fact linear or not.

It is interesting to consider the case in which the vector system has different transition variables for each equation but we test the linearity jointly using a single transition variable. Does the joint test still have power? To answer this question, consider a single-transition bivariate VLSTAR model with a different transition variable for each equation. It can be written equivalently as a double-transition VLSTAR model with a single transition variable for each nonlinear component. From this, it is clear that a joint test using either transition variable is a joint test for the whole system with neglected nonlinearity.
1.5 Estimation of parameters

In this section, we consider both the least squares estimation and the maximum likelihood estimation of the parameters in the VLSTAR model. These involve the design of efficient nonlinear optimization algorithms. In both the NLS estimation and the ML estimation, we employ the grid search algorithm to find a suitable starting value for the following numerical optimization. The basic principle of the grid search algorithm is to convert the objective function, average residual sum of squares function or the log-likelihood function, into a concentrated one.

1.5.1 Nonlinear least squares estimation

After selecting the transition variable \( s_t \) or \( S_t = \text{diag}(s_{1t}, \ldots, s_{pt}) \), assuming that the transition function takes the logistic form, we estimate the VLSTAR model by nonlinear least squares. The model \( (1.2.6) \) contains the parameters \( \theta = \{B, \Gamma, C\} \), where \( B = (B_1, B_2, \ldots, B_m) \), \( B_1 = F_1 \), and \( B_i = F_i - F_{i-1}, i = 2, \ldots, m \), \( \Gamma = \{\gamma_{ij}\} \), \( C = \{c_{ij}\}, i = 1, \ldots, m - 1 \) and \( j = 1, \ldots, p \). The NLS estimators are obtained by solving the optimization problem

\[
\hat{\theta}_{NLS} = \arg \min_{\theta} \sum_{t=1}^{T} (y_t - \Psi'_tB'x_t)'(y_t - \Psi'_tB'x_t).
\]  

(1.5.1)

In practice, finding the optimum may be problematic in the sense that the convergence to the optimum may be slow and the algorithm may converge to some local minimum which is not the global one. This is due to the shape of the sum of squares function in (1.5.1): the function can be rather "flat" in some directions and possess many local optima. For this reason, finding a suitable starting-value of \( \theta \) for the algorithm is crucial.

This can be done by means of the so-called "grid search" algorithm. The basic idea is to construct a discrete grid in the parameter space of \( \Gamma \) and \( C \) and estimate the parameters in \( B \) conditionally on each pair of \( \Gamma \) and \( C \) in the grid. For fixed \( \Gamma \) and \( C \), the model is linear in parameters. Choosing the pair of \( \Gamma \) and \( C \) which produces the smallest residuals sum of squares yields a starting-value for the nonlinear optimization. More specifically, the corresponding average residual sum of squares function, which can be also called the NLS objective function, is concentrated with respect to the linear \( B \) parameters. And therefore, the optimization is only performed with respect to the parameters \( \Gamma \) and \( C \).

The grid search works very well in the univariate case. It solves the nonlinear optimization problem successfully by converting it to a series of simple linear regressions using a discrete grid, while the dimension of parameter space of \( \Gamma \) and \( C \), that is, the dimension of the nonlinear estimation problem, is low. When \( m = 2 \), there are only two nonlinear parameters in the model.

In the vector STAR model, consider fixing the parameters \( \Gamma \) and \( C \) in (1.5.1). Set \( Q_T = \sum_{t=1}^{T} (y_t - \Psi'_tB'x_t)'(y_t - \Psi'_tB'x_t) \). For a fixed \( \Psi_t \), the conditional minimizer of \( Q_T \) can be obtained by solving the first-order condition equations

\[
\sum_{t=1}^{T} x_t (y_t - \Psi'_tB'x_t)' \Psi'_t = 0.
\]
Chapter 1. Vector smooth transition autoregressive models

It follows that
\[ \sum_{t=1}^{T} x_t y'_t \Psi'_t = \sum_{t=1}^{T} x_t' B \Psi_t \Psi'_t \] (1.5.2)

The equation (1.5.2) leads to the following closed form of the NLS estimator of \( B \) conditional on \( \Gamma \) and \( C \):

\[
\text{vec}(\hat{B})_{NLS} = \left[ T^{-1} \sum_{t=1}^{T} (\Psi_t \Psi'_t) \otimes (x_t x'_t) \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} \text{vec} (x_t y'_t \Psi'_t) \right]
\] (1.5.3)

where \( \text{vec}(\cdot) \) is the vectorization operator, see for example Lütkepohl (1996).

Thus, given \( \Gamma \) and \( C \), we have the conditional NLS estimators:

\[
\text{vec}(\hat{B})_{NLS} = (M'M)^{-1} M' \text{vec}(Y')
\] (1.5.4)

\[
\hat{\Omega}_{NLS} = T^{-1} \hat{E}' \hat{E}
\] (1.5.5)

where \( M = (Y_1, Y_2, ..., Y_T)' \) is a \( Tp \times mp(kp + q) \) matrix, \( Y_t = \Psi_t \otimes x_t \) is an \( mp(kp + q) \times p \) matrix, \( \hat{E} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_T)' \) is a \( T \times p \) matrix, and \( \hat{\epsilon}_t = y_t - \Psi'_t \hat{B}_{NLS}' x_t \) is a column vector of residuals. Appendix A contains the detailed derivation and the proof of the equivalence of (1.5.3) and (1.5.4).

It is seen from (1.5.4) that \( T \geq m(kp + q) \) is a necessary condition for \( M'M \) to have full rank and, consequently, to ensure that both (1.5.3) and (1.5.4) have a unique solution. When estimating a restricted model such as (1.2.10), we replace \( \Psi_t \) by \( H' \Psi_t \).

Since the error covariance matrix \( \Omega \) does not enter the objective function \( Q_T \), it is convenient to estimate the parameters equation by equation. The corresponding estimation algorithm for each equation is the same as the one in univariate STAR models. For equation \( j \), we denote the corresponding parameters \( \Gamma_j, C_j \) and \( B_j \), and the corresponding residual sum of squares \( Q_{j,T} \). A discrete grid is constructed in the parameter space of \( \Gamma_j \) and \( C_j \). For fixed \( \Gamma_j \) and \( C_j \), we regress the \( j \)th equation and obtain the estimates of \( \hat{B}_{j,NLS} \). We still employ (1.5.5) to compute the NLS estimates of the error covariance matrix.

We apply the following algorithm for obtaining the starting-values for equation \( j \):

1. Construct a grid in the parameter space of \( \Gamma_j \) and \( C_j \).
2. For each pair of \( \Gamma_j \) and \( C_j \) in the grid, compute the corresponding \( \hat{B}_{j,NLS} \) and the corresponding residual sum of squares \( Q_{j,T} \).
3. Find the smallest \( Q_{j,T} \), and choose the corresponding pair of \( \Gamma_j \) and \( C_j \) as suitable starting values.

In constructing the grid, one has to choose combinations of \( \Gamma_j \) and \( C_j \) such that the corresponding transition functions display a sufficient amount of variation for the grid points to be relevant. For example, if the location parameter \( c \) in a logistic function is chosen outside the observed support of the transition variable, the corresponding slope parameter \( \gamma \) has to be small enough to compensate. A large \( \gamma \) would lead to a transition function that equals either zero or
one for all observed values of the transition function. In multiple-regime case, especially when \( m \) is large, \( \Gamma_j \) and \( C_j \) may contain a large number of parameters. How to construct a grid that works without a heavy computational effort is discussed in Section 1.5.4.

After selecting the starting-values, the parameters can be estimated using nonlinear least squares (NLS). In order to alleviate the computational burden, it is advisable to follow the suggestion of Sollis et al. (1999) made for univariate STAR models. The first iteration consists of re-estimating the parameters in \( \Gamma_j \) and \( C_j \), given the starting values for \( B_j \). This is done by NLS. Following this, the new value of \( B_j \), \( \hat{B}_j^{(2)} \) say, is calculated as in Step 2 of the above algorithm. In the next iteration \( \hat{B}_j^{(2)} \) is used to obtain new estimates for the parameters in \( \Gamma_j \) and \( C_j \). Iteration is continued until convergence. Dividing each iteration into these two components reduces the dimension of the nonlinear estimation problem and thus saves computation time. If the grid is dense, the initial step-length of the optimization algorithm must be sufficiently short so that optimization with a high probability leads to a local maximum which is the closest to the value found using the grid approach.

### 1.5.2 Maximum likelihood estimation

In this section, we develop an algorithm to obtain the Maximum Likelihood (ML) estimators based on the NLS ones. The ML estimator of \( \theta \) is obtained from the solution to the optimization problem

\[
\hat{\theta}_{ML} = \arg \max_{\theta} \log L(\theta)
\]

where \( \log L(\theta) \) is the log-likelihood function of VLSTAR model, which has been given in (1.2.8).

By fixing \( \Gamma \) and \( C \), we can get the conditional estimates of \( B \) by solving the first-order condition problem of the log-likelihood function. The first-order condition is:

\[
\sum_{t=1}^{T} x_t (y_t - \Psi_t' x_t)' \Omega_t^{-1} \Psi_t' = 0
\]

It follows that

\[
\sum_{t=1}^{T} x_t y_t' \Omega_t^{-1} \Psi_t' = \sum_{t=1}^{T} x_t x_t' B \Psi_t \Omega_t^{-1} \Psi_t'
\]

We have the closed form of the ML estimator of \( B \) conditional on \( \Gamma \), \( C \) and \( \Omega \):

\[
\text{vec}(\hat{B})_{ML} = \left[ T^{-1} \sum_{t=1}^{T} (\Psi_t \Omega_t^{-1} \Psi_t') \otimes (x_t x_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} \text{vec}(x_t y_t' \Omega_t^{-1} \Psi_t') \right]
\]

(1.5.6)

Thus, the ML estimator of \( B \) conditional on \( \Gamma \), \( C \) and \( \Omega \) can also be written as:

\[
\text{vec}(\hat{B})_{ML} = \left[ T^{-1} \sum_{t=1}^{T} (Y_t \Omega_t^{-1} Y_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} (Y_t \Omega_t^{-1} y_t) \right] = \left( M' (I_T \otimes \Omega^{-1}) M \right)^{-1} \left( M' (I_T \otimes \Omega^{-1}) \text{vec}(Y) \right),
\]

(1.5.7)

The ML estimator of \( \Omega \) is the one in (1.5.5). See Appendix A for details.
In contrast to the NLS estimator, the conditional ML estimator of $B$ is not only conditional on $\Gamma$ and $C$, but on $\Omega$ as well, which has to be taken into account. We cannot estimate $B$ equation by equation, and the ML estimation cannot be conducted equation by equation. We have to use either (1.5.6) or (1.5.7). We conduct the grid search algorithm to find a suitable starting-value for the subsequent optimization. And therefore, the corresponding grid is built in the space of $\Gamma$ and $C$.

In practice, however, (1.5.6) and (1.5.7) have different numerical implications in the sense that (1.5.6) calculates matrix products within loops, while (1.5.7) constructs the matrix $M$ and stacks elements within loops instead and calculates the matrix products outside the loops.

Our proposal is to first carry out the NLS estimation and use the NLS estimate of $\Omega$ as the input in the subsequent ML estimation. This results in the following algorithm for the ML estimation.

1. Carry out the NLS estimation, and obtain $\hat{\theta}_{NLS}$.
2. Set $\Gamma_0 = \hat{\Gamma}_{NLS}$, $C_0 = \hat{C}_{NLS}$, $\Omega_0 = \hat{\Omega}_{NLS}$.
3. Compute $\hat{B}_{ML}$ using formula (1.5.6) by setting $\Gamma = \Gamma_0$, $C = C_0$ and $\Omega = \Omega_0$. Solve $\hat{\theta}_{ML} = \arg \max \log L(\theta)$ numerically w.r.t. $\Gamma$ and $C$ given $\hat{B}_{ML}$, and obtain $\hat{\Gamma}_{ML}$ and $\hat{C}_{ML}$. Calculate $\hat{\Omega}_{ML}$.
4. Update $\Gamma_0 = \hat{\Gamma}_{ML}$, $C_0 = \hat{C}_{ML}$, $\Omega_0 = \hat{\Omega}_{ML}$.
5. Repeat steps 3 and 4 until convergence.

$\Gamma$ and $C$ may have a large number of parameters, and hence the corresponding grid may have a huge number of elements. We suggest to use a grid with a zoom, which will be discussed in the following subsection.

1.5.3 Asymptotic properties of the estimators

The purpose of this subsection is to set forth sufficient conditions for the consistency and asymptotic normality of both the NLS estimator and the ML estimator. First, we have the following lemma:

**Lemma 1.5.1.** Given the model (1.2.6), under Assumption 1.2.1 to 1.2.3, the limiting average residual sum of squares function has a unique global minimum at $\theta = \hat{\theta}$, and the limiting average log-likelihood function has a unique global maximum at $\theta = \hat{\theta}$, where $\hat{\theta}$ is the vector of true parameters.

**Proof.** See the first part of Appendix B. In Appendix B, we just focus on the ML estimator, since the NLS estimation is a special case when $\Omega = I$. \qed

We see that our model (1.2.6) is assumption (a) and Lemma 1.5.1 is actually assumption (b) in Jennrich (1969). Note that the vector of parameters $\theta$ here does not contain $\Omega$, but in Appendix B, $\Omega$ is included and of course consistent as well. Thus, we have the following theorem:
Chapter 1. Vector smooth transition autoregressive models

Theorem 1.5.2. Given the model (1.2.6) and Lemma 1.5.1, both the NLS estimator and the ML estimator are consistent.

Proof. See the second part of Appendix B or Theorem 6 in Jennrich (1969).

Theorem 1.5.3. Given the model (1.2.6) and Lemma 1.5.1, under Assumption 1.2.1 to 1.2.3, and assuming that the true parameter vector \( \hat\theta_j \) of the \( j \)th equation is an interior point of the support \( \Theta_j \) and the corresponding information matrix \( i(\hat\theta_j) \) is non-singular,

\[
\sqrt{T}(\hat\theta_{j,NLS} - \hat\theta_j) \xrightarrow{d} N(0, \hat\sigma_j^2 i^{-1}(\hat\theta_j)),
\]

(1.5.8)

where \( \hat\theta_{j,NLS} \) is the NLS estimator, \( \hat\theta_j \) is the true parameters, \( \hat\sigma_j^2 \) is the true variance of the errors in equation \( j \), and \( \xrightarrow{d} \) stands for convergence in distribution. And \( T^{-1}i_T(\hat\theta_{j,NLS}) \) is a consistent estimator of \( i(\hat\theta_j) \).


1.5.4 Grid with a zoom

As we have discussed, optimization may be cumbersome in practice due to the shape of the sum of squares function: the function can be rather “flat” in some directions and possess more than one local optimum. A well functioning grid search algorithm is essential for finding a suitable starting-value for the optimization algorithm. However, in VLSTAR models the number of nonlinear parameters \( \Gamma \) and \( C \) increases rapidly with the increase of the dimension \( p \) and the number of transitions \( m - 1 \). Hence, a conventional grid contains a large number of points and it takes much time to find a reasonable starting-value.

Our solution to this problem is to build grid with a zoom. Figure 1.1 gives an example. For simplicity we assume that there are only two nonlinear parameters in the model. The \( x \)-axis represents the support of the first parameter, and the \( y \)-axis that of the second parameter. For both parameters, we choose a moderate number of points to build a first grid. In Figure 1.1, their number equals five, when the lower and upper bound points are included. This yields a grid with twenty points. The grid points are joined by the dash-dotted black lines.

Suppose that point A in Figure 1.1 is the point which produces the smallest residuals sum of squares. We then build a smaller grid using neighbouring points of A as the new lower and upper bound. In Figure 1.1, there are again five points for each variable, and the dashed blue lines define the points (except the four corner points). We search for the smallest residuals sum of squares in the next round. Suppose that Point B in Figure 1.1 is the point which now produces the smallest residuals sum of squares and build another zoom grid whose points except four corner points are joined by solid red lines in Figure 1.1. Supposing Point C is the optimal point, the next zoom will be focused around C.

From this example, we see that the grid with a zoom is an ad hoc algorithm designed in order to improve the grid search algorithm in VLSTAR models. It splits the grid search algorithm into several sequential stages. At each stage, it reduces the number of points inside a grid, which
Chapter 1. Vector smooth transition autoregressive models

saves much of the computing time, and achieves accuracy to some extent by means of zooming. However, one can imagine that this algorithm may mislead in some special cases, for example the optimizer of the objective function is located inside a very steep area. During each stage, the grid is somewhat sparse. The grid with a zoom algorithm works well when the global optimizer is located within the neighbourhood of the sparsely distributed points in a grid. It will desirable to have some information about the true parameters, and build the grid with a zoom based on that. At each stage, zooming using more than one point can also be considered to avoid the possible misleading problem. Unfortunately, as it is the first time to employ this method in multivariate nonlinear models, until now the properties of this algorithm remain unclear and we leave this to future work.

1.6 Evaluation

There are many possible ways for a VLSTAR model to be misspecified. Finding out whether the model appears to satisfy the assumptions under which it was estimated should be an integral part of a valid modelling procedure. Our suggestion is to employ the three multivariate misspecification tests developed in Chapter 2. They are the test of no serial correlation test, the test of no additive nonlinearity test and the parameter constancy test. All of them are the extensions of the three univariate misspecification tests developed in Eitrheim and Teräsvirta (1996). The multivariate serial correlation test is derived from the autocorrelation VARMA form, see (Godfrey, 1988, pp. 117-118). The multivariate additive nonlinearity test is the just the joint linearity test for the case \( m > 2 \). The multivariate parameter constancy test is a special case of the joint linearity test in which the transition variable is the normalized time \( \tau = t/T \).

For the inspection of the residuals and residual autocorrelations, one may check the empirical autocorrelation function and the empirical partial autocorrelation function as well. Other model evaluation devices will be be discussed below.

1.6.1 Constancy of the error covariance matrix

Lütkepohl (2004) argues that the error variance of a model may be nonconstant even when the conditional mean is correctly specified. It would be desirable to have test at hand for investigating parameter constancy of the error covariance matrix in vector models. Our suggestion is to employ the tests developed in Eklund and Teräsvirta (2007).

Consider the following LVSTAR model

\[
y_t = \Psi' B' x_t + \epsilon_t
\]  

(1.6.1)

where the \( p \)-dimensional vector error process \( \{ \epsilon_t \} \) is a sequence of random variables with zero mean and conditional \((p \times p)\) covariance matrix \( \Omega_t = [\omega_{ij}] \), \( i, j = 1, \ldots, p \). More specifically, we assume that the way in which the covariance matrix evolves through time satisfies the following assumption
Chapter 1. Vector smooth transition autoregressive models

Assumption 1.6.1. The time-varying covariances $\omega_{ijt}$ have the form

$$\omega_{ijt} = \rho_{ij}(\omega_{iit}\omega_{jjt})^{1/2}, \quad i, j = 1, \ldots, p,$$

(1.6.2)

where $\rho_{ij}$ is the correlation coefficient of $\varepsilon_{it}$ and $\varepsilon_{jt}$.

As is seen from (1.6.2), the tests in Eklund and Teräsvirta (2007) are based on the assumption of constant correlations. The simulations in Eklund and Teräsvirta (2007) showed that the test is quite robust against the case where only the correlations are time-varying.

The null hypothesis to be tested is

$$H_0: \quad \omega_{iit} = \omega_{ii}, \quad i = 1, \ldots, p.$$  

(1.6.3)

There are a number of alternatives as far as the alternative hypothesis is concerned. We only consider the case where the variances change smoothly over time. The time-varying variances are

$$\omega_{iit} = h_i(\sigma_i^2 + \lambda_i F_i(s_{it})), \quad i = 1, \ldots, p,$$

(1.6.4)

where $\lambda_i$ is a scalar parameter which satisfies $\min_{s_{it}}[\sigma_i^2 + \lambda_i F_i(s_{it})] > 0$ and $F_i(\cdot)$ a real-valued function of a transition variable $s_{it}$. Function $h_i(x)$, $i = 1, \ldots, p$, is a general positive-valued function and assumed to be at least second-order differentiable.

Under Assumption 1.6.1, the covariance matrix $\Omega_t$ can be written in the form

$$\Omega_t = D_t PD_t$$

(1.6.5)

where $D_t = \text{diag}(\omega_{11t}^{1/2}, \ldots, \omega_{ppt}^{1/2})$ and $P = [\rho_{ij}]$ is the corresponding correlation matrix. Notice that $\rho_{ii} = 1$.

Denote $\theta$ to be a vector of all the parameters in $D_t$ and $P$. Eklund and Teräsvirta (2007) gave the corresponding block of the \textit{average score vector} and the corresponding block of the inverse of the \textit{average population information matrix}:

$$s_T(\theta) = T^{-1} \sum_{t=1}^T \left\{ \frac{\partial \text{vec}(D_t^{-1})'}{\partial \theta} \text{vec} \left( D_t - \frac{1}{2} \varepsilon_i \varepsilon_i' D_t^{-1} P^{-1} - \frac{1}{2} P^{-1} D_t^{-1} \varepsilon_i \varepsilon_i' \right) \right\},$$

$$+ \frac{1}{2} \frac{\partial \text{vec}(P)'D_t^{-1}P^{-1}D_t^{-1}P^{-1}}{\partial \theta} \left\{ D_t \otimes D_t + \frac{1}{2} (P^{-1} \otimes \Omega_t) + \frac{1}{2} (\Omega_t \otimes P^{-1}) \right\},$$

(1.6.6)

$$I_T(\theta) = T^{-1} \sum_{t=1}^T \left\{ \frac{\partial \text{vec}(D_t^{-1})'}{\partial \theta} - \frac{1}{2} \frac{\partial \text{vec}(D_t^{-1})'}{\partial \theta} (D_t \otimes P^{-1} + P^{-1} \otimes D_t) \frac{\partial \text{vec}(P)}{\partial \theta} - \frac{1}{2} \frac{\partial \text{vec}(P)'}{\partial \theta} \right\} - \frac{1}{2} \frac{\partial \text{vec}(P)'}{\partial \theta} \left( D_t \otimes P^{-1} + P^{-1} \otimes D_t \right) \frac{\partial \text{vec}(D_t^{-1})}{\partial \theta} \right\}. \quad (1.6.7)$$

The vector model is estimated under the null hypothesis. The score $s_T(\bar{\theta})$ and the information matrix $I_T(\bar{\theta})$ are obtained by plugging in $\{\tilde{\varepsilon}_t\}_{t=1}^T \tilde{D}, \tilde{P}$ and $\tilde{\Omega}$. Notice that the subscript $t$ in $D_t$ and $\Omega_t$ are omitted under the null, since the variances are constant. The LM test statistic

$$LM = T s_T(\theta)'^{-1} s_T(\theta)$$

(1.6.8)
is asymptotically $\chi^2$ distributed under the null hypothesis, of which the degrees of freedom is the number of restrictions.

### 1.6.2 Stability of the system

The VLSTAR models are built by assuming weak stationarity of the underlying vector process, see Assumption 1.2.4. If the model is not stable in some sense, this assumption cannot hold. In the linear VAR case, the necessary and sufficient conditions for exponential stability and weak exogeneity coincide. There are no necessary and sufficient conditions for weak stationarity of the vector LSTAR models (such conditions are lacking for the general univariate STAR models as well). A naive approach for checking weak stationarity is that one can check the roots of all extreme regimes and see if they are all outside the complex unit disk. However, we do not suggest that, because our model is rather flexible. To see this, consider that, for example, our model allows one equation enters another regime while the other equations remain the same regime as before. The situation can be very complex such that all the roots are outside the unit disk, but still the system is unstable.

Exponential stability may be considered a necessary condition for weak stationarity, which makes checking stability of estimated VLSTAR models interesting. It may be studied numerically by generating paths of realisations by switching off the noise, starting from a number of initial points and seeing whether or not the paths of realisations converge. Convergence to a single stationary point is a necessary condition for exponential stability. Since the method is numerical, the conclusion following from repeated convergence to the same point can only be that the stability assumption is not contradicted by these calculations. In that case, the model may be respecified and re-estimated or abandoned. This diagnostic is employed in the examples of Section 7. The 'histories', i.e., sets of values of the time series in question are natural starting-values for stability calculations.

### 1.7 Heteroskedasticity-robust tests

Heteroskedasticity is a common feature in both financial and macroeconomic time series data. The presence of heteroskedasticity does matter the size of the tests developed in Chapter 2. From the results of the tests of the empirical applications in the following section, we will see that the tests are much more likely to reject the null of linearity, when the the covariance matrix is not constant over time. In such circumstances, it is important to have asymptotically valid tests that are reliable in finite samples.

A straightforward way to derive Lagrange multiplier type tests that are asymptotically heteroskedasticity-robust is to use heteroskedasticity-consistent covariance matrix estimates (HCCME) suggested in White (1980). Nevertheless, the findings in the Monte Carlo experiments in, for example Godfrey and Tremayne (2005), indicates that asymptotic critical values fail to give good control of finite sample significance levels of heteroskedasticity-robust versions
of the standard Lagrange multiplier tests. Godfrey and Tremayne (2005) also reported the results from the wild bootstrap method, which outperform the previous method in finite sample. Their argument is that computer-intensive methods like the wild bootstrap can help removing the size distortion caused by heteroskedasticity (even nonnormality).

In the following, we introduce the wild bootstrap method we will employ in the next section:

1. Estimate the model under the null hypothesis, obtain the corresponding estimates $\hat{\theta}$ and the residuals $\hat{\varepsilon}_t$, $t = 1, ..., T$ and compute the Lagrange multiplier test statistic.

2. Multiply each element of $\hat{\varepsilon}_t$ by $z_{jt}$, $j = 1, ..., p$, and obtain the new sequence of residuals $\tilde{\varepsilon}_t$, where $z_{jt}$ is independently drawn from a simple two-value distribution: $z_{jt} = 1$ with probability 0.5 and $z_{jt} = -1$ otherwise.

3. Given $\hat{\theta}$ and $\tilde{\varepsilon}_t$, $t = 1, ..., T$, generate the sequence of the new observations $y^*_t$ and compute the corresponding Lagrange multiplier test statistic.

4. Repeat step 2 and 3 $N$ times in order to estimate the $p$-values of the observed statistics. The null hypothesis is rejected for $p$-values that are sufficiently small.

### 1.8 Empirical applications

#### 1.8.1 Gasoline price and consumption

In this section, we give two illustrations to show how the VLSTAR modelling proceeds in practice. We begin by considering the interaction between the US real gasoline prices and gasoline consumption using a bivariate VLSTAR model. The time series are the monthly real gasoline prices and the monthly real gasoline consumption. The time period analysed extends from February 1973 to December 1998. A rather common, albeit not unanimous, view in the literature is that the effects of energy price shocks on some macroeconomic aggregates are asymmetric. In particular, energy price increases are perceived to have larger effects than energy price decreases. We want to find out whether this is also the case for the relationship between the gasoline price and the gasoline consumption.

This pair of time series has been analysed in Kilian and Vigfusson (2009). The original series are transformed into logarithms which are then differenced, so the two variables to be modelled are approximate monthly growth rates. Figure 1.2 shows the levels of both series. Let $y_{1,t}$ denote the real gasoline price change and $y_{2,t}$ the real gasoline consumption change. Set $y_t = (y_{1,t}, y_{2,t})'$.

In order to model the presumed asymmetry, they censored the price variable in the consumption equation such that it obtains value zero for negative values of the series, that is, $y_{1,t-d}^+ = \max(0, y_{1,t-d})$, given a suitable lag $d$. Following this, they estimated the following
Chapter 1. Vector smooth transition autoregressive models

multivariate VAR\((k)\) model

\[ y_{1t} = a_{10} + \sum_{i=1}^{p} a_{11,i} y_{1,t-d_i} + \sum_{i=1}^{p} a_{12,i} y_{2,t-d_i} + \varepsilon_{1t} \]

\[ y_{2t} = a_{20} + \sum_{i=0}^{p} a_{21,i} y_{1,t-d_i} + \sum_{i=1}^{p} a_{22,i} y_{2,t-d_i} + \sum_{i=0}^{p} f_{21,i} y_{1,t-d_i}^+ + \varepsilon_{2t} \]  

(1.8.1)

where \(E(\varepsilon_{1t}) = 0, E(\varepsilon_{2t}^2) = \sigma_i^2, i = 1, 2\), and \(E(\varepsilon_{1t}\varepsilon_{2t}) = 0\). The focus of Kilian and Vigfusson (2009) is on testing the hypothesis \(f_{21,i} = 0\) for \(i = 0, ..., p\) in (1.8.1). They are interested in possible asymmetry in the effect of a price change on consumption, and they find no compelling evidence against the symmetry.

Without presuming the asymmetry, we begin by fitting a VAR model to the series and testing linearity of the two equations against the VLSTAR model. In doing so, we treat lags of both variables as potential transition variables, as we do not know in advance which variable controls potential nonlinearity. It appears that \(k = 2\) is a sufficient lag length, as the multivariate Godfrey-Breusch LM test with ten lags produces a \(p\)-value of 9.05%. However, the multivariate Lomnicki-Jarque-Bera test in Lütkepohl (2006) rejects strongly the null hypothesis of multivariate normality, and the vector heteroskedasticity test in both Eklund and Teräsvirta (2007) and Chapter 3 also convincingly rejects the null hypothesis of constant covariance matrix of residuals over time. These rejections may indicate misspecification of the conditional mean, or the presence of outliers in the linear model, or both.

Linearity of the VAR model is tested equation by equation. For both equations, we test using the transition variables \(y_{j,t-d}, j = 1, 2, d = 1, ..., 6\). We normalize the transition variable by dividing it by its standard deviation. This normalization makes it convenient to compare the values of the slope parameter, or the smoothness of transition, \(\gamma_{ij}\), of different transition variables.

Table 1.1 contains results of the linearity tests. The \(p\)-values (multiplied by 100) of Wilks’s \(\Lambda\) test, Rao’s \(F\) test and wild bootstrapped LM test are reported. In this application, Wilks’s and Rao’s tests produce almost identical results, whereas the wild bootstrap tests perform differently comparing to the other two. Linearity is strongly rejected for many transition variables in the first two tests, whereas only several transition variables can be rejected from the wild bootstrap results. The presence of heteroskedasticity explains why the tests perform so differently.

We choose the transition variable according to the \(p\)-values of the third order wild bootstrap tests. Thus, we choose \(y_{1,t-4}\) for the price equation and \(y_{1,t-1}\) for the consumption equation from Table 1.1. The linear vector model is extended to a 2-regime VLSTAR model. After estimating the 2-regime VLSTAR model, we conduct the misspecification test of no additive nonlinearity and find that the price equation still has possible nonlinear additive term which is controlled by the transition variable \(y_{1,t-5}\), whereas there seems no other possible nonlinear term in the consumption equation. Therefore, the price equation is a 3-regime (2-transition) logistic STAR model with transition variable \(y_{1,t-4}\) and \(y_{1,t-5}\), while the consumption equation is a 2-regime (1-transition) logistic STAR model with transition variable \(y_{1,t-1}\).
Chapter 1. Vector smooth transition autoregressive models

Removing the variables corresponding to insignificant parameter estimates and imposing some restrictions based on the discussion in Section 1.2.2, we end up with the following estimated model:

\[
y_{1,t} = 0.290 \frac{y_{1,t-1}}{0.006} - 0.283 \frac{y_{1,t-2}}{0.006} + 0.139 \frac{y_{2,t-2}}{0.005} + 0.202 \frac{y_{2,t-2}}{0.010} \\
+ \left( \frac{0.088}{0.006} y_{1,t-1} + \frac{0.134}{0.003} y_{2,t-1} + \frac{0.283}{0.007} y_{1,t-2} - \frac{0.372}{0.007} y_{2,t-2} \right) \\
\times \left( 1 + \exp \left\{ \frac{-97.24}{13.653} \left( y_{1,t-4} - 0.160 \right) / \sigma_s \right\} \right)^{-1} \\
+ \left( \frac{0.463}{0.006} y_{1,t-1} + \frac{0.134}{0.003} y_{2,t-1} - 0.018 \frac{y_{1,t-2}}{0.007} + 0.105 \frac{y_{2,t-2}}{0.013} - 0.460 \right) \\
\times \left( 1 + \exp \left\{ \frac{-94.491}{65.695} \left( y_{1,t-5} + 0.604 \right) / \sigma_s \right\} \right)^{-1} + \hat{\varepsilon}_{1,t}
\]

\[
y_{2,t} = -0.214 \frac{y_{1,t-1}}{0.007} - 1.098 \frac{y_{2,t-1}}{0.009} + 0.123 \frac{y_{1,t-2}}{0.006} - 0.249 \frac{y_{2,t-2}}{0.010} + 0.496 \\
+ \left( \frac{0.051}{0.008} y_{1,t-1} + \frac{0.726}{0.010} y_{2,t-1} - 0.123 \frac{y_{1,t-2}}{0.006} + 0.071 \frac{y_{2,t-2}}{0.011} - 0.216 \right) \\
\times \left( 1 + \exp \left\{ \frac{-21.677}{4.086} \left( y_{1,t-1} + 1.582 \right) / \sigma_s \right\} \right)^{-1} + \hat{\varepsilon}_{2,t}
\]

\[
T = 311, \text{tr}\{\hat{\Theta}_{ML}\} = 6.77, \sigma_s = 2.43
\]

The reduced multiple-transition VLSTAR model (1.8.2) is evaluated by means of a series of misspecification tests which are developed in Chapter 2, and the results can be seen in Table 1.2. The model passes almost all the wild bootstrap tests at significance level 5%. The p-value 4% for \(y_{1,t-4}\) in the consumption is not so small enough reject the null hypothesis of no additive nonlinearity, as it may be due to the random variation caused by bootstrapping. The joint wild bootstrap error serial correlation tests and the joint wild bootstrap parameter constancy tests look fine enough. In addition, we also plot the ACF and PACF functions in Figure 1.3 as another evidence to show that there is no serial correlation. Thus, we believe that the model is well described by (1.8.2).

Table 1.3 to 1.5 give p-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2). Table 1.4 reports the tests using the smooth transition specification and suggests that the conditional error covariance matrix is time-varying and controlled by the price \(y_{1,t-4}\). The strong rejection using White’s specification in Table 1.5 strengthens the findings in the tests using the smooth transition specification. Due to the presence of heteroskedasticity, one can understand why the wild bootstrap tests perform so differently compared to the others in Table 1.2.

Figure 1.4 depicts the estimated transition functions for the gasoline price change equation and the consumption change equation. The transition in the consumption equation is smoother than the two transitions in the price equation, since it has more observations in between the smooth transition region. As a whole, the observations cover the whole range of values of the transition functions from zero to one. Figure 1.5 depicts the values of the transition functions, or the regime-switching, through time for both equations.
Figure 1.6 contains the results of checking the stability of our estimated model as discussed in Section 1.6.2. We use all histories in the data set as the initial values, and the paths are depicted in Figure 1.6. The price and consumption observation sequences converge to -0.302 and 0.218, respectively. As the nonlinear components in the two equations are controlled by the price change, the upper second extreme regime in the price equation is the stable state of the price dynamics and the same regime in the consumption equation is the stable state of the consumption dynamics. No matter where the initial values are located, the trajectories end up at the same stationary point.

As is the case in autoregressive models in general, it is not possible to interpret single coefficients of the model. In order to interpret the results and shed light on the question of possible asymmetry of the gasoline prices and consumption to shocks, we compute generalized multivariate impulse functions as suggested in Koop et al. (1996) and represent them using highest density regions (HDR) of Hyndman (1996); see also Teräsvirta et al. (2010, Section 15.3). The HDRs, which in this example are unimodal, are illustrated using boxplots. Responses to positive shocks are shown separately from responses to negative ones.

The lower panel of Figure 1.7 shows that the response of consumption change to price change shocks is asymmetric. A negative price shock causes a stronger response than a positive one in the sense that the density has greater dispersion in the former case than in the latter. This difference lasts two months before disappearing. Negative price shocks thus cause greater uncertainty in the consumer behaviour than positive ones. The directions are as expected: a negative price shock on the average increases consumption growth, whereas a positive shock decreases it. The other three responses, the one of price shocks to the price in the upper panel of Figure 1.7 and the two consumption shocks in Figure 1.8, are symmetric around zero, as the responses to positive and negative shocks are mirror images of each other.

We would like to stress again here that the asymmetry is found by applying the series of nonlinear modelling introduced in this chapter but not any economic assumptions a priori. To understand this kind of asymmetry, consider that gasoline is a special commodity in the sense that people will always consume it no matter how high the price will be. We see from the estimated model (1.8.2), the gasoline price is the transition variable for both equations, which governs the transition across different regimes. One can imagine that a good strategy for an individual to save the gasoline cost every month is to tank less when the price is high, and more when low. However, daily commuting, which occupies most of the aggregate gasoline cost, will not allow people to tank too little, and will normally not be affected greatly even when people tank a lot. Thus, it is easy to understand that a negative price shock causes a stronger positive response and is followed by a negative response in consumption.

### 1.8.2 Icelandic river flow

In our second application we model daily Icelandic river flow series from the years 1972–1974, measured in cubic metres per second. The series are from the Hydrological Survey of the National
Energy Authority of Iceland and were first analyzed using nonlinear models by Tong, Thanoon and Gudmundsson (1985). Tsay (1998) considered the two rivers, Jökulsá eystra and Vatndalsá jointly, and fitted a bivariate threshold autoregressive model to the flow series.

Tong et al. (1985) describe the rivers and the observation station. Jökulsá is the bigger river of the two, with a large drainage basin that includes a glacier. Vatndalsá has a much smaller drainage area, and some of the flow is due to groundwater. The weather station lies between the two drainage basins at about 650 meters. The temperatures measured there are higher than the ones on the glacier of Jökulsá, which affects the results of modelling. Tong et al. (1985) point out that measuring the rainfall accurately is difficult because of high winds in the area. This may also explain some of the empirical results.

Before modelling, the precipitation series is shifted forward by one day due to the way of recording the rainfall, see Tong et al. (1985). The precipitation and the temperature are both graphed in Figure 1.9. The river flow series can also be found in Figure 1.9. It is seen that the flow is strongest in the spring when the snow is melting and slows down in the summer. The spring peak is more pronounced in the Vatndalsá flow than the Jökulsá one, because the drainage area of the latter contains the glacier, which smoothes the flow.

We denote Jökulsá by $y_1t$, Vatndalsá by $y_2t$, precipitation by $x_t$ and temperature by $z_t$. And we complement our model by a seasonal component that contains an annual half-cycle (a sine function) and a seasonal frequency component. It has the form

$$\delta_1 \sin(n\pi/365) + \delta_2 \sin(2n\pi/365) + \delta_3 \cos(2n\pi/365) \quad (1.8.3)$$

where $n \in \mathbb{N}$ represents the $n^{th}$ day of the year.

We begin modelling by testing linearity. The column "Lin to R2" in Table 1.6 reports the results from the wild bootstrap linearity tests of order 3. The null hypothesis of linearity is rejected very strongly for both flows and all transition variables; some of the $p$-values lie below $10^{-16}$.

The test results are thus inconclusive, so follow Tsay (1998) and select the temperature as the transition variable for both flows. Since the flows are heavily autocorrelated, choosing a lag of the flow would have been another possibility. After estimating the 2-regime (1-transition) VLSTAR model, we report the results from the wild bootstrap misspecification tests of no additive nonlinearity in the column "R2 to R3" in Table 1.6. We find that the nonlinearity caused by the temperature has been well described, but still there are many strong rejections there. The model can be extended to a 3-regime (2-transition) VLSTAR model. We select the precipitation of lag one $x_{t-1}$ as the new transition variable for both equations. The column "R3 to R4" in Table 1.6 reports the results from the wild bootstrap misspecification tests of no additive nonlinearity. There are still many rejections, especially in Vatndalsá flow. This suggests that there is remaining nonlinearity controlled by the flows to be modelled. Since we

\[1\] The numbers stored in a computer system are not continuous. And there exits a positive number $\epsilon$ such that $1 + \epsilon \neq 1$, and for any $x < \epsilon$, $1 + x = 1$, and hence this number $\epsilon$ is the smallest positive floating-point number. In R, $\epsilon = 2.220446e^{-16}$. 

23
Chapter 1. Vector smooth transition autoregressive models

are only interested in the nonlinear behaviour caused by the temperature and the precipitation, we stop the modelling procedure here.

After removing redundant variables and imposing some parameter restrictions, the estimated model has the following form:

\[
y_{1t} = 0.93y_{1t-1} + 0.34y_{1t-2} - 0.21y_{1t-3} - 0.7y_{1t-4} + 0.21y_{1t-5} \\
+ 0.43y_{1t-6} - 0.14y_{1t-7} - 0.10y_{1t-8} - 0.08y_{1t-10} \\
+ 0.08y_{1t-12} + 0.24y_{1t-14} + 0.09y_{1t-15} \\
- 0.18y_{2t-1} - 0.05y_{2t-5} \\
+ 0.26y_{2t-7} + 0.12y_{2t-8} - 0.34y_{2t-9} - 0.07y_{2t-10} \\
+ 0.07y_{2t-13} + 0.06y_{2t-15} \\
- 0.06z_{t-1} + 4.94 \\
+ (0.06y_{1t-1} + 3.01y_{1t-2} + 0.09y_{1t-3} - 0.7y_{1t-4} - 0.17y_{1t-5} \\
- 0.86y_{1t-6} - 0.07y_{1t-7} + 1.43y_{1t-8} + 0.07y_{1t-9} - 1.1y_{1t-10} \\
+ 0.25y_{1t-12} - 0.14y_{1t-13} - 0.05y_{1t-14} \\
- 0.96y_{2t-1} + 0.77y_{2t-3} + 0.17y_{2t-4} + 0.83y_{2t-5} \\
+ 0.07y_{2t-6} + 0.5y_{2t-7} - 0.39y_{2t-8} - 0.18y_{2t-10} \\
- 0.17y_{2t-11} + 0.16y_{2t-13} + 0.09y_{2t-14} - 1.21y_{2t-15} \\
+ 0.88x_{t-1} + 0.06x_{t-2} - 0.41x_{t-3} - 0.54z_{t} + 1.67z_{t-1} + 2.91) \\
\times \left(1 + \exp\left\{-0.59(z_t - 4.52)\right\}\right)^{-1} \\
+ \left(-0.14y_{1t-1} - 0.46y_{1t-2} - 0.08y_{1t-3} + 1.08y_{1t-4} + 0.07y_{1t-5} \\
- 1.1y_{1t-6} + 0.09y_{1t-7} + 0.72y_{1t-8} - 0.25y_{1t-9} + 0.78y_{1t-10} \\
+ 0.22y_{1t-11} + 0.19y_{1t-12} - 0.06y_{1t-14} + 0.11y_{1t-15} \\
- 0.72y_{2t-1} - 0.26y_{2t-2} + 0.5y_{2t-3} + 0.14y_{2t-4} \\
- 0.07y_{2t-6} - 0.7y_{2t-7} + 0.07y_{2t-8} + 0.95y_{2t-9} + 0.14y_{2t-10} \\
- 0.98y_{2t-11} - 0.24y_{2t-12} + 0.24y_{2t-13} + 0.52y_{2t-15} \\
+ 0.35x_{t-1} - 0.17z + 10.45) \\
\times \left(1 + \exp\left\{-12.77(x_{t-1} - 5.17)\right\}\right)^{-1} \\
- 4.01 \sin\left(\frac{n\pi}{365}\right) + 0.53 \sin\left(\frac{2n\pi}{365}\right) - 2.16 \cos\left(\frac{2n\pi}{365}\right) + \hat{e}_{1t} \tag{1.8.4}
\begin{align*}
y_{2t} &= -0.12y_{1,t-1} + 0.71y_{1,t-2} \\
&
+ 0.23y_{1,t-6} - 0.06y_{1,t-8} - 0.14y_{1,t-10} \\
&
+ 0.1y_{1,t-12} - 0.15y_{1,t-14} \\
&
+ 0.05y_{2,t-1} \\
&
- 0.02y_{2,t-6} + 0.08y_{2,t-9} \\
&
- 0.09y_{2,t-11} + 0.06y_{2,t-15} \\
&
+ 0.31x_{t-2} + 3.09 \\
&
+ \left( 0.1y_{1,t-1} + 0.62y_{1,t-2} - 0.64y_{1,t-4} \\
&
+ 0.12y_{1,t-6} + 0.12y_{1,t-8} + 0.04y_{1,t-10} \\
&
- 0.05y_{1,t-11} - 0.12y_{1,t-12} + 0.19y_{1,t-14} \\
&
- 0.19y_{2,t-1} + 0.2y_{2,t-5} \\
&
- 0.27y_{2,t-7} - 0.09y_{2,t-9} \\
&
+ 0.35y_{2,t-11} - 0.23y_{2,t-13} \\
&
+ 0.05x_{t-1} - 0.04x_{t-2} + 0.06x_{t-3} - 2.49 \right) \\
&\times \left( 1 + \exp \left\{ -2.63 \left( z_t - 0.31 \right) \frac{0.411}{0.066} \right\} \right)^{-1} \\
&+ \left( 0.05y_{1,t-1} + 0.09y_{1,t-2} - 0.07y_{1,t-3} - 0.1y_{1,t-4} + 0.02y_{1,t-5} \\
&
- 0.1y_{1,t-6} + 0.36y_{1,t-10} \\
&
- 0.34y_{1,t-12} + 0.67y_{1,t-14} \\
&
- 0.74y_{2,t-1} + 0.63y_{2,t-3} - 0.83y_{2,t-5} \\
&
+ 0.68y_{2,t-7} - 0.28y_{2,t-9} \\
&
- 0.14y_{2,t-11} + 0.38y_{2,t-13} - 0.18y_{2,t-15} \\
&
+ 0.05x_{t-1} - 0.33x_{t-2} - 1.82 \right) \\
&\times \left( 1 + \exp \left\{ -9.37 \frac{4.122}{0.081} \left( x_{t-1} - 1.41 \right) \right\} \right)^{-1} \\
&- \frac{0.72}{0.110} \sin \left( \frac{n\pi}{365} \right) + 0.45 \sin \left( \frac{2n\pi}{365} \right) - \frac{0.28}{0.067} \cos \left( \frac{2n\pi}{365} \right) + 3.18 + \varepsilon_{2,t}
\end{align*}

(1.8.5)

and \( \text{tr}\{\Omega\} = 17.13. \)

Before interpreting the results, we evaluate the model. The wild bootstrap tests of no error autocorrelation up to lag 10 have been reported in Table 1.7, and they look fine. The wild bootstrap tests of parameter constancy are all accepted. These indicate that the model specification is adequate. Table 1.9 to 1.11 report the results from the error covariance matrix
Chapter 1. Vector smooth transition autoregressive models

(ECM) constancy tests developed in Chapter 3. We see that all of them are very small, which indicates the presence of heteroskedasticity.

Nevertheless, we focus on the effect of the temperature and the precipitation on the flow and take a look at the estimated model and the transitions that were estimated separately for the two rivers.

The transition in the Jökulsá equation driven by the temperature, shown in Figure 1.10, is very smooth. The regime change begins at the temperature around 0°C and is completed when the temperature approaches 10°C. This is due to the glacier. When the temperature increases, so does the flow from the glacier. However, since the glacier is located at 1000 – 1800 metres, higher than the weather station, the flow increases gradually as a function of the temperature. Vatndalsá, without a glacier, has a much more rapid transition driven by the temperature, and the estimate of the location parameter equals 0.4°C, see Figure 1.11.

Figure 1.12 and 1.13 show the transitions driven by the precipitation in the two rivers. These nonlinearities are not yet found in the literature. It is interesting to see that besides the temperature, the precipitation will also produce nonlinearity. Since we are using the same precipitation for both rivers and the turning point of transition in Jökulsá is much greater than the one in Vatndalsá, it seems that the small river is affected more greatly by the rain than the big one, which can be easily understood. Notice that though the transition of the small river driven by the precipitation has a very small turning point 1.41, still there are more than half of the observations are located below the turning point.

It is seen from the estimates of the coefficients of the lagged flow that the flow does increase with the temperature as is expected. A somewhat mysterious thing is that there are cross-effects even if the rivers are separate. They may be attributed to common omitted variables or the situation in the area where the drainage basins are nearest each other. Tsay (1998) also found them in his multiple TAR model.

The temperature itself does enter the Jökulsá flow equation when it is above zero, but the coefficients are hard to interpret. This may be due to the fact that there is only one weather station, and its readings may therefore not be useful other than in the transition function. When the temperature is above freezing, precipitation does impact the flow of Jökulsá with a large drainage area, whereas it only has a minor effect on the flow of Vatndalsá. In that river, given the estimate of the location parameter, the most significant contribution to changes in the flow is the melting snow. Difficulties in measuring the precipitation may also have contributed to this outcome.

In order to illustrate the dynamic behaviour of the estimated model, we estimate generalised impulse response functions. For the transition variable of the temperature, we define two separate sets of histories: one when the temperature \( z_t \geq 0.4 \) at the moment of shock, and the other when \( z_t < 0.4 \). This value equals the estimate of the location parameter in the transition function of the Vatndalsá equation. It also represents the point in which melting of snow begins (the flow begins to increase) in the Jökulsá drainage basin. The shocks are divided to positive and negative ones. This gives eight different types of shocks according to the flow shocked, the
temperature, and the sign of the shock. As in the previous application, the impulse response functions are described using HDRs and boxplots. The results are in Figures 1.14–1.17.

The effects of shocks generally last longer in the 'summer', $z_t \geq 0.4$, than in the 'winter', $z_t < 0.4$. Although the mode of the HDR converges to zero quite quickly in the summer, the densities shrink towards a point much more slowly than in the winter. Most of the responses are close to being symmetric: the most pronounced asymmetry can be found in the flow of Vatndalsá in the summer, see Figure 1.17. A negative shock to that flow causes a much stronger response than a positive one. The most puzzling finding is the strong response of the Jökulsá flow to shocks to the Vatndalsá equation in the summer in Figure 1.16. It is also asymmetric and even stronger than the response of the flow of this river to own shocks. The HDRs of the latter can be found in Figure 1.14. But then, as Figure 1.15 shows, the flow of Vatndalsá is not much affected by shocks to the flow of Jökulsá.

The reasons for effects of shocks to the Vatndalsá flow on the flow of Jökulsá can only be guessed. It may be that the dynamic effects partly reflect those of precipitation which itself is not well measured. Thus dry summer weather (and therefore decreased flow in Vatndalsá) would also indicate dry weather in the drainage basin of Jökulsá and a reduced flow in this river. Interestingly, the response begins with a one-day lag, which may be due to the inertia caused by the glacier. That the opposite cross-effect is very small may be explained by the size difference of these rivers. Shocking a big flow (Jökulsá when the snow is melting, say) does not have a similar dynamic indicator effect on the flow of the smaller river.

For the transition variable of the precipitation, we also define two separate sets of histories: one when the precipitation $x_{t-1} > 5$ at the moment of shock, and the other when $x_{t-1} < 1.4$. Note that the observations in between $x_{t-1} \in [1.4, 5]$ are discarded, which ensures that both the rivers are located approximately in the same extreme regime. The impulse response functions are also described using HDRs and boxplots and reported in Figure 1.18–1.21.

Different from the findings in the transitions driven by the temperature, there is no clear sign that the effects of shocks last longer during the rain season.

The asymmetry before and after the rain season can be easily found. The responses to the Jökulsá shocks before the rain season are stronger, which shows that the big river is affected more greatly by the temperature than by the precipitation, while the Vatndalsá flow looks symmetric except the responses to negative shocks. Asymmetry arises in both of the two rivers when there are negative shocks in the small river. Also, most of the responses look symmetric between positive and negative shocks, except when Vatndalsá shocks, especially during the rain season. It seems that a negative shock in Vatndalsá will have stronger responses to both flows. It is also not easy to explain the reason, since the two rivers are separate to each other. One possible explanation is that Vatndalsá river is so small that its water body cannot be reduced too much.

It should be kept in mind, however, that the dynamic behaviour of this river system is not adequately captured by the estimated VLSTAR model with two single transitions. Tests of no additive nonlinearity strongly reject the specification, unless the transition variable is the temperature (see column "R3 to R4" in Table 1.6). Extending the model by, say, transitions
Chapter 1. Vector smooth transition autoregressive models

controlled by lagged flow variables would be a possibility.

1.9 Concluding remarks

We generalize previous versions of vector LSTAR or LSTR models to the case in which the model can have a different transition variable or variables for each equation, and the model can contain more than one transition. This is the case in our first application, and the second one could be augmented in the same way. We devise a modelling strategy for this class of nonlinear models, consisting of specification, including testing linearity, estimation and evaluation stages. Parameter estimation is carried out by maximum likelihood, and the main tool at the evaluation stage is a set of misspecification tests as in Teräsvirta (1996), Teräsvirta (1998) and Camacho (2004). The dynamic behaviour of the model is characterized by generalized impulse response functions. The two applications show how the strategy works in practice.

In this work the VLSTAR model is assumed stationary, unless the transition variable is time, but generalizing the approach to nonstationary linearly cointegrated series, as in Rothman et al. (2001), appears straightforward. This is true as long as the short-run dynamic behaviour of the model, including the drift towards the equilibrium, is characterized using nonlinearity of STAR type. Accommodating a nonlinear equilibrium correction mechanism is from a statistical point of view a much more complicated problem. Some discussion can be found in Ripatti and Saikkonen (2001). We leave these extensions to further work.
Appendix A: The two equivalent forms of the maximum likelihood estimator

The problem is to find $\hat{B}$ satisfying the first-order condition

$$
\sum_{t=1}^{T} x_t y_t' \Omega^{-1} \Psi_t' = \sum_{t=1}^{T} x_t x_t' B \Psi_t \Omega^{-1} \Psi_t'.
$$

(1.9.1)

We need the following identities:

$$
\text{vec}(\sum_{t=1}^{T} A_t) = \sum_{t=1}^{T} \text{vec}(A_t) \quad (1.9.2)
$$

$$
\text{vec}(AXB) = (B' \otimes A)\text{vec}(X) \quad (1.9.3)
$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \quad (1.9.4)
$$

From (1.9.1), (1.9.2), (1.9.3) and (1.9.4), it follows that

$$
\text{vec} \left( \sum_{t=1}^{T} x_t y_t' \Omega^{-1} \Psi_t' \right) = \sum_{t=1}^{T} \text{vec}(x_t y_t' \Omega^{-1} \Psi_t')
$$

$$
= \sum_{t=1}^{T} \text{vec}(x_t x_t' B \Psi_t \Omega^{-1} \Psi_t')
$$

$$
= \sum_{t=1}^{T} \left[ (\Psi_t \Omega^{-1} \Psi_t') \otimes (x_t x_t') \text{vec}(B) \right].
$$

This yields the solution

$$
\text{vec}(\hat{B}) = \left[ T^{-1} \sum_{t=1}^{T} (\Psi_t \Omega^{-1} \Psi_t') \otimes (x_t x_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} \text{vec}(x_t y_t' \Omega^{-1} \Psi_t') \right]. \quad (1.9.5)
$$

Using (1.9.4) one obtains

$$(\Psi_t \Omega^{-1} \Psi_t') \otimes (x_t x_t') = (\Psi_t \Omega^{-1} \otimes x_t) (\Psi_t \Omega^{-1} \otimes x_t)' = (\Psi_t \otimes x_t) \Omega^{-1} (\Psi_t \otimes x_t)'.
$$

Furthermore, applying (1.9.3) and (1.9.4) yields

$$
\text{vec} (x_t y_t' \Omega^{-1} \Psi_t') = (\Psi_t \Omega^{-1} \otimes x_t) \text{vec}(y_t') = (\Psi_t \otimes x_t) \Omega^{-1} y_t.
$$

Set $Y_t = \Psi_t \otimes x_t$, which is a $mp(kp + q) \times p$ matrix, and $M = (Y_1, Y_2, ..., Y_T)'$ which is a $Tp \times mp(kp + q)$ matrix. The vector (1.9.5) has the following compact form

$$
\text{vec}(\hat{B}) = \left[ T^{-1} \sum_{t=1}^{T} (Y_t \Omega^{-1} Y_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} (Y_t \Omega^{-1} y_t) \right]
$$

$$
= \left( M' (I_T \otimes \Omega^{-1}) M \right)^{-1} \left( M' (I_T \otimes \Omega^{-1}) \text{vec}(Y) \right). \quad (1.9.6)
$$

Moreover, when $\Omega = I$, (1.9.5) becomes

$$
\text{vec}(\hat{B}) = \left[ T^{-1} \sum_{t=1}^{T} (\Psi_t \Psi_t') \otimes (x_t x_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} \text{vec}(x_t y_t' \Psi_t') \right] \quad (1.9.7)
$$

and (1.9.6) can be written as

$$
\text{vec}(\hat{B}) = \left[ T^{-1} \sum_{t=1}^{T} (Y_t Y_t') \right]^{-1} \left[ T^{-1} \sum_{t=1}^{T} (Y_t Y_t) \right] = (M' M)^{-1} M' \text{vec}(Y'). \quad (1.9.8)
$$

(29)
Appendix B: Existence and consistency of the NLS and ML estimators

Proof. Suppose that (1.2.6) is the data generating process, and the true parameters are $\hat{\theta} = \{\hat{B}, \hat{\Omega}, \hat{\Gamma}, \hat{C}\}$.

The limits of the normalized log-likelihood function and the average NLS function

We define the normalized log-likelihood function

$$\ell_T(\theta) = -2T^{-1} \log L(\theta) - p \log 2\pi$$

$$= \log |\Omega| + T^{-1} \sum_{t=1}^{T} (y_t - \Psi_t' \hat{B} \hat{x}_t)' \Omega^{-1} (y_t - \Psi_t' \hat{B} \hat{x}_t)$$

(1.9.9)

which converts the maximization to a minimization.

It is easy to find the expression

$$\ell_T(\theta) = \log |\Omega| + T^{-1} \sum_{t=1}^{T} (x_t'(\hat{B} \hat{\Psi}_t - B \Psi_t) \Omega^{-1} (\hat{B} \hat{\Psi}_t - B \Psi_t)' x_t +$$

$$2x_t'(\hat{B} \hat{\Psi}_t - B \Psi_t) \Omega^{-1} \epsilon_t + \epsilon_t' \Omega^{-1} \epsilon_t)$$

(1.9.10)

Applying the law of large numbers, (1.9.9) yields the following probability limit

$$\ell_T(\theta) \xrightarrow{p} \ell(\theta) = \log |\Omega| + E \left\{ x_t'(\hat{B} \hat{\Psi}_t - B \Psi_t) \Omega^{-1} (\hat{B} \hat{\Psi}_t - B \Psi_t)' x_t + \epsilon_t' \Omega^{-1} \epsilon_t \right\}$$

(1.9.11)

because $E \left\{ 2x_t'(\hat{B} \hat{\Psi}_t - B \Psi_t) \Omega^{-1} \epsilon_t \right\} = 0$, due to $E(x_t' \epsilon_t) = 0$. Setting $\mu_t = (\hat{B} \hat{\Psi}_t - B \Psi_t)' x_t$, we have

$$\ell(\theta) = \log |\Omega| + E \left( \mu_t' \Omega^{-1} \mu_t \right) + E (\epsilon_t' \Omega^{-1} \epsilon_t)$$

$$= \log |\Omega| + E \left( \mu_t' \Omega^{-1} \mu_t \right) + \text{tr} \left( \hat{\Omega} \Omega^{-1} \right)$$

(1.9.12)

Note that $\Omega^{-1}$ is a symmetric positive definite matrix. We have $\mu_t' \Omega^{-1} \mu_t \geq 0$ and equality holds only when $\mu_t = 0$ for all $t = 1, \ldots$. Thus, we have $E (\mu_t' \Omega^{-1} \mu_t) \geq 0$ and equality holds only when $\mu_t = 0$ for all $t = 1, \ldots$. Obviously, all $\mu_t = 0$ if and only if all $B \hat{\Psi}_t \equiv B \Psi_t$. We assume that each element in $\hat{\Gamma}$ is nonzero. Therefore, $B$ and $\hat{\Psi}_t$ are both nonzero for all $t = 1, \ldots$, $B \hat{\Psi}_t = B \Psi_t$ if and only if $B = \bar{B}$, $\Gamma = \bar{\Gamma}$ and $C = \bar{C}$. That is

$$\ell(\theta) \geq \log |\Omega| + \text{tr} (\hat{\Omega} \Omega^{-1})$$

(1.9.13)

in which equality holds when $B = \bar{B}$, $\Gamma = \bar{\Gamma}$ and $C = \bar{C}$. And we have

$$\ell(\theta) \geq \log |\Omega| + \text{tr} (\hat{\Omega} \Omega^{-1}) \geq \log |\hat{\Omega}| + p$$

(1.9.14)

in which the equality holds when $\Omega = \hat{\Omega}$.
Hence, we conclude that the probability limit of the normalized log-likelihood function $\ell_T(\theta)$ has a unique global minimum $\ell(\hat{\Omega})$ at the true parameters $\hat{\theta}$.

Furthermore, in the special case that $\Omega = I$, (1.9.9) becomes the average nonlinear least square (NLS) function to be minimized with respect to $\theta$. Analogously, we conclude that the probability limit of the average NLS function has a unique global minimum $\ell(\hat{\Omega})$ at the true parameters $\hat{\theta}$.

**Consistency of the estimator**

In the following, $\ell_T(\theta)$ represents both the average NLS function and the normalized log-likelihood function. The convergence of the sequence of the function $\ell_T(\theta)$ is uniform in the parameters if the parameters are bounded in a compact set $\mathcal{K}$ of which the true parameters are interior point. That is,

$$\Pr \left\{ \max_{\theta \in \mathcal{K}} |\ell_T(\theta) - \ell(\theta)| \geq \eta \right\} \to 0$$ (1.9.15)

as $T \to \infty$, for any small $\eta > 0$.

Define $N(\hat{\theta}, \delta)$ to be an open neighborhood in $\mathcal{K}$ for a small $\delta > 0$

$$N(\hat{\theta}, \delta) = \{ \theta \in \mathcal{K} : |\pi - \hat{\pi}| < \delta \},$$ (1.9.16)

where $\pi$ is an element in any one of the matrices: $B, \Gamma, C$ and $\Omega$.

For $\theta \notin N(\hat{\theta}, \delta)$, there exists $\delta_0 > 0$ which satisfies

$$\ell(\theta) \geq \ell(\hat{\theta}) + \delta_0$$ (1.9.17)

From (1.9.15) it follows that for any small $\zeta > 0$ and $\delta_0/3 > 0$, we can find $T_0$ such that for any $T > T_0$ we have

$$\Pr \left\{ \max_{\theta \in \mathcal{K}} |\ell_T(\theta) - \ell(\theta)| \leq \delta_0/3 \right\} \geq 1 - \zeta$$ (1.9.18)

This implies that on the set $\mathcal{Y} = \{ y_1, y_2, ..., y_T : |\ell_T(\theta) - \ell(\theta)| \leq \delta_0/3 \}$, we have

$$\ell_T(\theta) \leq \ell(\hat{\theta}) + \delta_0/3$$ (1.9.19)

On the other hand, on the set $\mathcal{Y}$, for $\theta \notin N(\hat{\theta}, \delta)$, we find that

$$\ell_T(\theta) = \ell(\theta) + \ell_T(\theta) - \ell(\theta) \geq \ell(\theta) - \delta_0/3,$$ (1.9.20)

and from (1.9.17), we have

$$\ell_T(\theta) \geq \ell(\hat{\theta}) + \delta_0 - \delta_0/3 = \ell(\hat{\theta}) + 2\delta_0/3.$$ (1.9.21)

Thus, for $\theta \in N(\hat{\theta}, \delta)$, $\ell_T(\theta)$ takes a value $\leq \ell(\hat{\theta}) + \delta_0/3$, whereas for $\theta \notin N(\hat{\theta}, \delta)$, $\ell_T(\theta)$ is $\geq \ell(\hat{\theta}) + 2\delta_0/3$, on the set $\mathcal{Y}$ for $T > T_0$. Hence the minimum has to be attained inside the neighborhood $N(\hat{\theta}, \delta)$. From (1.9.18), we have on $\mathcal{Y}$, $\hat{\theta} \in N(\hat{\theta}, \delta)$, so that

$$\Pr \left\{ \hat{\theta} \in N(\hat{\theta}, \delta) \right\} \geq \Pr \left\{ \max_{\theta \in \mathcal{K}} |\ell_T(\theta) - \ell(\theta)| \leq \delta_0/3 \right\} \geq 1 - \zeta$$ (1.9.22)

for $T > T_0$ which shows that $\hat{\theta}$ is consistent.
Table 1.1: P-values of linearity tests for the VAR(2) model against the VLSTAR model in Section 1.8.1. The upper part is the linearity tests in the gasoline price change equation, and the lower part is the linearity tests in the gasoline consumption change equation. 3e represents the parsimonious test of order three. tran. var. stands for transition variables, WLK for Wilks’s statistics, RAO for Rao’s statistics and WB for wild bootstrapping. The sample size of the wild bootstrapping is 5000. Note that all of the p-values are multiplied by 100 and hence read percent. We denote $\tau = t/T$, where $t = 1, \ldots, T$.

<table>
<thead>
<tr>
<th>tran. var.</th>
<th>$y_{1,t-1}$</th>
<th>$y_{1,t-2}$</th>
<th>$y_{1,t-3}$</th>
<th>$y_{1,t-4}$</th>
<th>$y_{1,t-5}$</th>
<th>$y_{2,t-1}$</th>
<th>$y_{2,t-2}$</th>
<th>$y_{2,t-3}$</th>
<th>$y_{2,t-4}$</th>
<th>$y_{2,t-5}$</th>
<th>$y_{2,t-6}$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order</strong></td>
<td><strong>Test</strong></td>
<td><strong>Gasoline price change equation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>WLK</td>
<td>0.10</td>
<td>0.14</td>
<td>0.25</td>
<td>0.00</td>
<td>0.01</td>
<td>31.82</td>
<td>0.86</td>
<td>9.41</td>
<td>6.65</td>
<td>0.00</td>
<td>54.91</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>0.10</td>
<td>0.14</td>
<td>0.25</td>
<td>0.00</td>
<td>0.01</td>
<td>31.83</td>
<td>0.86</td>
<td>9.42</td>
<td>6.66</td>
<td>0.00</td>
<td>54.92</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>2.68</td>
<td>2.52</td>
<td>18.12</td>
<td>4.66</td>
<td>6.42</td>
<td>11.26</td>
<td>23.92</td>
<td>24.40</td>
<td>69.00</td>
<td>19.24</td>
<td>50.22</td>
</tr>
<tr>
<td>3</td>
<td>WLK</td>
<td>0.00</td>
<td>0.69</td>
<td>0.06</td>
<td>0.05</td>
<td>0.02</td>
<td>27.26</td>
<td>22.50</td>
<td>7.30</td>
<td>30.66</td>
<td>0.17</td>
<td>39.20</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>0.00</td>
<td>0.71</td>
<td>0.06</td>
<td>0.05</td>
<td>0.02</td>
<td>27.53</td>
<td>22.75</td>
<td>7.43</td>
<td>30.95</td>
<td>0.17</td>
<td>39.50</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>5.38</td>
<td>26.44</td>
<td>2.28</td>
<td>0.62</td>
<td>2.30</td>
<td>56.80</td>
<td>54.64</td>
<td>11.34</td>
<td>88.98</td>
<td>3.46</td>
<td>88.08</td>
</tr>
<tr>
<td>3e</td>
<td>WLK</td>
<td>0.24</td>
<td>0.02</td>
<td>0.22</td>
<td>0.00</td>
<td>0.01</td>
<td>42.27</td>
<td>1.64</td>
<td>16.67</td>
<td>9.05</td>
<td>0.01</td>
<td>58.84</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>0.25</td>
<td>0.02</td>
<td>0.22</td>
<td>0.00</td>
<td>0.01</td>
<td>42.29</td>
<td>1.64</td>
<td>16.69</td>
<td>9.06</td>
<td>0.01</td>
<td>58.86</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>5.38</td>
<td>2.24</td>
<td>7.52</td>
<td>1.72</td>
<td>5.92</td>
<td>34.50</td>
<td>36.06</td>
<td>12.62</td>
<td>78.42</td>
<td>12.62</td>
<td>55.68</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td><strong>Test</strong></td>
<td><strong>Gasoline consumption change equation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>WLK</td>
<td>3.24</td>
<td>0.22</td>
<td>0.09</td>
<td>21.41</td>
<td>21.42</td>
<td>23.23</td>
<td>0.00</td>
<td>19.77</td>
<td>20.23</td>
<td>44.91</td>
<td>44.71</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>3.25</td>
<td>0.22</td>
<td>0.09</td>
<td>21.42</td>
<td>21.43</td>
<td>23.24</td>
<td>0.00</td>
<td>19.79</td>
<td>20.24</td>
<td>44.93</td>
<td>44.73</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>1.06</td>
<td>0.24</td>
<td>0.10</td>
<td>32.74</td>
<td>82.66</td>
<td>15.64</td>
<td>4.84</td>
<td>36.34</td>
<td>50.28</td>
<td>70.72</td>
<td>28.70</td>
</tr>
<tr>
<td>3</td>
<td>WLK</td>
<td>0.00</td>
<td>0.43</td>
<td>0.28</td>
<td>19.86</td>
<td>6.46</td>
<td>56.26</td>
<td>0.00</td>
<td>17.24</td>
<td>10.58</td>
<td>73.68</td>
<td>81.04</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>0.00</td>
<td>0.44</td>
<td>0.29</td>
<td>20.10</td>
<td>6.58</td>
<td>56.54</td>
<td>0.01</td>
<td>17.47</td>
<td>10.75</td>
<td>73.89</td>
<td>81.21</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>0.04</td>
<td>0.56</td>
<td>0.12</td>
<td>3.36</td>
<td>14.34</td>
<td>28.22</td>
<td>4.30</td>
<td>47.88</td>
<td>27.14</td>
<td>78.89</td>
<td>44.62</td>
</tr>
<tr>
<td>3e</td>
<td>WLK</td>
<td>4.30</td>
<td>0.27</td>
<td>0.11</td>
<td>32.94</td>
<td>19.12</td>
<td>16.65</td>
<td>0.01</td>
<td>27.48</td>
<td>31.90</td>
<td>60.63</td>
<td>55.47</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>4.31</td>
<td>0.27</td>
<td>0.11</td>
<td>32.96</td>
<td>19.14</td>
<td>16.67</td>
<td>0.01</td>
<td>27.50</td>
<td>31.92</td>
<td>60.65</td>
<td>55.49</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>2.52</td>
<td>0.26</td>
<td>0.04</td>
<td>42.00</td>
<td>49.48</td>
<td>14.28</td>
<td>13.78</td>
<td>43.94</td>
<td>70.38</td>
<td>39.30</td>
<td>34.86</td>
</tr>
</tbody>
</table>
Chapter 1. Vector smooth transition autoregressive models

Table 1.2: P-values of the evaluation tests of the estimated model (1.8.2). GP stands for gasoline price change equation, and GC stands for gasoline consumption change equation. The tests of no additive nonlinearity are based on the third-order Taylor expansion. WLK stands for Wilks’s statistics, RAO for Rao’s statistics and WB for wild bootstrapping. The sample size of the wild bootstrapping is 100. Notice that p-values are all multiplied by 100. We denote \( \tau = t/T \), where \( t = 1, \ldots, T \).

<table>
<thead>
<tr>
<th>lags(J)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>eq. GP</td>
<td>WLK</td>
<td>70.93</td>
<td>19.99</td>
<td>35.31</td>
<td>25.43</td>
<td>36.51</td>
<td>18.43</td>
<td>25.74</td>
<td>5.05</td>
<td>16.05</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>70.93</td>
<td>9.17</td>
<td>20.00</td>
<td>25.45</td>
<td>36.54</td>
<td>18.47</td>
<td>25.80</td>
<td>5.09</td>
<td>16.44</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>90.00</td>
<td>11.00</td>
<td>35.90</td>
<td>60.10</td>
<td>24.50</td>
<td>57.20</td>
<td>35.60</td>
<td>46.60</td>
<td></td>
</tr>
<tr>
<td>eq. GC</td>
<td>WLK</td>
<td>52.74</td>
<td>10.53</td>
<td>27.37</td>
<td>5.05</td>
<td>16.05</td>
<td>5.05</td>
<td>16.05</td>
<td>7.41</td>
<td>16.14</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>54.34</td>
<td>11.45</td>
<td>29.48</td>
<td>5.05</td>
<td>16.05</td>
<td>5.05</td>
<td>16.05</td>
<td>7.41</td>
<td>16.14</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>28.80</td>
<td>35.60</td>
<td>46.60</td>
<td>5.05</td>
<td>16.05</td>
<td>5.05</td>
<td>16.05</td>
<td>7.41</td>
<td>16.14</td>
</tr>
</tbody>
</table>

Joint parameter constancy

<table>
<thead>
<tr>
<th>lags(J)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>eq. GP</td>
<td>WLK</td>
<td>3.46</td>
<td>17.85</td>
<td>0.40</td>
<td>1.75</td>
<td>1.59</td>
<td>20.85</td>
<td>8.75</td>
<td>57.22</td>
<td>66.47</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>3.47</td>
<td>17.88</td>
<td>0.40</td>
<td>1.76</td>
<td>1.60</td>
<td>20.88</td>
<td>8.77</td>
<td>57.23</td>
<td>66.48</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>45.10</td>
<td>73.40</td>
<td>13.10</td>
<td>25.20</td>
<td>24.10</td>
<td>67.00</td>
<td>31.20</td>
<td>11.20</td>
<td>59.40</td>
</tr>
<tr>
<td>eq. GC</td>
<td>WLK</td>
<td>5.81</td>
<td>64.24</td>
<td>37.79</td>
<td>64.33</td>
<td>26.74</td>
<td>8.10</td>
<td>30.22</td>
<td>31.05</td>
<td>34.43</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>5.83</td>
<td>64.27</td>
<td>37.83</td>
<td>64.35</td>
<td>26.76</td>
<td>8.12</td>
<td>30.23</td>
<td>31.09</td>
<td>34.47</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>63.00</td>
<td>70.00</td>
<td>30.00</td>
<td>40.00</td>
<td>63.00</td>
<td>70.00</td>
<td>30.00</td>
<td>40.00</td>
<td>63.00</td>
</tr>
</tbody>
</table>

No additive nonlinearity

<table>
<thead>
<tr>
<th>lags(J)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>eq. GP</td>
<td>WLK</td>
<td>5.81</td>
<td>64.24</td>
<td>37.79</td>
<td>64.33</td>
<td>26.74</td>
<td>8.10</td>
<td>30.22</td>
<td>31.05</td>
<td>34.43</td>
</tr>
<tr>
<td></td>
<td>RAO</td>
<td>5.83</td>
<td>64.27</td>
<td>37.83</td>
<td>64.35</td>
<td>26.76</td>
<td>8.12</td>
<td>30.23</td>
<td>31.09</td>
<td>34.47</td>
</tr>
<tr>
<td></td>
<td>WB</td>
<td>63.00</td>
<td>70.00</td>
<td>30.00</td>
<td>40.00</td>
<td>63.00</td>
<td>70.00</td>
<td>30.00</td>
<td>40.00</td>
<td>63.00</td>
</tr>
</tbody>
</table>
Table 1.3: \( P \)-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using ARCH specification. LM stands for Lagrange-multiplier test. Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>lags(J)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>6.43</td>
<td>3.00</td>
<td>9.02</td>
<td>7.13</td>
<td>10.66</td>
<td>18.70</td>
<td>18.04</td>
<td>22.96</td>
<td>13.96</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Table 1.4: \( P \)-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using smooth transition (ST) specification. LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>t.v.</th>
<th>( \tau )</th>
<th>( y_{1,t-1} )</th>
<th>( y_{1,t-2} )</th>
<th>( y_{1,t-3} )</th>
<th>( y_{1,t-4} )</th>
<th>( y_{1,t-5} )</th>
<th>( y_{1,t-6} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>23.82</td>
<td>13.83</td>
<td>52.23</td>
<td>16.26</td>
<td>0.67</td>
<td>16.32</td>
<td>13.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t.v.</th>
<th>( y_{2,t-1} )</th>
<th>( y_{2,t-2} )</th>
<th>( y_{2,t-3} )</th>
<th>( y_{2,t-4} )</th>
<th>( y_{2,t-5} )</th>
<th>( y_{2,t-6} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>6.07</td>
<td>76.60</td>
<td>88.57</td>
<td>32.24</td>
<td>94.26</td>
<td>84.49</td>
</tr>
</tbody>
</table>

Table 1.5: \( P \)-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.2) using White specification. LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y_{1,t-1} )</th>
<th>( y_{1,t-2} )</th>
<th>( y_{1,t-3} )</th>
<th>( y_{1,t-4} )</th>
<th>( y_{1,t-5} )</th>
<th>( y_{1,t-6} )</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>4.80</td>
<td>34.64</td>
<td>8.43</td>
<td>1.67</td>
<td>10.32</td>
<td>96.80</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y_{2,t-1} )</th>
<th>( y_{2,t-2} )</th>
<th>( y_{2,t-3} )</th>
<th>( y_{2,t-4} )</th>
<th>( y_{2,t-5} )</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>1.71</td>
<td>66.11</td>
<td>74.08</td>
<td>49.34</td>
<td>81.89</td>
<td>69.59</td>
</tr>
</tbody>
</table>
Table 1.6: p-values of misspecification tests for the VAR(15) model against additive nonlinearities. \( Jokul \) stands for the Jökulsá flow equation, and \( Vatns \) for the Vatnsdalsá flow equation. Only the wild bootstrap results from the linearity tests of order 3 are reported. \( t.v. \) stands for transition variables. \( y_{1,t-d} \) stands for the the Jökulsá flow at time \( t-d \), \( y_{2,t-d} \) stands for the the Vatnsdalsá flow at time \( t-d \), \( x_{t-d} \) for the precipitation at time \( t-d \) and \( z_{t-d} \) for the temperature at time \( t-d \). The sample size of the wild bootstrapping is 5000. Note that all of the p-values are multiplied by 100 and hence read percent.

<table>
<thead>
<tr>
<th>t.v.</th>
<th>( Jokul )</th>
<th>( Vatns )</th>
<th>( Jokul )</th>
<th>( Vatns )</th>
<th>( Jokul )</th>
<th>( Vatns )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{1,t-1} )</td>
<td>0.00</td>
<td>3.06</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>( y_{1,t-2} )</td>
<td>0.00</td>
<td>4.38</td>
<td>0.10</td>
<td>0.00</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{1,t-3} )</td>
<td>0.00</td>
<td>1.70</td>
<td>0.80</td>
<td>0.00</td>
<td>7.40</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{1,t-4} )</td>
<td>0.00</td>
<td>15.82</td>
<td>3.70</td>
<td>0.10</td>
<td>10.80</td>
<td>0.90</td>
</tr>
<tr>
<td>( y_{1,t-5} )</td>
<td>0.00</td>
<td>59.20</td>
<td>0.60</td>
<td>0.10</td>
<td>16.80</td>
<td>1.20</td>
</tr>
<tr>
<td>( y_{1,t-6} )</td>
<td>0.00</td>
<td>7.58</td>
<td>5.70</td>
<td>0.00</td>
<td>42.50</td>
<td>0.80</td>
</tr>
<tr>
<td>( y_{1,t-7} )</td>
<td>0.00</td>
<td>10.60</td>
<td>0.10</td>
<td>0.00</td>
<td>6.70</td>
<td>0.10</td>
</tr>
<tr>
<td>( y_{2,t-1} )</td>
<td>1.46</td>
<td>0.00</td>
<td>0.40</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{2,t-2} )</td>
<td>2.18</td>
<td>0.16</td>
<td>0.30</td>
<td>0.00</td>
<td>7.90</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{2,t-3} )</td>
<td>1.54</td>
<td>2.14</td>
<td>0.70</td>
<td>0.00</td>
<td>3.90</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{2,t-4} )</td>
<td>5.36</td>
<td>13.24</td>
<td>3.60</td>
<td>0.00</td>
<td>12.20</td>
<td>0.00</td>
</tr>
<tr>
<td>( y_{2,t-5} )</td>
<td>6.40</td>
<td>33.44</td>
<td>5.50</td>
<td>0.00</td>
<td>1.90</td>
<td>0.20</td>
</tr>
<tr>
<td>( y_{2,t-6} )</td>
<td>4.16</td>
<td>12.80</td>
<td>3.30</td>
<td>0.00</td>
<td>4.60</td>
<td>0.10</td>
</tr>
<tr>
<td>( y_{2,t-7} )</td>
<td>9.28</td>
<td>7.76</td>
<td>9.30</td>
<td>0.00</td>
<td>26.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( x_{t-1} )</td>
<td>0.00</td>
<td>0.02</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>1.40</td>
</tr>
<tr>
<td>( x_{t-2} )</td>
<td>0.18</td>
<td>0.00</td>
<td>11.90</td>
<td>0.00</td>
<td>0.10</td>
<td>1.60</td>
</tr>
<tr>
<td>( x_{t-3} )</td>
<td>8.68</td>
<td>1.04</td>
<td>0.30</td>
<td>0.00</td>
<td>17.40</td>
<td>65.00</td>
</tr>
<tr>
<td>( z_{t} )</td>
<td>0.00</td>
<td>0.00</td>
<td>82.60</td>
<td>10.60</td>
<td>53.70</td>
<td>80.50</td>
</tr>
</tbody>
</table>

Table 1.7: P-values of the joint serial correlation tests of the estimated model (1.8.4) and (1.8.5). \( WB \) stands for the wild bootstrapping. The sample size of the wild bootstrapping is 1000. Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>lags(J)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>WB</td>
<td>36.60</td>
<td>1.20</td>
<td>13.50</td>
<td>14.00</td>
<td>27.60</td>
<td>3.40</td>
<td>12.40</td>
<td>12.10</td>
<td>24.10</td>
<td>10.70</td>
</tr>
</tbody>
</table>
Table 1.8: P-values of the partial and joint parameter constancy tests of the estimated model (1.8.4) and (1.8.5). WB stands for the wild bootstrapping. The sample size of the wild bootstrapping is 1000. Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>Partial and joint parameter constancy test</th>
<th>Jokul</th>
<th>Vatns</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>WB</td>
<td>5.30</td>
<td>45.4</td>
<td>9.70</td>
</tr>
</tbody>
</table>

Table 1.9: P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using ARCH specification. LM stands for Lagrange-multiplier test. Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>Joint test of ECM constancy using ARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>lags(J)</td>
</tr>
<tr>
<td>LM</td>
</tr>
</tbody>
</table>

Table 1.10: P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using smooth transition (ST) specification. LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>Joint test of ECM constancy using ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.v.</td>
</tr>
<tr>
<td>LM</td>
</tr>
<tr>
<td>t.v.</td>
</tr>
<tr>
<td>LM</td>
</tr>
</tbody>
</table>

Table 1.11: P-values of the error covariance matrix (ECM) constancy tests of the estimated model (1.8.4) and (1.8.5) using White specification. LM stands for Lagrange-multiplier test. \( \tau = t/T \). Notice that p-values are all multiplied by 100.

<table>
<thead>
<tr>
<th>Joint test of ECM constancy using White</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
</tr>
<tr>
<td>LM</td>
</tr>
<tr>
<td>x</td>
</tr>
<tr>
<td>LM</td>
</tr>
</tbody>
</table>
Figure 1.1: An example of the grid with a zoom. The first grid points are joined by the dash-dotted black lines, and point A produces the smallest Residuals Sum of Squares (RSS). The second grid points are joined by the dashed blue lines, and point B produces the smallest RSS. The third grid points are joined by the solid red lines, and the point C produces the smallest RSS.
Figure 1.2: Top panel: monthly growth rate of gasoline price (log difference); Bottom panel: monthly growth rate of gasoline consumption (log difference). The sample period is from Feb. 1973 to Dec. 1998, and 311 observations in all.
Figure 1.3: Autocorrelation functions (ACF) and partial autocorrelation function (Partial ACF) of the residuals of the estimated model (1.8.2) for the monthly gasoline price change (GP) and the monthly gasoline consumption change (GC).
Figure 1.4: Transition functions for the gasoline price change equation and the consumption change equation: (Top) the first transition driven by $y_{1,t-4}$ in the price equation, (Middle) the second transition driven by $y_{1,t-5}$ in the price equation, and (Bottom) the transition driven by $y_{1,t-1}$ in the consumption equation. Each point represents an observation.
Figure 1.5: Regime-switching through time for the gasoline price change equation and the consumption change equation: (Top) the first regime-switching driven by $y_{1,t-4}$ in the price equation, (Middle) the second regime-switching driven by $y_{1,t-5}$ in the price equation, and (Bottom) the regime-switching driven by $y_{1,t-1}$ in the consumption equation.
Figure 1.6: Stability of the two processes. The upper plot is the price change equation which converges to about -0.302, and the lower plot is the consumption change equation which converges to about 0.218. We use all the history in the data set as the initial values of the DGP.
Figure 1.7: Generalized impulse response functions of the VLSTAR model (1.8.2) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: gasoline price percent change to positive price shocks (left); gasoline price percent change to negative price shocks (right). Bottom two boxplots: gasoline consumption percent change to positive price shocks (left); gasoline consumption percent change to negative price shocks (right).
Figure 1.8: Generalized impulse response functions of the VLSTAR model (1.8.2) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: gasoline price percent change to positive consumption shocks (left); gasoline price percent change to negative consumption shocks (right). Bottom two boxplots: gasoline consumption percent change to positive consumption shocks (left); gasoline consumption percent change to negative consumption shocks (right).
Chapter 1. Vector smooth transition autoregressive models

Figure 1.9: River flow, precipitation and temperature series.

(a) The daily river flow of the Jökulsá Eystri River $y_{1t}$ measured in $m^3s^{-1}$

(b) The daily river flow of the Vatnsdalsá River $y_{2t}$ measured in $m^3s^{-1}$

(c) The daily precipitation $x_t$ measured in $mm$

(d) The daily temperature $z_t$ measured in °C
Chapter 1. Vector smooth transition autoregressive models

(a) The transition function

(b) Temperature, the transition variable; the turning point (solid) and the interval for the smooth transition (dashed).

(c) Regime-switchings through time.

Figure 1.10: The regime-switching of the Jökulsá flow driven by the temperature.
Figure 1.11: The regime-switching of the Vatnsdalsá flow driven by the temperature.
Chapter 1. Vector smooth transition autoregressive models

(a) The transition function

(b) Precipitation, the transition variable; the turning point (solid) and the interval for the smooth transition (dashed).

(c) Regime-switchings through time.

Figure 1.12: The regime-switching of the Jökulsá flow driven by the precipitation.
(a) The transition function

(b) Precipitation, the transition variable; the turning point (solid) and the interval for the smooth transition (dashed).

(c) Regime-switchings through time.

Figure 1.13: The regime-switching of the Vatnsdalsá flow driven by the precipitation.
Chapter 1. Vector smooth transition autoregressive models

Figure 1.14: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to **positive** Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to **negative** Jökulsá shock below freezing (left) and above freezing (right).
Chapter 1. Vector smooth transition autoregressive models

Figure 1.15: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Jökulsá shock below freezing (left) and above freezing (right).
Chapter 1. Vector smooth transition autoregressive models

Figure 1.16: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Vatnadalshá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Vatnadalshá shock below freezing (left) and above freezing (right).
Chapter 1. Vector smooth transition autoregressive models

Figure 1.17: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Vatnsdalsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Vatnsdalsá shock below freezing (left) and above freezing (right).
Figure 1.18: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Jökulsá shock below freezing (left) and above freezing (right).
Figure 1.19: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Jökulsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Jökulsá shock below freezing (left) and above freezing (right).
Figure 1.20: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Jökulsá to positive Vatnsdalsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Jökulsá to negative Vatnsdalsá shock below freezing (left) and above freezing (right).
Figure 1.21: Generalized impulse response functions of the VLSTAR model in (1.8.4) and (1.8.5) represented with 50 percent (box) and 80 percent (whisker) highest density regions. Top two boxplots: Vatnsdalsá to positive Vatnsdalsá shock below freezing (left) and above freezing (right). Bottom two boxplots: Vatnsdalsá to negative Vatnsdalsá shock below freezing (left) and above freezing (right).
Chapter 2

Linearity and Misspecification Tests for Vector Smooth Transition Regression Models
Linearity and Misspecification Tests for Vector Smooth Transition Regression Models

abstract

In this paper, we propose LM-type misspecification tests in vector smooth transition models. We report the results of the simulation studies, in which the size and power of the proposed tests are considered for high-dimensional systems. We found that the tests in high-dimensional case have great size-distortion, which cannot be overlooked. We suggest several solutions to the size-distortion problem. The Wilks’ A statistic and the Rao’s statistic, which have satisfying size properties, are recommended in empirical use.

2.1 Introduction

Hypothesis testing is an essential part of building and evaluating nonlinear time series models. Many nonlinear models such as the smooth transition regression or switching regression model or their univariate counterparts nest a linear model and are not identified if in fact the linear model has generated the observations. This is why testing linearity is essential before fitting a nonlinear model. Evaluating any time series model before using it, typically for forecasting, is important to ensure the relevance of the empirical results. In this paper we consider testing a linear vector autoregressive (VAR) model against a nonlinear vector logistic smooth transition autoregressive (VLSTAR) or regression (VLSTR) models. Furthermore, we derive various misspecification tests for estimated VLSTAR or VLSTR models.

Eitrheim and Teräsvirta (1996) constructed misspecification tests for univariate STAR models. They include the test of no error autocorrelation, based on considerations in Godfrey (1988, Section 4.4), a test of the hypothesis of no additional nonlinearity, and a third test against parameter nonconstancy. The last two test contained the linearity test and the parameter constancy test in the linear VAR model as special cases. They build on the idea of circumventing the identification problem present in testing as in Luukkonen et al. (1988). Camacho (2004) generalised the test of no error autocorrelation to a bivariate STAR model.

In this paper we work further on linearity and misspecification tests in the VLSTR framework. We allow the dimension of the model exceed two and, furthermore, do not restrict the number of transitions to one. We focus on two cases. In the first one, the VLSTR model only has one transition variable, that is, the same transition variable is controlling nonlinearity in all equations. In the second case, the transition variable need not be the same in all equations, but the set of transition variables is known. This means that it is known which variable belongs

\[1\]This chapter is joint work with Timo Teräsvirta.
to which smooth transition equation. In some applications, there may be underlying theory propositions determining these transition variables and thus justifying this type of test.

When the VLSTR model is extended beyond the bivariate one considered by Camacho (2004), the problem of size distortion emerges. The standard tests tend to be oversized in small samples, sometimes very badly. This is a well known problem in testing vector models, see for example Laitinen (1978), Meisner (1979), Bera et al. (1981) and Edgerton and Shukur (1999) and Shukur and Edgerton (2002). In the former article the authors conducted a large simulation study of tests of no error autocorrelation in linear regression models and found that Rao’s F-test, see Rao (1951) and Rao (1965, Section 8c.5), designed to correct the size, had the best performance. In Shukur and Edgerton (2002) they studied the functional form specification test, RESET, by Ramsey (1969) and came up with a similar conclusion. In this work we simulate our tests and, like the previous authors, consider various remedies to size distortion. This is important because some of our tests can have a large number of parameters in the null hypothesis, and size problems are likely to emerge.

The asymptotic theory of our tests requires that the log-likelihood function is at least twice continuously differentiable in a neighbourhood of the null hypothesis. Theoretically this means that the tests are not valid for vector threshold autoregressive models such as the model by Tsay (1998). In practice, however, our tests do have power even against threshold-type alternatives, see Strikholm and Teräsvirta (2006) and Teräsvirta et al. (2010, Section 16.4) for a discussion of this in the univariate threshold autoregressive case. The tests are principally designed, however, to be applied in the VLSTAR or VLSTR modelling framework of Chapter 1.

The plan of the paper is as follows. In Section 2, we introduce the VLSTAR model developed in Chapter 1. In Section 3, we develop the linearity tests in two cases. First, this is done when a single transition variable controls transitions. Second, tests are derived for a situation in which each equation also adjusts its own transition function and transition variable. In Section 4, we discuss about the size distortion problem often arising in testing vector models and propose test statistics that alleviate this problem. In Section 5, several misspecification tests for model evaluation are derived. In Section 5, we design simulation experiments to investigate the size properties of the tests and report the results. Section 5 concludes.

2.2 The vector logistic smooth transition model

In Chapter 1, we define the Vector Logistic Smooth Transition AutoRegressive (VLSTAR) model with lag \( k \) as follows:

\[
y_t = \left\{ \sum_{i=1}^{m} (G_i^{i-1} - G_i^i)F_i^t \right\} x_t + \varepsilon_t
\]  \hspace{1cm} (2.2.1)

where \( y_t \) is a \( p \times 1 \) column vector, \( x_t = (y_{t-1}', ..., y_{t-k}', d_t')' \) is a \( (kp + q) \times 1 \) vector in which \( d_t \) is a \( q \times 1 \) vector of intercept and any exogenous variables. \( F_i \) is a \( (kp + q) \times p \) matrix of linear parameters: \( F_i = (A_{i1}', ..., A_{ik}', \Phi_i')' \), where each \( A_{ij}, i = 1, ..., m, j = 1, ..., k \), is a \( p \times p \) matrix,
and each $\Phi_i$, $i = 1, \ldots, m$, is a $q \times p$ matrix. The $p \times 1$ error vector $\varepsilon_t$ is assumed i.i.d. $N(0, \Omega)$, where the covariance matrix $\Omega$ is positive definite.

$G_i$ is a diagonal matrix of transition functions which takes the form:

$$G_i^t = \text{diag} \{ g(s_{1it} | \gamma_{1i}, c_{1i}), \ldots, g(s_{pit} | \gamma_{pi}, c_{pi}) \} \quad (2.2.2)$$

for $i = 1, \ldots, m - 1$, and $G_i^0 = I_p$. The diagonal elements of $G_i$ in (2.2.2) are logistic functions of their transition variables:

$$g(s_{ijt} | \gamma_{ij}, c_{ij}) = (1 + \exp\{-\gamma_{ij} (s_{ijt} - c_{ij})\})^{-1}, \quad \gamma_{ij} > 0 \quad (2.2.3)$$

for $i = 1, \ldots, m - 1$ and $j = 1, \ldots, p$.

We consider the VLSTAR model (2.2.1) using the following reparametrization:

$$y_t = (B_1' + G_1^t B_2' + \ldots + G_{m-1}^t B_m') x_t + \varepsilon_t = \Psi_t B' x_t + \varepsilon_t \quad (2.2.4)$$

where $\Psi_t = (I_p, G_1^t, \ldots, G_{m-1}^t)'$ is a $mp \times p$ full rank matrix. Furthermore, $B = (B_1, B_2, \ldots, B_m)$ is a $(kp + q) \times mp$ matrix, where $B_1 = F_1$, and $B_i = F_i - F_{i-1}$, $i = 2, \ldots, m$.

As $\varepsilon_t \sim i.i.d. N(0, \Omega)$, the conditional log-likelihood function takes the form

$$\log L(\theta) = -(T/2) \log 2\pi - (T/2) \log |\Omega|$$

$$- (1/2) \sum_{t=1}^{T} (y_t - \Psi_t B' x_t)' \Omega^{-1} (y_t - \Psi_t B' x_t) \quad (2.2.5)$$

The set of parameters to be estimated is $\theta = \{B, \Omega, \Gamma, C\}$, where $\Gamma = [\gamma_{ij}]$ and $C = [c_{ij}]$ contain the parameters in the transition functions.

Hereafter, we assume in addition that:

**Assumption 2.2.1.** The log-likelihood function $\log L(\theta)$ is second-order differentiable with respect to the parameters $\theta$.

**Assumption 2.2.2.** Both the first-order derivative and second-order derivative of the log-likelihood function $\log L$ with respect to the parameters $\theta \in \Theta$ are $L_T$-integrable.

**Assumption 2.2.3.** We assume that the following limiting information exists

$$\lim_{T \to \infty} T^{-1} E_\theta(i_T(\theta)) = i(\theta) < \infty \quad (2.2.6)$$

where $i_T(\theta) = -\partial^2 \log L / \partial \theta^2$.

These are the regularity conditions introduced by Feigin (1976).

### 2.3 Joint linearity tests

#### 2.3.1 Testing linearity against VLSTAR with single transition variable

When we consider hypothesis testing in the VLSTAR framework, we restrict ourselves to the case in which the diagonal elements of the transition matrix $G_i$ are logistic functions. In this
section, we consider the special case in which there is only one single transition variable, that is, \( s_{1t} = s_{2t} = ... = s_{pt} = s_t \). In the more general case where each equation may have a different transition variable, testing has to be first carried out equation by equation as in in Luukkonen et al. (1988) and Camacho (2004), and if necessary, testing different combinations of transition variables may be carried out.

As already indicated, testing linearity against STAR is complicated by the presence of unidentified nuisance parameters under the null hypothesis. As a result, the asymptotic null distribution of the classical likelihood ratio, the Lagrange multiplier and the Wald type tests remains unknown. The solution of Saikkonen and Luukkonen (1988) and Luukkonen et al. (1988) to this problem is to replace the transition function by a suitable Taylor series approximation. We generalise this approach to the multivariate case with a single transition variable.

Consider the \( p \)-dimensional 2-regime logistic VSTAR model in (2.2.4): 

\[
y_t = B'_1 x_t + G_t B'_2 x_t + \epsilon_t
\]

where the sequence \( \{ \epsilon_t \} \) is \( i.i.d. \ N(0, \Omega) \). The null hypothesis of linearity can be written as \( H_0 : \gamma_j = 0, j = 1, ..., p \). When the null holds true, \( G_t \equiv (1/2)I_p \) and (2.3.1) becomes linear, while the location parameters \( c_j \) in the logistic functions and the linear combination \( B_1 + (1/2)B_2 \) are not identified. The alternative hypothesis is \( H_1 : \) at least one \( \gamma_j > 0, j = 1, ..., p \).

In order to solve the identification problem, we approximate the logistic function \( g(s_t|\gamma_j, c_j) \) with a first-order Taylor approximation around \( \gamma_j = 0 \) as proposed in Luukkonen et al. (1988). This gives 

\[
g(s_t|\gamma_j, c_j) = (1/2) + (1/4)\gamma_j (s_t - c_j) + r_{jt} = a_j s_t + b_j + r_{jt}
\]

where \( a_j = \gamma_j / 4, b_j = 1/2 - a_j c_j \), and \( r_{jt} \) is the remainder term of the Taylor expansion. We rewrite the \( G_t \) matrix as follows:

\[
G_t = \text{diag}\{ a_1 s_t + b_1 + r_{1t}, ..., a_p s_t + b_p + r_{pt} \}
\]

\[
= A s_t + B + R_t
\]

where \( A = \text{diag}(a_1, ..., a_p) \), \( B = \text{diag}(b_1, ..., b_p) \), and \( R_t = \text{diag}(r_{1t}, ..., r_{pt}) \).

Consider now the reparametrization

\[
y_t = B'_1 x_t + (A s_t + B + R_t) B'_2 x_t + \epsilon_t
\]

\[
= (B'_1 + BB'_2) x_t + A B'_2 x_t s_t + R_t B'_2 x_t + \epsilon_t
\]

\[
= \Theta'_0 x_t + \Theta'_1 x_t s_t + \epsilon^*_t
\]

(2.3.3)

where \( \Theta_0 = B_1 + B_2 B, \Theta_1 = B_2 A \), and \( \epsilon^*_t = R_t B'_2 x_t + \epsilon_t \). The null hypothesis implies \( A = 0, B = (1/2)I_p \), and \( R_t = 0 \) in (2.3.2). Thus, under the null hypothesis, the model (2.3.3) is linear, with \( \Theta_0 = B_1 \) and \( \Theta_1 = 0 \). The linearity test is therefore equivalent to the test \( H_0 : \Theta_1 = 0 \) in (2.3.3). Moreover, due to the fact that \( R_t = 0 \) under the null hypothesis, the error term \( \epsilon^*_t = \epsilon_t \). Since the Lagrange multiplier test only requires estimating the model under the null
Chapter 2. Linearity and misspecification tests

hypothesis, the remainder term does not affect the normality of the errors or the standard asymptotic inference.

The corresponding Lagrange multiplier test under the null is derived from the score

$$\frac{\partial \log L(\hat{\theta})}{\partial \Theta} = \sum_{t=1}^{T} x_t s_t \left( y_t - \hat{B}' x_t \right)' \hat{\Omega}^{-1} = Z_1' \left( Y - X \hat{B}_1 \right) \hat{\Omega}^{-1}$$

(2.3.4)

where

$$Y = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_T \end{bmatrix}, \quad X = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_T \end{bmatrix}, \quad Z_1 = \begin{bmatrix} x'_1 s_1 \\ x'_2 s_2 \\ \vdots \\ x'_T s_T \end{bmatrix}$$

and \( \hat{B}_1 \) and \( \hat{\Omega} \) are estimates from the restricted model. Under regularity conditions, the score converges to a matrixvariate normal distribution with zero mean and variance \( Z_1 (I_T - P_x) Z_1 \otimes \Omega^{-1} \), where \( P_x \equiv X(X'X)^{-1}X' \) is the projection matrix.

**Theorem 2.3.1.** Consider the model in (2.3.1) and its reparametrization (2.3.3). Under the null hypothesis that \( H_0 : \gamma_j = 0, j = 1, ..., p \) in (2.3.1), or equivalently, \( H_0 : \Theta_1 = 0 \) in (2.3.3), the LM test statistic has the form:

$$LM_1 = \text{tr}\{ \hat{\Omega}^{-1} (Y - X \hat{B}_1)' Z_1 [Z_1'(I_T - P_x)Z_1]^{-1} Z_1'(Y - X \hat{B}_1) \}. \quad (2.3.5)$$

Under the null hypothesis, the test statistic has an asymptotic \( \chi^2 \) distribution with \( p(k_p + q) \) degrees of freedom.

**Proof.** See Appendix A. \( \square \)

The subscript 1 in \( Z_1 \) indicates that the test is based on the first-order Taylor expansion of the transition function.

Recall that the explanatory variable \( x_t \) may contain the intercept, seasonal dummies, trend and other deterministic terms. In the case that \( x_t \) contains the intercept, and the transition variable \( s_t = y_{t-d,j} \) for some \( 1 \leq d \leq k \) and \( 1 \leq j \leq p \), the column vector \( (s_1, ..., s_T)' \) must be omitted from \( Z_1 \) to avoid collinearity. Thus the degrees of freedom under the null change to \( p(k_p + q - 1) \). The number of degrees of freedom equals the number of restrictions, i.e., \( p \) multiplied by the column dimension of \( Z_1 \).

The test can be performed as follows:

1. Estimate the restricted model: regress \( Y \) on \( X \). Collect the residuals \( \hat{E} = (I_T - P_x) Y \), and the matrix residual sum of squares \( \text{RSS}_0 = \hat{E}' \hat{E} \).
2. Run an auxiliary regression of \( \hat{E} \) on \( (X, Z_1) \). Collect the residuals \( \hat{\Xi} \), and compute the matrix residual sum of squares \( \text{RSS}_1 = \hat{\Xi}' \hat{\Xi} \).
3. Compute the test statistic

$$LM = T \text{tr}\{ \text{RSS}_0^{-1}(\text{RSS}_0 - \text{RSS}_1) \} \quad (2.3.6)$$

$$= T(p - \text{tr}\{ \text{RSS}_0^{-1}\text{RSS}_1 \}) \quad (2.3.6)$$

64
The joint test statistic defined in Theorem 2.3.1 collapses into the univariate LM-type linearity test statistic when \( p = 1 \). This joint test can also be applied to any subset of equations in the system, for instance, to check whether some equations in the system are nonlinear with a common transition variable.

### 2.3.2 Testing linearity against VLSTAR with different transition variables

Now we turn to the VLSTAR model in which every equation can have its own transition variable, \( s_{jt} \), \( j = 1, \ldots, p \), and do not exclude the possibility that some equations in the system are linear. Linearity testing will then be carried out equation by equation.

As there may be more than one transition variable candidate for any equation, we test the linearity for each of them. The basic rule of the VLSTAR model specification is to select the transition variable that produces the strongest rejection of linearity measured by the \( p \)-value of the tests. However, sometimes there may be transition variables that should not be immediately excluded from the considerations, because testing with them also produces \( p \)-values close to the smallest one. It is common in applications that one equation may have more than one reasonable candidate of transition variable left after all linearity tests have been carried out.

If some equations have more than one reasonable transition variables after testing linearity, combinations of them could be considered. Note that each STR equation merely contains one transition variable. One way of modelling the nonlinearity is to estimate all relevant combinations and select the best model by means of evaluation tests. But then, a joint test of linearity against the VLSTAR model with a predetermined combination of transition variables may be advisable. In that case, the combination leading to the strongest rejection of the linearity hypothesis will be selected.

The solution of Saikkonen and Luukkonen (1988) and Luukkonen et al. (1988) is applicable even here. We replace the transition function by first-order Taylor expansion, which yields

\[
G_t = S_t A + B + R_t,
\]

(2.3.7)

where \( A \), \( B \) and \( R_t \) have been defined in (2.3.2), and \( S_t = \text{diag}(s_{1t}, s_{2t}, \ldots, s_{pt}) \). This nests the case where \( s_{kt} = s_{lt} \) for \( k \neq l \). Reparameterizing (2.3.3) gives

\[
y_t = \Theta_0 x_t + S_t \Theta_1 x_t + \varepsilon_t^*,
\]

(2.3.8)

where \( \Theta_0 \), \( \Theta_1 \) and \( \varepsilon_t^* \) are defined as in (2.3.3).

The null hypothesis of linearity, that is, \( \gamma_{ij} = 0 \) for all \( i = 1, \ldots, m - 1 \) and \( j = 1, \ldots, p \), can be equivalently stated as \( H_0 : \Theta_1 = 0 \). Then we have an LM test similar to that in (2.3.5), with the difference that \( Z_1 = (\text{vec}(S_1 \otimes x_1), \ldots, \text{vec}(S_T \otimes x_T))^\prime \), see Appendix B.

One can carry out this test by using the procedure given in the previous section, just setting \( Z_1 = (\text{vec}(S_1 \otimes x_1), \ldots, \text{vec}(S_T \otimes x_T))^\prime \). The drawback of doing so, however, is that the dimension of the null hypothesis quickly becomes large, which may affect the empirical size of the test. This issue is considered in Section 2.4.
This is a general joint test of the test given in the previous section. However, some of the transition variables can still be equal, that is, \( s_{kt} = s_{lt} \) for some \( k \neq l \). This implies that the matrix \( Z_1 \) has reduced rank and some of its columns must be removed to avoid collinearity.

In addition, we suggest a more parsimonious way of carrying out the test:

1. Estimate the restricted model under the null hypothesis. Collect the residuals \( \tilde{\varepsilon}_t = y_t - B_1'x_t \) and form \( \tilde{\mathbf{E}} = (\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_T)' \). Compute the matrix residual sum of squares \( \text{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}} \).

2. Run an auxiliary regression for each column of \( \tilde{\mathbf{E}} \), that is, regress \( \tilde{\varepsilon}_{jt} \), \( j = 1, ..., p \), on \( x_t \) and \( x_ts_{jt} \), which yields the residual vectors \( \tilde{\xi}_t = (\tilde{\xi}_{1t}, ..., \tilde{\xi}_{pt})' \). Form the residual matrix \( \tilde{\mathbf{\xi}} \), and compute the matrix residual sum of squares \( \text{RSS}_1 = \tilde{\mathbf{\xi}}'\tilde{\mathbf{\xi}} \).

3. Compute the test statistic in (2.3.6).

The corresponding degrees of freedom of the limiting distribution of the test statistic equal the number of parameters in \( \Theta_1 \). The disadvantage of this short-cut is that the limiting null distribution of the LM statistic remains unknown and the empirical null distribution has to be obtained by simulation.

This means in practice that the critical value is obtained using a standard bootstrap assuming that the errors are independent. This is done as follows. First, we estimate the restricted model under the null hypothesis. Then we generate \( y_t \) using \( x_t \), the estimates of the parameters from the restricted model and the bootstrapped residuals, and compute the test statistic using the parsimonious algorithm suggested above. Repeating the sampling procedure and the computation of the test statistic for \( N \) times, we obtain \( N \) values of the test statistics, under the null hypothesis. After ordering them we select the relevant quantile to be the critical value at a given significance level.

A simple simulation-based size experiment is conducted here. We choose the simple bivariate data generating process as follows:

\[
y_{it} = 0.8y_{i,t-1} + \varepsilon_{it} \tag{2.3.9}
\]

where \( i = 1, 2 \). Let \( T = 50, 100, 200, 500 \). The covariance matrix of the errors is

\[
\Omega = \begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix} \tag{2.3.10}
\]

where \( \rho = 0.9, 0, -0.9 \). We choose \( N = 10000 \) to obtain the empirical null distribution of the test. We repeat the bootstrapping test for 2000 times and report the empirical size in Table 2.1. We see that they are satisfactory.

### 2.3.3 Higher-order linearity tests

Luukkonen et al. (1988) pointed out that in the situations where \( s_t = y_{t-d, j} \) for some \( 1 \leq d \leq k \), \( 1 \leq j \leq p \), and only the intercept fluctuates across regimes, the \( LM_1 \) statistic only has trivial
Table 2.1: The empirical size of the bootstrapping test.

<table>
<thead>
<tr>
<th>Nom. size</th>
<th>$\rho = 0.9$</th>
<th>$\rho = 0$</th>
<th>$\rho = -0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>1.35% 5.90% 10.90%</td>
<td>0.90% 5.00% 9.65%</td>
<td>1.10% 4.50% 8.50%</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>1.05% 5.15% 9.85%</td>
<td>1.25% 4.95% 10%</td>
<td>0.9% 5.35% 9.85%</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>1.00% 5.25% 10.40%</td>
<td>0.80% 4.65% 9.55%</td>
<td>0.9% 4.95% 9.20%</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>1.05% 5.60% 10.95%</td>
<td>1.35% 5.60% 9.85%</td>
<td>1.40% 5.25% 9.55%</td>
</tr>
</tbody>
</table>

power against the alternative. They solved this problem by applying the third-order Taylor expansion instead of the first-order one. This method is also applicable here and leads to the following auxiliary regression

$$y_t = \Theta_0' x_t + \Theta_1' x_t s_t + \Theta_2' x_t s_t^2 + \Theta_3' x_t s_t^3 + \varepsilon_t^*.$$ (2.3.11)

Although the third-order linearity test of $H_0: \gamma_j = 0$ becomes, after the Taylor approximation, $H_0': \Theta_1 = \Theta_2 = \Theta_3 = 0$. Similarly, we have the third order LM test statistic

$$LM_3 = \text{tr}\{\tilde{\Omega}^{-1}(Y - X\tilde{B}_1)'Z_3 [Z_3'(I - P_x)Z_3]^{-1}Z_3'(Y - X\tilde{B}_1)\},$$

where

$$Z_3 = \begin{bmatrix} x_1's_1 & x_1's_1^2 & x_1's_1^3 \\ x_2's_2 & x_2's_2^2 & x_2's_2^3 \\ \vdots & \vdots & \vdots \\ x_T's_T & x_T's_T^2 & x_T's_T^3 \end{bmatrix}$$

Under the null hypothesis, the test statistic has an asymptotic $\chi^2$ distribution with $3(p(kp+q)$ degrees of freedom. If the explanatory variable $x_t$ contains the intercept, and $s_t = y_{t-d,j}$ for some $1 \leq d \leq k$, $1 \leq j \leq p$, several column vectors should be omitted from the $Z_3$ matrix to avoid the perfect multicollinearity.

Although the third-order test statistic $LM_3$ is designed for the particular case when only the intercept differs across regimes, it can be applied more generally. A parsimonious version of the $LM_3$ statistic that only works for the situation in which the intercept vector is the only nonlinear component of the model, can be obtained by augmenting the first order test of (2.3.3) with regressors $s_t^2$ and $s_t^3$. Thus, we have the parsimonious third-order $LM$ test, which is denoted as $LM^e_3$

$$LM^e_3 = \text{tr}\{\tilde{\Omega}^{-1}(Y - X\tilde{B}_1)'Z_3^e [Z_3^e'(I - P_x)Z_3^e]^{-1}Z_3^e'(Y - X\tilde{B}_1)\}$$

where

$$Z_3^e = \begin{bmatrix} x_1's_1 & s_1^2 & s_1^3 \\ x_2's_2 & s_2^2 & s_2^3 \\ \vdots & \vdots & \vdots \\ x_T's_T & s_T^2 & s_T^3 \end{bmatrix}$$
Chapter 2. Linearity and misspecification tests

The test statistic has an asymptotic $\chi^2$ distribution and its number of degrees of freedom under the null is $p(kp + q + 2)$. These two higher-order linearity tests can be carried out using the steps in (2.3.6) and replacing $Z_1$ by $Z_3$ or $Z_5$, depending on the alternative hypothesis.

The joint linearity test with different transition variables may also have the higher-order variation. However, practically we do not use the joint test with different transition variables in modelling. Instead, we test equation by equation. The joint test will be used when the model has the same but unknown transition variable throughout equations. Furthermore, the limiting distribution of the joint test with different transition variables remains unknown even for order one. We leave it to future work.

2.4 Linearity test statistic with improved size

In small samples, the test of the $\chi^2$-family in Sections 2.3.1 and 2.3.3 are likely to suffer from the problem that the number of observations does not suffice for the asymptotic inference to be an adequate approximation to the unknown finite sample null distribution. This leads to positive size distortion: the empirical size of the test exceeds the corresponding asymptotic size. This is the case in particular when $p$, the dimension of $y_t$, is high. In single-equation models, it is often recommended to use the $F$-version of the $LM$ statistic, but this idea is not applicable in the multivariate case.

One straightforward solution to this problem is to employ the bootstrap testing procedure. The bootstrap method is a sort of intensive computational methods which requires large sample size in simulation so that the accuracy of the $p$-values can be achieved. Although the cost of computation (in time and space or memory) becomes cheaper and cheaper recently, it still can be rather time-consuming and cannot be afforded in high-dimension complex models. There is obviously a trade-off between applying the bootstrap procedure and improving the test statistic. If the improvement of the test statistic can be satisfactory to some extent, we prefer this kind of improvement.

Bartlett and Bartlett-type corrections have been widely used as a possible remedy to the size problem of $LM$ type tests. We consider the Laitinen-Beisner correction as an approximation to the exact test. It consists of a degrees of freedom rescaling of the form $(pT - K)/(G \cdot pT)$, where $p$ and $T$ are as before, $K$ is the number of parameters, and $G$ the number of restrictions, see Laitinen (1978) and Meisner (1979). The $F$-type $LM$ test statistic, or rescaled $LM$ test statistic, can be computed as follows

$$F = \frac{(pT - K)}{G \cdot pT} \frac{LM}{LM}$$

where $LM$ represents any of the three tests $LM_1$, $LM_3$ and $LM_5$ designed in previous sections. The rescaled test statistic is assumed to follow an $F(G, pT - K)$ distribution. In the following, it will be called the rescaled LM test.

The Monte Carlo results of Bera et al. (1981) show that the Laitinen-Beisner correction is likely to overcorrect the size. We propose two improvements. The first one is based on the
so-called Wilks’s Λ-distribution, and we shall call it Wilks’s statistic. Before introducing this statistic, we state the following result:

**Theorem 2.4.1.** Let $\mathbf{RSS}_j$, $j = 0, 1$, be the $p \times p$ residual sum of squares matrix from the restricted regression ($j = 0$) and the auxiliary regression ($j = 1$). Furthermore, let $\mathbf{W}_1 = \mathbf{RSS}_0 - \mathbf{RSS}_1$, and $\mathbf{W}_2 = \mathbf{RSS}_1$. Under the null hypothesis of linearity, $\mathbf{W}_1$ and $\mathbf{W}_2$ are two independent Wishart distributed random matrices:

\[
\mathbf{W}_1 \sim \mathcal{W}_p(\Omega, \cd(\mathbf{Z})) \quad \mathbf{W}_2 \sim \mathcal{W}_p(\Omega, T - \cd(\mathbf{X}) - \cd(\mathbf{Z}))
\]  

where $\cd(\cdot)$ is the column dimension of a matrix.

**Proof.** See Appendix C.

Matrix $\mathbf{Z}$ in (2.4.2) can be any of the three matrices $\mathbf{Z}_1$, $\mathbf{Z}_3$, and $\mathbf{Z}_5$, and $\Omega$ is the covariance matrix of errors under $H_0$. It is worth stressing that, in the special case $p = 1$, the two independent Wishart variables $\mathbf{W}_1$ and $\mathbf{W}_2$ become scalars and $\chi^2$-distributed, which implies an $F$ test.

We define Wilks’s Λ-distribution as follows:

**Definition 2.4.2.** When $\mathbf{A} \sim \mathcal{W}_p(\Sigma, m)$ and $\mathbf{B} \sim \mathcal{W}_p(\Sigma, n)$ are independent, $\Sigma$ is a $p \times p$ positive definite matrix, $m \geq p$,

\[
\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A} + \mathbf{B}|} = |\mathbf{I}_p + \mathbf{A}^{-1}\mathbf{B}|^{-1} \sim \mathcal{L}(p, m, n)
\]

has a Wilks’s Λ-distribution with parameters $p$, $m$, and $n$.

The above definition is a variant of Definition 3.7.1 in Mardia et al. (1979). Anderson (1958, Section 8.3) and Mardia et al. (1979) contain a detailed discussion of the Wilks’s Λ distribution. The distribution is invariant under changes in the covariance matrix $\Sigma$. Wilks’s Λ statistic has the following form:

\[
\Lambda = \frac{|\mathbf{W}_2|}{|\mathbf{W}_2 + \mathbf{W}_1|} = \frac{|\mathbf{RSS}_1|}{|\mathbf{RSS}_0|}.
\]  

It follows Wilks’s Λ-distribution $\mathcal{L}(p, T - \cd(\mathbf{X}) - \cd(\mathbf{Z}), \cd(\mathbf{Z}))$ under the null hypothesis of linearity. If $T$ is large, we may use Bartlett’s approximation

\[
\lambda = \left(\frac{1}{2}(p + \cd(\mathbf{Z}) + 1) + \cd(\mathbf{X}) - T\right) \log \Lambda \sim \chi^2_{\cd(\mathbf{Z})p},
\]

see Bartlett (1954) and Anderson (1958, Section 8.3). The value of the test statistic can be computed by performing steps 1 and 2 outlined in the algorithm in Section 2.3.1 but computing the value of the test statistic defined in (2.4.4) and (2.4.5) instead of step 3. It is worth stressing that, in the special case $p = 1$, replacing $\log \Lambda$ in (2.4.5) by its first-order Taylor expansion around $\Lambda = 1$, we obtain the common $\chi^2$-type LM test.
Chapter 2. Linearity and misspecification tests

Rao (1951) and Rao (1965, Section 8c.5) define yet another test statistic. It provides a useful approximation to the unknown null distribution in small samples, if it is used with critical values from an F-distribution. The statistic is defined as follows:

\[ F_{Rao} = \left( \frac{\left| RSS_0 \right|}{RSS_1} \right)^{1/s} - 1 \left( \frac{Ns - (1/2)cd(Z)p + 1}{cd(Z)p} \right) \tag{2.4.6} \]

where

\[ s = \left( \frac{cd^2(Z)p^2 - 4}{p^2 + cd^2(Z) - 5} \right)^{1/2}, \quad N = T - cd(X) - (1/2) (p + cd(Z) + 1). \]

The corresponding degrees of freedom of (2.4.6) are \( cd(Z)p \) and \( Ns - (1/2)cd(Z)p + 1 \). Similarly to Wilks’s \( \Lambda \), the test can be carried out by performing steps 1 and 2 outlined in the algorithm in Section 2.3.1, and then computing the value of the statistic (2.4.6). When \( p = 1 \) and \( s = 1 \), (2.4.6) becomes the F-type LM test.

2.5 Evaluation tests

The evaluation stage of the VLSTAR modelling strategy designed and applied in Chapter 1 makes use of misspecification tests of the estimated VLSTAR model. In this section we consider three such tests that are either Lagrange multiplier or Lagrange multiplier type tests. All of them are likely to suffer from the problem that the number of observations does not suffice for the asymptotic inference, which can lead to significant size distortion. Fortunately, Wilks’s \( \Lambda \) and Rao’s F statistic considered in section 4, can be applied even here to alleviate the size distortion problem.

2.5.1 Serial correlation in the error process

First, we extend the Lagrange multiplier test of no serial correlation of Eitheim and Teräsvirta (1996) to the multivariate case. Camacho (2004) considered this extension in the bivariate STR model. Assume the \( p \)-dimensional \( m \)-regime nonlinear VLSTAR model with autocorrelated errors:

\[ y_t = \{ \sum_{i=1}^{m} (G_i^{-1} - G_i^L) F_i^t \} x_t + u_t = \Psi_t' B' x_t + u_t \tag{2.5.1} \]

where

\[ u_t = \sum_{i=1}^{J} P'_i u_{t-i} + \varepsilon_t = P(L) u_t + \varepsilon_t \tag{2.5.2} \]

In (2.5.2), \( P_i \) is a \( p \times p \) matrix, \( P(L) = \sum_{i=1}^{J} P'_i L^i \), \( L \) is the lag operator, \( J \) is the lag length, and \( \varepsilon_t \sim i.i.d. N(0, \Omega) \) is a \( p \times 1 \) vector. We assume that the roots of the polynomial \( |I_p - \sum_{i=1}^{J} P'_i z^i| \) lie outside the unit circle. Furthermore, we assume that the sequence of \( \{y_t\} \) is stationary and ergodic such that the parameters can be estimated consistently under the null hypothesis of no serial correlation \( \{u_t\} \), that is, when

\[ H_0 : P_1 = P_2 = ... = P_J = 0 \]
holds. Left-multiplying (2.5.1) by \( P(L) \) yields
\[
y_t = P(L)(y_t - \Psi' B' x_t + \varepsilon_t)
\]
which gives
\[
y_t = P'Lz_t + \Psi' B' x_t + \varepsilon_t
\]

(2.5.3)

see Godfrey (1988), where

\[
P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_J \end{bmatrix}, \quad z_t = \begin{bmatrix} y_{t-1} - \Psi'_{t-1} B' x_{t-1} \\ y_{t-2} - \Psi'_{t-2} B' x_{t-2} \\ \vdots \\ y_{t-j} - \Psi'_{t-j} B' x_{t-j} \end{bmatrix} = \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-j} \end{bmatrix}.
\]

Thus we have the corresponding log-likelihood function
\[
\log L = -(T-J)p/2) \log 2\pi - ((T-J)/2) \log |\Omega|
\]

\[
-(1/2) \sum_{t=J+1}^{T} (y_t - P' z_t - \Psi' B' x_t)' \Omega^{-1} (y_t - P' z_t - \Psi' B' x_t)
\]

(2.5.4)

The Lagrange multiplier test is based on the score under the null hypothesis
\[
\frac{\partial \log L(\hat{\theta})}{\partial P} = \sum_{t=J+1}^{T} \begin{bmatrix} \hat{z}_t \hat{u}_t' \Omega^{-1} \end{bmatrix} = \hat{Z}' \hat{U} \Omega^{-1}
\]

(2.5.5)

where

\[
\hat{Z} = \begin{bmatrix} \hat{z}_{J+1}' \\ \hat{z}_{J+2}' \\ \vdots \\ \hat{z}_T' \end{bmatrix}, \quad \hat{U} = \begin{bmatrix} \hat{u}_{J+1}' \\ \hat{u}_{J+2}' \\ \vdots \\ \hat{u}_T' \end{bmatrix}
\]

and \( \hat{z}_t, \hat{u}_t \) and \( \hat{\Omega} \) are estimates under the null hypothesis. Notice that the subscript of \( \hat{Z} \) and \( \hat{U} \) should start from \( J + 1 \), because usually \( \hat{u}_0, \hat{u}_{-1}, \ldots, \hat{u}_{1-J} \) are not available. Following Luukkonen et al. (1988), we have the following theorem:

**Theorem 2.5.1.** Consider the model (2.5.1). In order to test the null hypothesis \( P_1 = P_2 = \ldots = P_J = 0 \), we have the LM test statistic

\[
LM = \text{tr}\{\hat{\Omega}^{-1} \hat{U}' \hat{Z} (I_{T-J} - \hat{K}' \hat{K}^{-1} \hat{K}') \hat{Z}' \hat{U} \}
\]

(2.5.6)

where

\[
\hat{K} = \begin{bmatrix} \text{vec}[\partial (\hat{\Psi}'_{J+1} \hat{B}' x_{J+1})/\partial \theta]' \\ \text{vec}[\partial (\hat{\Psi}'_{J+2} \hat{B}' x_{J+2})/\partial \theta]' \\ \vdots \\ \text{vec}[\partial (\hat{\Psi}'_T \hat{B}' x_T)/\partial \theta]' \end{bmatrix}.
\]
Assume that the estimates of the parameters in \( \theta \), defined in Appendix D, are consistent. Under the null hypothesis the test statistic (2.5.5) has an asymptotic \( \chi^2 \) distribution with \( Jp^2 \) degrees of freedom.

Proof. See Appendix B.

The vectorised first order derivatives of \( \Psi' B' x_t \) w.r.t. parameters \( \theta \) can be easily found in both univariate and multivariate cases, see Eitrheim and Teräsvirta (1996). We summarize them in Appendix D. We have to mention that if the error terms are not normally distributed, there are moment conditions for the asymptotic distribution theory to be valid. In univariate case, a STAR model with logistic type transition functions must satisfy the condition \( E(u_t^4) < \infty \). However, in multivariate case, the moment condition is not so straightforward. A sufficient one can be \( E(u_{jt}^4) < \infty \) in \( u_t \), for \( j = 1, ..., p \).

In the multivariate case, the column dimension of the \( \tilde{K} \) matrix is \( (pk + q)m + 2(m - 1) \), which grows very fast as a function of \( m \) and \( p \). It is seen that \( T - J \geq (pk + q)m + 2(m - 1) \) is a necessary condition for \( \tilde{K}' \tilde{K} \) to have full rank, and consequently, the existence of the LM statistic. If either \( p \) or \( m \) or both are large, inverting \( \tilde{K}' \tilde{K} \) requires care. The matrix may be near-singular and the inversion slow. This is bound to restrict the size of the null hypothesis, in practice the lag length \( J \), in small and moderate samples, unless the dimension of the model is sufficiently low.

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.3.1 by replacing \( X \) by \( \tilde{K} \) and \( Z_1 \) by \( \tilde{Z} \). It contains the following steps:

1. Estimate the VLSTAR model under the null hypothesis of no serial correlation. Choose the lag length \( J \), and collect the residual vectors. Form the matrix residual sum of squares \( \text{RSS}_0 = \tilde{U}' \tilde{U} \), where \( \tilde{U} = (\tilde{u}_{J+1}, ..., \tilde{u}_T)' \).

2. Run the auxiliary regression of \( \tilde{U} \) on \( (\tilde{K}, \tilde{Z}) \) from \( t = J + 1 \) to \( T \). Collect the residuals \( \tilde{\Xi} \) and form the matrix residual sum of squares \( \text{RSS}_1 = \tilde{\Xi}' \tilde{\Xi} \).

3. Compute the test statistic

\[
LM = (T - J) \text{tr} \{ \text{RSS}_0^{-1} (\text{RSS}_0 - \text{RSS}_1) \}
\]

\[
= (T - J)(p - \text{tr} \{ \text{RSS}_0^{-1} \text{RSS}_1 \})
\]

After obtaining \( \text{RSS}_0 \) and \( \text{RSS}_1 \), we can apply the tests with improved size suggested in Section 2.4. Nevertheless, there is the possibility of positive size distortion even in the tests with improved empirical size. This is due to the fact that in practice \( \tilde{U} \) may not be completely orthogonal to the gradient matrix \( \tilde{K} \), which distorts the size. This is because the LM test statistic obtained from the two-step auxiliary regression is equivalent to (2.5.5) only when \( \tilde{U}' \tilde{K} = 0 \). To remedy the situation in univariate case, Eitrheim and Teräsvirta (1996) suggested to replace \( \tilde{U} \) by its orthogonal part to the space spanned by \( \tilde{K} \), i.e. \( \tilde{V} = (I - \tilde{K}' \tilde{K}^{-1} \tilde{K}') \tilde{U} \). In the multivariate case, following Eitrheim and Teräsvirta (1996) \( \tilde{Z} \) leads to the following procedure:
1. Estimate the VLSTAR model under the null hypothesis of no serial correlation. Choose a lag length $J$ of the serial correlation in residuals. Regress the residuals $\tilde{U}$ on $\tilde{K}$ from $t = J + 1$ to $T$. Collect the residuals $\tilde{V}$ and compute the matrix residual sum of squares $RSS_0 = \tilde{V}'\tilde{V}$.

2. Run the auxiliary regression of $\tilde{V}$ on $(\tilde{K}, \tilde{Z})$ from $t = J + 1$ to $T$. Collect the residuals $\tilde{\Xi}$, and form the matrix residual sum of squares $RSS_1 = \tilde{\Xi}'\tilde{\Xi}$.

3. Compute the value of the test statistic (2.5.7).

2.5.2 Additive nonlinearity

We shall now consider the alternative hypothesis that after fitting a VLSTAR model, there is still nonlinearity left unmodelled. For simplicity, following Eitrheim and Teräsvirta (1996), it is specified as another logistic smooth transition component that enters the model additively. When the null hypothesis of no additive nonlinearity is rejected, there are at least two alternatives. First, one may accept the alternative and estimate a VSTAR model with two transitions. Since the reason of a rejection usually remains unknown, it is also possible to conclude that the model does not fit the data well and either respecify the whole model or switch to another family of models.

In order to derive the test statistic, consider the additive VLSTAR model

$$ y_t = B'_1x_t + G'_1tB'_{m+1}x_t + \varepsilon_t + G'_mB'_{m+1}x_t + \varepsilon_t $$

where $\varepsilon_t$ is i.i.d. normally distributed. We estimate the $m$-regime VLSTAR model in which the additive nonlinear term $G'_mB'_{m+1}x_t$ has been omitted and test $H_0 : \gamma_{mj} = 0, j = 1, \ldots, p$, against the alternative $H_1 : \text{at least one } \gamma_{mj} > 0$.

In order to carry out the tests, we have to form the set of the potential transition variables. If there is no theory available for doing that, the set used in testing linearity can be re-employed. Similarly, there are two ways of doing this: first, if economic theory does not suggest a single transition variable, we carry out the test of no additive nonlinearity equation by equation as in Teräsvirta (1998). But then, if the system should have a single transition variable, a joint test of no additive nonlinearity test is applicable. In the following, we develop such a test. As in Section 2.3, this joint test can also be applied to a subset of equations and it will be identical to a univariate LM-type test when $p = 1$.

Analogously to the situation in Section 2.3.1, the alternative model is not identified under the null hypothesis. We again employ the Taylor approximation of the transition function to deal with the problem. The model is reparameterised and approximated using either the first-order Taylor expansion around the null hypothesis, which gives

$$ y_t = \Psi'_1B'x_t + \Theta'_1x_t\gamma_t + \varepsilon'_t $$

(2.5.9)
or the third-order one, leading to

\[ y_t = \Psi'_t B'_t x_t + \Theta'_1 x_t s_t + \Theta'_2 x_t s_t^2 + \Theta'_3 x_t s_t^3 + \varepsilon^*_t \]

(2.5.10)

The error vector \( \varepsilon^*_t \) contains the remainder term of the Taylor expansion. The corresponding null hypotheses are:

\[ H_0 : \Theta_1 = 0 \]

(2.5.11)

for (2.5.9), and

\[ H_0 : \Theta_1 = \Theta_2 = \Theta_3 = 0 \]

(2.5.12)

for (2.5.10). Let again

\[
X = \begin{bmatrix}
x'_1 \\
x'_2 \\
\vdots \\
x'_T
\end{bmatrix},
Z_N = \begin{bmatrix}
x'_1 s_1 \\
x'_2 s_2 \\
\vdots \\
x'_T s_T
\end{bmatrix},
\]

Without loss of generality, we only consider testing the null hypothesis (2.5.11) in equation (2.5.9). The corresponding score evaluated under the null hypothesis is

\[
\frac{\partial \log L(\hat{\theta})}{\partial \Theta} = \sum_{t=1}^{T} \left[ x_t s_t \left( y_t - \hat{\Psi}'_t \hat{B}'_t x_t \right) \hat{\Omega}^{-1} \right]
\]

where \( \hat{\Psi} \), \( \hat{B} \), and \( \hat{\Omega} \) are estimates under the null hypothesis. Denote \( \tilde{\varepsilon}_t = y_t - \hat{\Psi}'_t \hat{B}'_t x_t \), and \( \hat{\varepsilon} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_T)' \). Following Luukkonen et al. (1988), the LM statistic for (2.5.9) or (2.5.10) can be written as

\[
LM = \text{tr}\{ \hat{\Omega}^{-1} \hat{\varepsilon}' Z_N (I_T - \hat{K}' \hat{K})^{-1} \hat{K}' Z_N^{-1} Z_N \hat{\varepsilon} \}
\]

(2.5.13)

where \( \hat{K} \) has been defined in (2.5.6) with \( J = 0 \) and \( N \) is the order of the Taylor expansion. Appendix D contains the details. The inequality \( T \geq ((pk + q)m + 2(m - 1))p^2 \) is a necessary condition for \( \hat{K}' \hat{K} \) to have full rank, and consequently, for the existence of the LM test. We have the following theorem:

**Theorem 2.5.2.** Consider the model in (2.5.8). Under the null hypothesis in (2.5.11) or (2.5.12), we have the LM test statistic.

\[
LM = \text{tr}\{ \hat{\Omega}^{-1} \hat{\varepsilon}' Z_N (I_T - \hat{K}' \hat{K})^{-1} \hat{K}' Z_N^{-1} Z_N \hat{\varepsilon} \}
\]

(2.5.14)

Assume that the parameter estimators are consistent. The test statistic then has an asymptotic \( \chi^2 \) distribution with degrees of freedom \( pN (kp + q) \) under the null hypothesis of no additive nonlinearity.

**Proof.** See Appendix B. \( \square \)
Note that if \( G_1^t = ... = G_{m-1}^t = 0 \) in (2.5.8), the test collapses into the linearity test discussed in Section 3.

Again, we have to mention that if the error terms are not normally distributed, there are moment conditions for the asymptotic distribution theory to be valid. In univariate case, a STAR model with logistic type transition functions must satisfy the condition \( E(\varepsilon_t^8) < \infty \). A sufficient condition in multivariate case can be \( E(\varepsilon_t^{8j}) < \infty \) in \( \varepsilon_t \), for \( j = 1, ..., p \).

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.3.1 by replacing \( X \) by \( \tilde{K} \). However, it also suffers from the empirical size distortion problem due to the non-orthogonality between \( \tilde{E} \) and \( \tilde{K} \). Thus we have the following algorithm:

1. Estimate the VLSTAR model under the null hypothesis of no additional nonlinearity. Regress the residuals \( \tilde{E} \) on \( \tilde{K} \). Collect the residuals \( \tilde{V} \), and compute the matrix residual sum of squares \( \text{RSS}_0 = \tilde{V}'\tilde{V} \).

2. Run the auxiliary regression of \( \tilde{V} \) on \( (\tilde{K}, \tilde{Z}_N) \). Collect the residuals \( \tilde{\Xi} \) and form the matrix residual sum of squares \( \text{RSS}_1 = \tilde{\Xi}'\tilde{\Xi} \).

3. Compute the test statistic (2.3.6).

### 2.5.3 Parameter constancy

One of the central assumptions of the linear VAR as well the VLSTAR models is that the parameters of the model are constant over time. Since the estimation of parameters is based on this assumption it must be tested, and this holds for both linear and nonlinear models. In this section we shall discuss testing parameter constancy of the VLSTAR model. As in the univariate case, considered in Eitrheim and Ter"asvirta (1996), a useful alternative is that the parameters change smoothly over time. In the alternative model the parameter change is characterised using a logistic function. In this case, we have a single transition variable for all equations, and it is \( \tau = t/T \). Consider the following VLSTAR model:

\[
y_t = \Psi'tB(\tau)'x_t + \varepsilon_t \tag{2.5.15}
\]

where \( B(\tau) = B_a + B_b \lambda(\tau|\gamma, c) \) is a function of rescaled (normalized) time \( \tau = t/T \). In this work, \( \lambda(\tau|\gamma, c) \) may take one of the three forms below

\[
\lambda(\tau|\gamma, c) = (1 + \exp(-\gamma(\tau - c)))^{-1} - 1/2 \tag{2.5.16}
\]

\[
\lambda(\tau|\gamma, c) = (1 + \exp(-\gamma(\tau - c_1)(\tau - c_2)))^{-1} - 1/2 \tag{2.5.17}
\]

\[
\lambda(\tau|\gamma, c) = \left(1 + \exp \left(-\gamma \prod_{j=1}^{3} (\tau - c_j) \right) \right)^{-1} - 1/2. \tag{2.5.18}
\]

The null hypothesis of parameter constancy is \( H_0 : \lambda(\tau|\gamma, c) = 0 \). Function (2.5.16) postulates a smooth monotonic parameters change and function (2.5.17) a nonmonotonic change that is symmetric around \( (c_1 + c_2)/2 \). Function (2.5.18) describes an even more flexible parameter
change which is generally nonmonotonic, but monotonic change appears as a special case when $c_1 = c_2 = c_3$.

When $\gamma \rightarrow \infty$ in (2.5.16), the alternative becomes a single structural break. When the same occurs in (2.5.17), one obtains a special case of a double break if $c_1 \neq c_2$, whereas (2.5.18) implies a triple break but only two regimes if $c_1 \neq c_2 \neq c_3$.

The equations (2.5.16)–(2.5.18) can be even more general, i.e. not necessarily the same for all equations. For notational simplicity, we only consider the function (2.5.16) here. The first order Taylor expansion of (2.5.16) becomes

$$\lambda(\tau|\gamma, c) = (\gamma/4)(\tau - c) + r_t$$  \hspace{1cm} (2.5.19)

where $r_t$ is the remainder. The model (2.5.15) is approximated and reparameterised as follows:

$$y_t = \Psi'_{i}B'a_{i}x_{t} + \Psi'_{i}B'b_{i}x_{t}\tau + \varepsilon^*_t$$  \hspace{1cm} (2.5.20)

where $\varepsilon^*_t$ also contains the remainder $r_t$ from the Taylor expansion (2.5.19). Note, however, that under $H_0$, $\varepsilon^*_t = \varepsilon_t$. The new null hypothesis is

$$H_0 : B_b = 0.$$  \hspace{1cm} (2.5.21)

The Lagrange multiplier test is derived from the score evaluated under the null hypothesis

$$\frac{\partial \log L(\theta)}{\partial B_b} = \sum_{t=1}^{T} x_t \tau \cdot \left( y_t - \hat{\Psi}'_{i}B'a_{i}x_{t} \right)' \Omega^{-1} \hat{\Psi}'_{i}$$

where $\hat{\Psi}$, $\hat{B}_a$, and $\hat{\Omega}$ are estimates under the null. Denote $\tilde{\varepsilon}_t = y_t - \hat{\Psi}'_{i}B'a_{i}x_{t}$, and $\tilde{E} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, ..., \tilde{\varepsilon}_T)'$. Let

$$\tilde{Z} = \begin{bmatrix} \text{vec}(\hat{\Psi}_1 \otimes x_1 \tau_1)'
\text{vec}(\hat{\Psi}_2 \otimes x_2 \tau_2)'
\vdots
\text{vec}(\hat{\Psi}_T \otimes x_T \tau_T)'
\end{bmatrix}, \quad \tilde{K} = \begin{bmatrix} \text{vec}[(\partial \hat{\Psi}'_{i}B'a_{i}x_{t}/\partial \theta)']
\text{vec}[(\partial \hat{\Psi}'_{i}B'b_{i}x_{t}/\partial \theta)']
\vdots
\text{vec}[(\partial \hat{\Psi}'_{i}B'a_{i}x_{t}/\partial \theta)']
\end{bmatrix}$$

where $\tau_i = i/T$, $i = 1, ..., T$. We have

**Theorem 2.5.3.** Consider the model in (2.5.15) where $\lambda(\tau|\gamma, c)$ takes the form (2.5.16). Under the null hypothesis $H_0 : \lambda(\tau) = 0$ or equivalently in (2.5.21), the LM-type test statistic has the form:

$$LM = \text{tr}\{\hat{\Omega}^{-1}\tilde{E}'\tilde{Z}[I_T - \hat{K}(\hat{K}'\hat{K})^{-1}\hat{K}']^{-1}\tilde{Z}'\hat{E}\}$$  \hspace{1cm} (2.5.22)

Assume that the estimators of the parameters in $\theta$ are consistent. The test statistic has an asymptotic $\chi^2$ distribution with degrees of freedom $mp^2(kp + q)$ under the null hypothesis of parameter constancy.

**Proof.** See Appendix B.  \hspace{1cm} \Box
Chapter 2. Linearity and misspecification tests

If the error terms are not normally distributed, the corresponding moment condition for the validity of the asymptotic distribution theory in univariate case is \( E(\varepsilon_t^6) < \infty \). A sufficient condition in multivariate case can be \( E(\varepsilon_{jt}^6) < \infty \) in \( \varepsilon_t \), for \( j = 1, \ldots, p \).

Similarly to the test in the previous section, \( T \geq [(pk + q)m + 2(m - 1)]p^2 \) is a necessary condition for \( \tilde{K}'\tilde{K} \) to have full rank, and consequently, for the existence of the LM test.

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.3.1 by replacing \( X \) by \( \tilde{K} \) and \( Z_1 \) by \( \tilde{Z} \). In order to alleviate the empirical size distortion problem due to the non-orthogonality between \( \tilde{E} \) and \( \tilde{K} \), we apply the following algorithm:

1. Estimate the VLSTAR model under the null hypothesis of constant parameters over time. Regress the residuals \( \tilde{E} \) on \( \tilde{K} \). Collect the residuals \( \tilde{V} \), and compute the matrix residual sum of squares \( RSS_0 = \tilde{V}'\tilde{V} \).

2. Run the auxiliary regression of \( \tilde{V} \) on \((\tilde{K}, \tilde{Z})\). Collect the residuals \( \tilde{\Xi} \), and the matrix residual sum of squares \( RSS_1 = \tilde{\Xi}'\tilde{\Xi} \).

3. Compute the test statistic (2.3.6).

This test can also be applied to subsets, which does not only mean subsets of equations or even single equations. It is often useful to focus on certain types of coefficients. For example, in a single equation it may be useful to test the constancy of the intercepts or other linear or nonlinear parameters separately, see Teräsvirta (1998) for discussion. This helps the modeller to locate possible weaknesses in the specification of the estimated model. This is particularly useful when the joint test rejects parameter constancy. Moreover, when the conditional mean (2.5.15) is linear, the test collapses into the corresponding parameter constancy test in a linear VAR model.

2.5.4 Curse of dimensionality

All of the three tests introduced in previous sections suffer from the “curse of dimensionality”. This is when the combination of the number of equations \( p \), the number of regimes \( m \), and the number of lags \( k \) is sufficiently large. The necessary conditions for the existence of these tests are: \( T - J \geq cd(\tilde{K}) \) for the test of no serial correlation test and \( T \geq cd(\tilde{K}) \) for the other two joint tests, where \( cd(\tilde{K}) \) is the column dimension of the matrix \( \tilde{K} \). The closer \( cd(\tilde{K}) \) is to \( T - J \), the more the size of the standard LM test will be distorted. A partial solution to this problem is to carry out the tests equation by equation. This way the modeller does not control the overall significance level of the test but is nevertheless able to gather information about the validity of the model.
2.6 Simulation study

2.6.1 P-value plot and p-value discrepancy plot

As already noted, the LM test statistic has advantage that estimation of the alternative model is avoided. This makes the test relatively easy to simulate. We begin by introducing the graphical methods we employ in reporting our results. By those we mean the techniques of Davidson and MacKinnon (1998). They suggested p-value plots and p-value discrepancy plots for the purpose. Consider a Monte Carlo experiment in which \( N \) realizations of some test statistic \( \tau \) are generated using a data-generating process (DGP) that is a special case of the null hypothesis (size experiments), or of the alternative (power experiments). Let \( p_j = p(\tau_j) \) denote the p-value evaluated using the \( j^{th} \) test statistic \( \tau_j \), \( j = 1, ..., N \), in the nominal distribution, which can be the asymptotic, or limiting, distribution of \( \tau \), or an approximation to the finite sample distribution of \( \tau \).

Both of the p-value plot and the p-value discrepancy plot are based on the empirical distribution function (EDF) of the p-values of the test statistic

\[
\hat{F}(x_i) = \frac{1}{N} \sum_{j=1}^{N} I(p_j \leq x_i)
\]  

(2.6.1)

where \( I(p_j \leq x_i) \) is an indicator function that takes the value 1 if its argument is true and 0 otherwise. The value \( x_i \) belongs to the \((0,1)\) interval. The EDF is a function of \( x_i \) given \( N \) realizations of \( \tau \). We construct a discrete grid \( \{x_i\}_{i=1}^{M} \) in advance so as to provide a reasonable snapshot of the \((0,1)\) interval. In this work we focus on the \((0,0.2)\) interval, because it covers the most commonly applied significance levels. We define

\[
\{x_i\}_{i=1}^{M} = \{0.001, 0.002, ..., 0.010, 0.015, ..., 0.195, 0.200\}
\]

where \( M = 68 \). This grid is not equidistant as the distance between two neighbouring elements for \( x_i < 0.01 \) is 0.001, otherwise it equals 0.005. The denser grid makes it less likely to miss any unusual behaviour in the left tail of the EDF.

The p-value plot is a scatter plot of \( \hat{F}(x_i) \) against \( x_i \), and the p-value discrepancy plot is a scatter plot of \( \hat{F}(x_i) - x_i \) against \( x_i \). If the actual distribution of the test statistic under the null hypothesis is very close to the nominal distribution, the p-value plot should be an approximately 45° straight line, and in the p-value discrepancy plot the points should stay close to zero, given a large number of realizations \( N \). The latter plot is suitable for reporting results of size experiments.

The p-value plot and the p-value discrepancy plot are continuous but nondifferentiable everywhere. For extreme cases in which the number of realizations \( N \) is not large, Davidson and MacKinnon (1998) suggested to smoothen the graphs. This will not be necessary here, as \( N = 5000 \) in our experiments.

We have to consider the experimental randomness caused by finite \( N \) in the p-value discrepancy plot. Davidson and MacKinnon (1998) employed the Kolmogorov-Smirnov (KS) test
statistic for the purpose. In practice this choice implies drawing a sample of realizations from the nominal distribution, and plotting two horizontal lines calculated using the formula above on the \( p \)-value discrepancy plot. The KS tends to be rather conservative, however, and as such may sometimes mislead us to think that there is no under- or over-rejection near \( x_i = 0 \).

Instead, we employ a two-sided asymptotic normal confidence band at the 95\% level which is derived from the assumption that the empirical and the nominal distributions of the test statistic are the same. Under this assumption, for each \( j = 1, \ldots, N \), the value of the indicator function \( I(p_j \leq x_i) \) is a realization of a Bernoulli distributed random variable with parameter \( x_i \). By applying the central limit theorem, for large number of realizations \( N \) the \( p \)-value discrepancy \( \hat{F}(x_i) - x_i \) can be approximated by a normal distribution with zero mean and variance \( N^{-1}x_i(1 - x_i) \).

### 2.6.2 Size experiments of linearity tests and misspecification tests of the evaluation

In this section, we shall investigate the finite sample size behaviour of our test statistics. We conduct Monte Carlo experiments in which 5000 realizations of a test statistics are generated using a DGP that is a special case of the null hypothesis which is a member of the family of linear VAR models. We shall focus on the size comparison among four different types of test statistics: the LM test in Theorem 2.3.1, the rescaled LM test in (2.4.1), Wilks's A-test in (2.4.5) and Rao's F-test in (2.4.6). The first two tests are included mainly to demonstrate the magnitude of size distortion, whereas the last two are serious contenders for use in applications. For simplicity we mainly focus on the first-order Taylor expansion based tests, but some results of the third-order variants will be presented as well.

**Linearity tests**

The basic DGP is a \( p \)-dimensional vector autoregressive time series \( y_t \) of lag order \( k \) without additional nonlinearity. The linear parameters are chosen to keep the dynamics asymptotically stationary. We test linearity using the transition variable \( s_t = t/T \) to check whether the coefficients of the VAR(k) model is time-varying. The true DGP is as follows:

\[
y_{t,t} = \sum_{j=1}^{k} \rho_j y_{t,t-j} + \varepsilon_{t,t}
\]

where \( \rho = 0.4 \) such that the vector model is stationary given a suitable set of initial values \((y_{t,1-k}, \ldots, y_{t,0})\), for any \( k = 1, 2, \ldots \). In our experiments, since there is no intercept in the true GDP, we just set the initial values to be all zeros. Different combinations of the values below will be considered:

\[
p \in \{2, 5, 10\}, \quad k \in \{1, 2, 5\}, \quad T \in \{30, 50, 100\}.
\]

We investigate three cases: the bivariate case \((p = 2)\), high-dimension \((p = 5)\) and extremely high-dimension case \((p = 10)\). The lag length \( k \) also matters, because it determines the column
dimension of the matrix $Z$, or in other words, the degrees of freedom of the nominal distribution. We choose a small sample size $T$, because we are interested in the behaviour of the empirical size of the tests in small samples.

Recall the necessary condition $T \geq m(kp + q)$ for the existence of the estimators discussed in Section 5 of Chapter 1. The following designs will be studied and the results of them will be plotted.

- **Design 1**: $p = 2$, $k = 1$ and $T = 30$;
- **Design 2**: $p = 2$, $k = 1$ and $T = 100$;
- **Design 3**: $p = 2$, $k = 2$ and $T = 30$;
- **Design 4**: $p = 2$, $k = 5$ and $T = 30$;
- **Design 5**: $p = 5$, $k = 1$ and $T = 50$;
- **Design 6**: $p = 10$, $k = 1$ and $T = 50$.

The $p$-value discrepancy plots for Designs 1 and 2 are given in Figures 2.1 and 2.2 respectively. Both designs are bivariate, and $k = 1$. The only difference between them is the sample size $T$.

As for Design 1, Figure 2.1 shows that both the LM test and the rescaled test are size-distorted. The former one over-rejects, whereas the latter under-rejects. This agrees with previous results. The empirical size of Wilks’s $\Lambda$ test is very close to that of Rao’s $F$-test, and neither is size-distorted. It seems that these two tests work well even for the smaller sample size $T = 30$.

Given this result, it is not surprising that when $T = 100$, Wilks’s $\Lambda$ and Rao’s $F$-test have a very similar performance. It is seen that the LM test now works well, which accords with the theoretical result that the LM statistic converges in distribution to the $\chi^2$ distribution. The rescaled test still under-rejects. This suggests that the Monte Carlo results of Bera et al. (1981) for univariate models also hold in multivariate models.

The $p$-value discrepancy plots for Designs 3 and 4 can be found in Figures 2.3 and 2.4, respectively. Both designs are bivariate and $T = 30$. The lag lengths are different, two and five.

In comparing Designs 3 and 4 we notice that the lag length has a strong impact on results. While the rescaled test and the LM test behave badly in both cases, Rao’s $F$-test is the only one that still has no size distortion when $k = 5$. Wilks’s $\Lambda$, a good performer in the first three experiments, is now rather strongly oversized.

In designs 5 and 6 the main object of interest is the dimension of the vector system $p$. In Design 5 $p = 5$, whereas $p = 10$ in Design 6, other things equal. Again, while the behaviour of Wilks’s $\Lambda$ is acceptable for the shorter lag length, the test over-rejects when $p = 10$. The empirical size of Rao’s $F$-test is practically unaffected by the change in the lag length from 2 to 10. Our conclusion is that among the tests inspected we should always choose Rao’s $F$-test when testing linearity using the tests described in Section 3. This accords with the results in Edgerton and Shukur (1999) who considered testing autocorrelation in a vector system.
Misspecification tests

The basic DGP is a $p$-dimensional vector autoregressive time series $y_t$ of lag order $k$ with one additive logistic nonlinear component, i.e., VLSTAR. The linear parameters are chosen to exclude unstable processes. The true transition variable $s_{it} = y_{i,t-1}$, $i = 1, ..., p$. So each equation has its own transition variable. We choose $c = 0$ to be the location parameter vector, and set $\gamma = 1$. The data generating process is as follows:

$$
 y_{i,t} = \left( \sum_{j=1}^{k} \rho_1^{i} y_{i,t-j} \right) (1 - g(s_{it})) + \left( \sum_{j=1}^{k} \rho_2^{i} y_{i,t-j} \right) g(s_{it}) + \varepsilon_{i,t} 
$$

(2.6.4)

where $\rho_1 = 0.4$, $\rho_2 = 0.2$. This model is ergodic, as all the roots of $\sum_{j=1}^{k} (\rho_1 (1 - g(s_{it})) + \rho_2 g(s_{it})) L^j = 0$ are outside the unit disk, for all $s_{it}$ and any $k = 1, 2, ..., $ see for example, Granger and Teräsvirta (1993, Chapter 2).

We generate 5000 realizations from our DGP. For each realization, we estimate the VLSTAR model (2.6.4). We compute the residual vector for each realization, and following the three algorithms in Section 2.5, we obtain the $p$-values of the three misspecification tests. The test statistics are the same as before: the LM test, the rescaled test, Wilks’s $\Lambda$ test and Rao’s $F$-test.

We consider the size distortion of the four tests. The performance of the misspecification tests depends on how accurate the parameter estimates are. In order to achieve reasonable accuracy, we choose $T = 200$ and $T = 500$.

$P$-value discrepancy plots of the test of no serial correlation is given in Figure 2.7. As can be expected, the LM test over-rejects, whereas the rescaled test under-rejects. The empirical size of Wilks’s $\Lambda$ test is very close to that of Rao’s $F$-test, and neither test is size-distorted. With the increase of sample size from $T = 200$ to $T = 500$, the performance of the LM test and the rescaled test improves. However, the improvement is not very large. The rescaled test seems to outperform the LM test at both sample sizes in the sense that the absolute size distortion is smaller in the former than in the latter.

Figure 2.8 shows the $p$-value discrepancy plots for test of no additive nonlinearity. Both Wilks’s $\Lambda$ test and Rao’s $F$-test work have a satisfactory empirical size. The rescaled test performs better than the LM test and shows no size distortion for $T = 500$.

Figure 2.9 shows the $p$-value discrepancy plots for test of parameter constancy. For $T = 200$, Wilks’s $\Lambda$ test is slightly oversized, while Rao’s $F$-test shows hardly any size distortion. The rescaled test and the LM test perform even worse.

Recall that the computation of the three misspecification tests requires construction of the matrices $Z$ and $K$. The column dimensions of the matrices $Z$ and $K$ affect the empirical size of the corresponding test statistic. In the three tests, $K$ has the same column dimension, whereas those of $Z$ are different. Choosing a large lag length $J$ for testing serial correlation slows down the convergence of the standard LM test statistic to its limiting distribution. The column dimension of the matrix $Z$ of the parameter constancy test is the squared size of a Kronecker product, and the test thus has the worst performance. For certain sample size $T$, the LM test and
the rescaled test of the three misspecification tests performs differently. This can be explained by different column dimension of the matrix $Z$. However, Rao’s F-test and the Wilks’s Λ test are still performing well.

2.7 Concluding remarks

In this chapter, we propose Lagrange-multiplier type linearity and misspecification tests in the VLSTR framework. We allow the dimension of the model exceed two and, furthermore, do not restrict the number of transitions to one. We consider both the case in which the VLSTR model only has a single transition variable, and the case in which every equation has its own (known) transition variable.

We consider three misspecification tests for possible model extensions: the test of no serial correlation, the test of no additive nonlinearity and the parameter constancy test. They are either Lagrange multiplier or Lagrange multiplier type tests. We generalize the univariate misspecification tests in Eitrheim and Teräsvirta (1996) to multivariate joint tests.

Small-sample properties of the tests must be carefully studied because they may be affected by the dimension of the model. We report the results of the simulation studies in which the size and power of the proposed tests are investigated in high-dimensional systems. We find that the standard LM tests are severely size-distorted when the dimension of the system increases. Wilks’ Λ statistic and Rao’s F statistic that have satisfying size properties are recommended for empirical use.
Appendix A: Proof of Theorem 2.3.1

Proof. Given Assumption 2.2.1–2.2.3, The Lagrange multiplier test under the null is derived from the score matrix

\[
\frac{\partial \log L(\tilde{\theta})}{\partial \Theta_1} = \sum_{t=1}^{T} \{x_t s_t \left( y_t - \tilde{\Theta}_0 x_t \right)' \tilde{\Omega}^{-1} \} = Z' \left( Y - XB_1 \right) \tilde{\Omega}^{-1} \tag{2.7.1}
\]

where

\[
Y = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix}, \quad Z = \begin{bmatrix} x_1's_1' \\ x_2's_2' \\ \vdots \\ x_T's_T' \end{bmatrix}
\]

and \( \tilde{\theta}, \tilde{\Theta}_0 = \tilde{B}_1 \) and \( \tilde{\Omega} \) are estimates under the null hypothesis \( H_0 \). The score converges in probability to a matrix-normal distribution with zero mean and variance \( Z' (I - P_x) Z \otimes \tilde{\Omega}^{-1} \) conditional on \( X \) and \( Z \), where \( P_x \equiv X(X'X)^{-1}X' \) is the projection matrix.

To see this, we write (2.7.1) as follows

\[
Q = \frac{\partial \log L(\tilde{\theta})}{\partial \Theta_1} = Z' \left( Y - XB_1 \right) \tilde{\Omega}^{-1} = Z'(Y - X(XX)^{-1}X'Y)\tilde{\Omega}^{-1} = Z'(I - P_x)(XB_1 + E)\tilde{\Omega}^{-1} = Z'(I - P_x)E\tilde{\Omega}^{-1}.
\]

Under the null hypothesis, \( Y = XB_1 + E \), where \( E = (\varepsilon_1, ..., \varepsilon_T)' \) and \( \text{vec}(E) \) follows a \( \mathcal{N}(0, I_T \otimes \Omega) \) distribution. Under the null hypothesis, \( \tilde{\Omega} \) will converge to \( \Omega \) in probability. Set

\[
S = (Z'(I - P_x)Z)^{-\frac{1}{2}} Q \tilde{\Omega}^\frac{1}{2}
\]

which will asymptotically converge to a matrix-normal distribution with zero mean and variance \( I \otimes I \). Thus we have the chi-square version LM test statistic

\[
LM = \text{tr}(S'S) = \text{tr}(\tilde{\Omega}^{-1}(Y - XB_1)'Z \left[ (I_T - P_x)Z \right]^{-1} Z'(Y - XB_1))
\]

which converges in probability to the \( \chi^2(p(kp + q)) \) distribution when the null hypothesis is valid.

Appendix B: LM test statistic against an additive component

Now consider the \( p \)-dimensional system of equations with an additive component to be tested:

\[
y_t = f(x_t, s_t|\Theta_0) + A'_t\tilde{\Theta}_1 z_t + \varepsilon_t \tag{2.7.2}
\]
Chapter 2. Linearity and misspecification tests

where $f$ is a vector of linear or nonlinear functions, in which $x_t$ and $s_t$ are vectors of independent variables, $s_t$ may be referred to as the vector of transition variables, $A_t$ is a time-varying matrix which contains some nuisance parameters, and $z_t$ is a vector of independent variables in the additive component. In many cases, $z_t = x_t$, see, for example, the joint test of linearity against the VLSTAR alternative in Section 2.3, but $z_t \neq x_t$ is allowed as well. The test of no error serial correlation test in Section 2.5.1 serves as an example.

The corresponding set of parameters in the model (2.7.2) is $\theta = \{\Theta_0, \Theta_1, \Omega\}$ where $\Theta_0$ is a parameter matrix in $f$, $\Theta_1$ is a parameter matrix in the additive component, and the positive definite matrix $\Omega$ is the covariance matrix of the errors. The existence of the additive component $A_t'\Theta_1'z_t$ is going to be tested, and then the null hypothesis is $H_0: \Theta_1 = 0$.

The case when $A_t = I_p$

Consider the special case when $A_t = I_p$, for example, the error serial correlation test in Section 2.5.1, the joint test of no additive nonlinearity in Section 2.5.2, and the joint test of linearity against the VLSTAR alternative with a single transition variable in Section 2.3. The corresponding block of the score matrix takes the form

$$\frac{\partial \log L(\tilde{\theta})}{\partial \Theta_1} = \sum_{t=1}^{T} \left\{ z_t \tilde{\varepsilon}_t' \tilde{\Omega}^{-1} \right\} = Z' \tilde{\varepsilon} \tilde{\Omega}^{-1}$$

where $\tilde{\varepsilon}_t = y_t - f(x_t, s_t|\tilde{\Theta}_0)$, $Z = (z_1, z_2, ..., z_T)'$ and $\tilde{E} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, ..., \tilde{\varepsilon}_T)'$. The tilde means estimates under the null hypothesis.

The vectorized LM test statistic is

$$LM = \text{vec}(\tilde{E}'Z)' \left( (Z'(I - P_K)Z) \otimes \tilde{\Omega} \right)^{-1} \text{vec}(\tilde{E}Z)'$$

where $P_K = K(K'K)^{-1}K'$ and

$$K = \begin{bmatrix} \text{vec} \left( \frac{\partial f(x_1, s_1|\tilde{\Theta}_0)}{\partial \Theta_0} \right)' \\ \text{vec} \left( \frac{\partial f(x_2, s_2|\tilde{\Theta}_0)}{\partial \Theta_0} \right)' \\ \vdots \\ \text{vec} \left( \frac{\partial f(x_T, s_T|\tilde{\Theta}_0)}{\partial \Theta_0} \right)' \end{bmatrix}$$

This is the general expression of LM test statistic in Luukkonen et al. (1988) written using the notation in Lütkepohl (2004, Chapter 4). Under Assumptions A. 1–3, the vectorized score matrix is asymptotically normally distributed with $p \text{cd}(Z)$ degrees of freedom, i.e. the number of elements in $\text{vec}(\tilde{E}'Z)$, or the number of restrictions. See for example Breusch and Pagan (1980).
Chapter 2. Linearity and misspecification tests

The statistic (2.7.4) can be written as follows:

\[
LM = \text{vec}(\tilde{E}'Z)' \left( (Z'(I - P_K)Z) \otimes \tilde{\Omega} \right)^{-1} \text{vec}(\tilde{E}'Z)
\]
\[
= \text{vec}(\tilde{E}'Z)' \left( (Z'(I - P_K)Z)^{-1} \right) \text{vec}(\tilde{E}'Z)
\]
\[
= \text{vec}(\tilde{E}'Z)' \text{vec} \left( \tilde{\Omega}^{-1} \tilde{E}'Z (Z'(I - P_K)Z)^{-1} \right)
\]
\[
= \text{tr} \left\{ Z'\tilde{E}\tilde{\Omega}^{-1} \tilde{E}'Z (Z'(I - P_K)Z)^{-1} \right\}
\]
\[
= \text{tr} \left\{ \tilde{\Omega}^{-1} \tilde{E}'Z (Z'(I - P_K)Z)^{-1} Z'\tilde{E} \right\}
\]  
(2.7.6)

Note that (2.7.6) avoids vectorization and Kronecker products. Furthermore, the value of (2.7.6) can be obtained by applying the following auxiliary regression:

1. Estimate the restricted model under the null hypothesis. Collect the residuals \( \tilde{e}_t = y_t - f(x_t, s_t; \hat{\Theta}_0) \) and form \( \tilde{E} \). Compute the matrix residual sum of squares \( \text{RSS}_0 = \tilde{E}'\tilde{E} \).

2. Run an auxiliary regression of \( \tilde{E} \) on \( (K, Z) \). Collect the residuals \( \tilde{\Xi} \), and form the matrix residual sum of squares \( \text{RSS}_1 = \tilde{\Xi}'\tilde{\Xi} \).

3. Compute the test statistic

\[
LM = T \text{tr} \{ \text{RSS}_0^{-1} (\text{RSS}_0 - \text{RSS}_1) \}
\]
\[
= T(p - \text{tr} \{ \text{RSS}_0^{-1} \text{RSS}_1 \})
\]  
(2.7.7)

The case when \( A_t \neq I_p \)

Consider the case when \( A_t \neq I_p \), for example the joint test of linearity against the VLSTAR alternative with different transition variables in Section 2.3 in which \( A_t = S_t \), and the test of parameter constancy in Section 2.5.3 in which \( A_t = \tilde{\Psi}_t \). The corresponding block of the score matrix takes the form

\[
\frac{\partial \log L(\tilde{\theta})}{\partial \Theta_1} = \sum_{t=1}^T \left\{ z_t \tilde{e}_t' \tilde{\Xi}_t^{-1} A_t' \right\}
\]  
(2.7.8)

where \( \tilde{e}_t \) has been defined in (2.7.3).

The LM statistic (2.7.4) is still valid for testing \( \Theta_1 = 0 \), when

\[
Z = (\text{vec}(A_1 \otimes z_1), \text{vec}(A_2 \otimes z_2), ..., \text{vec}(A_T \otimes z_T))'.
\]  
(2.7.9)

To see this, write

\[
A_t'\Theta_1'z_t = \text{vec}(z_t'\Theta_1 A_t) = (A_t \otimes z_t)'\text{vec}(\Theta_1) = \text{vec} \left( \text{vec}(\Theta_1)'(A_t \otimes z_t) \right)
\]
\[
= (I_p \otimes \text{vec}(\Theta_1))' \text{vec} (A_t \otimes z_t) = \tilde{\Theta}_1'\tilde{z}_t
\]  
(2.7.10)

where \( \tilde{\Theta}_1 = I_p \otimes \text{vec}(\Theta_1) \) and \( \tilde{z}_t = \text{vec} (A_t \otimes z_t) \). Note that \( \tilde{\Theta}_1 \) contains the same elements as \( \Theta_1 \), the remaining ones being equal to zero. Therefore, the corresponding number of degrees of freedom should be equal to the number of nonzero parameters in \( \tilde{\Theta}_1 \), that is, the number of parameters in \( \Theta_1 \), as only these parameters can vary freely.
Chapter 2. Linearity and misspecification tests

The null hypothesis can be rewritten as $H_0 : \hat{\Theta}_1 = 0$. The corresponding block of the score matrix is (2.7.3), with $Z$ is defined as in (2.7.9).

Suppose that $A_t$ is an $a \times p$ matrix, $\hat{\Theta}_1$ is a $b \times a$ matrix and $z_t$ is a $b \times 1$ vector. Moreover, write $A_t = (a_{1t}, ..., a_{pt})$, where $a_{jt}, j = 1, ..., p$, is an $a \times 1$ vector. We have

$$
\hat{\Theta}_1^T z_t = \begin{bmatrix}
\text{vec}(\Theta_1)' (a_{1} \otimes z_t)
\text{vec}(\Theta_1)' (a_{2} \otimes z_t)
\vdots
\text{vec}(\Theta_1)' (a_{p} \otimes z_t)
\end{bmatrix}_{p \times 1}
$$

and define $Z_j = \begin{bmatrix}
(a_{j1} \otimes z_1)'
(a_{j2} \otimes z_2)'
\vdots
(a_{jt} \otimes z_t)'
\end{bmatrix}$. \hfill (2.7.11)

If $A_t = S_t$, where $S_t$ is a diagonal matrix of $p$ transition variables $s_{1t}, ..., s_{pt}$, the statistic (2.7.4) is used for testing linearity against the VLSTAR model with these transition variables ($s_{jt}$ for the $j^{th}$ equation). A special case of this is $s_{1t} = ... = s_{pt}$ or $A_t = s_t I_p$, i.e., system has a single transition variable. Then (2.7.11) simplifies to

$$
\hat{\Theta}_1^T z_t = \begin{bmatrix}
\theta'_1 \cdot (z_t s_{1t})
\theta'_2 \cdot (z_t s_{2t})
\vdots
\theta'_p \cdot (z_t s_{pt})
\end{bmatrix}_{p \times 1}
$$

and define $Z_j = \begin{bmatrix}
(z_1 s_{j1})'
(z_2 s_{j2})'
\vdots
(z_t s_{jt})'
\end{bmatrix}$. \hfill (2.7.12)

where $\theta_j, j = 1, ..., p$, is a $b \times 1$ column vector.

We can still use the auxiliary regression to compute the value of (2.7.6). However, from (2.7.11), we see that the auxiliary regression should be carried out equation by equation. That is,

1. Estimate the restricted model under the null hypothesis. Collect the residuals $\hat{\varepsilon}_t = y_t - f(x_t, s_t|\hat{\Theta}_0)$ and form $\hat{E}$. Compute the matrix residual sum of squares $\text{RSS}_0 = \hat{E}'\hat{E}$.

2. Run an auxiliary regression for each column of $\hat{E}$, that is, regress column $j$ of $\hat{E}$, $j = 1, ..., p$, on $(K, Z_j)$ where $Z_j$ defined in (2.7.11) or in (2.7.12) if $A_t = S_t$. Collect the residuals $\hat{\eta}$, and form the matrix residual sum of squares $\text{RSS}_1 = \hat{\eta}'\hat{\eta}$.

3. Compute the test statistic in (2.7.7).

Appendix C: Proof of Theorem 2.4.1

Proof. The score matrix evaluated under the null hypothesis has the general form

$$
\frac{\partial \log L(\hat{\theta})}{\partial \Theta_1} = Z' (Y - XB_1) \hat{\Omega}^{-1}.
$$

Use of the auxiliary regression approach for computing the test statistic produces two residual sums of squares, $\text{RSS}_0$ and $\text{RSS}_1$.

The first one, $\text{RSS}_0$, is the residual sum of squares matrix from the restricted regression, i.e., $\text{RSS}_0 = \hat{E}'\hat{E}$, $\hat{E} = (I - P_z) Y$, where $P_z$ is the projection matrix of $X$. Notice that under
the null hypothesis, \( Y = XB_1 + E \), where \( E = (\varepsilon_1, ..., \varepsilon_T)' \) and vec(\( E' \)) follows a \( \mathcal{N}(0, I_T \otimes \Omega) \) distribution. So we have \( \tilde{E} = (I - P_x)Y = (I - P_x)E \).

The \( \text{RSS}_1 \) is the residual sum of squares matrix from the auxiliary regression, i.e., \( \text{RSS}_1 = \tilde{\Xi}'\tilde{\Xi}, \tilde{\Xi} = (I - P_{xz})\tilde{E} \), where \( P_{xz} \) is the projection matrix of the matrix \( [ X, Z ] \), i.e.,

\[
P_{xz} = \begin{bmatrix} X & Z \end{bmatrix} \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X' \\ Z' \end{bmatrix}.
\]

Let the \( p \times p \) matrix \( W_1 = \text{RSS}_0 - \text{RSS}_1 \) and the \( p \times p \) matrix \( W_2 = \text{RSS}_1 \).

For \( W_1 \), it follows that

\[
W_1 = \text{RSS}_0 - \text{RSS}_1 = \tilde{E}'\tilde{E} - \tilde{\Xi}'\tilde{\Xi} = \tilde{E}'(I - P_x)\tilde{E} = \tilde{E}'(I - P_x)(I - P_x)\tilde{E} = \tilde{E}'(I - P_x)\tilde{E}.
\]

Let \( I_T - P_x = RR' \), where \( R \perp X \) and \( R'R = I_{T - \text{cd}(X)} \). Then

\[
W_1 = E'RR'(Z'RR'Z)^{-1}Z'RR'E.
\]

Set \( V_1 = Z'RR'E \). So \( V_1 \sim \mathcal{N}(0, Z'RR'Z \otimes \Omega) \). It is seen that \( W_1 \) follows a Wishart distribution generated by \( V_1 \):

\[
W_1 = V_1'(Z'RR'Z)^{-1}V_1 \sim W_p(\Omega, \text{cd}(Z))
\]

For \( W_2 \), it follows that

\[
W_2 = \text{RSS}_1 = \tilde{\Xi}'\tilde{\Xi} = \tilde{E}'(I - P_{xz})\tilde{E} = \tilde{E}'\tilde{E} - \tilde{E}'P_{xz}\tilde{E} = \tilde{E}'(I - P_x)(I - P_x)\tilde{E} = \tilde{E}'R'(I - R)(Z'RR'Z)^{-1}Z'R'E = E'RR'(I - R)(Z'RR'Z)^{-1}Z'R'E.
\]

We have \( I_{T - \text{cd}(X)} - R'(Z'RR'Z)^{-1}Z'R' = QQ' \), where \( Q \perp R'Z \) and \( Q'Q = I_{T - \text{cd}(X) - \text{cd}(Z)} \).

Using this, \( W_2 = E'RQQ'R'E \). Set \( V_2 = Q'R'E \). So we have \( V_2 \sim \mathcal{N}(0, I \otimes \Omega) \). It is seen that \( W_2 \) follows a Wishart distribution generated by \( V_2 \):

\[
W_2 = V_2'V_2 \sim W_p(\Omega, T - \text{cd}(X) - \text{cd}(Z))
\]

Stacking the columns of \( V_1 \) and \( V_2 \) yields the random matrix

\[
U = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Z'R \\ Q' \end{pmatrix} R'E
\]

87
It follows that $U \sim N(0, \Sigma \otimes \Omega)$, where the row covariance matrix

$$\Sigma = \begin{pmatrix} Z'R \\ Q' \end{pmatrix} R'R (R'Z, Q) = \begin{pmatrix} Z'R'R'Z & Z'RQ \\ Q'R'Z & Q'Q \end{pmatrix} = \begin{pmatrix} Z'R'R'Z & 0 \\ 0 & I \end{pmatrix}$$

because $Q \perp R'Z$. We conclude that $V_1$ and $V_2$ are uncorrelated, and independent due to normality. It follows that $W_1$ and $W_2$ are independent as desired. \qed

Appendix D: The first-order partial derivatives of $\Psi' B' x_t$

The vectorized first order derivative of $\Psi' B' x_t$ w.r.t. parameters $\theta$ can be easily found in both univariate and multivariate cases, see Eitrheim and Ter"asvirta (1996). The set of parameters $\theta$ consists of $B$, $\Omega$, $\Gamma$ and $C$, where $B = [b_{ij}]$, $\Gamma = [\gamma_{ij}]$ and $C = [c_{ij}]$.

For parameter $B = [b_{ij}]$, we have

$$\frac{\partial \Psi' B' x_t}{\partial b_{ij}} = \Psi' H'_{ij} x_t \tag{2.7.14}$$

where $H_{ij} = [h_{kl}]$ is a matrix in which $h_{ij} = 1$ and $h_{kl} = 0$ for $k \neq i$ and $l \neq j$. Vector (2.7.14) is the directional derivative of the vector $\Psi' B' x_t$ with respect to the unit length matrix $H_{ij}$.

For the parameter matrices $\Gamma = [\gamma_{ij}]$ and $C = [c_{ij}]$, letting $\delta_{ij} = \gamma_{ij}$, $c_{ij}$, we have

$$\frac{\partial \Psi' B' x_t}{\partial \delta_{ij}} = \begin{pmatrix} 0_p, \ldots, \frac{\partial G_i^j}{\partial \delta_{ij}}, \ldots, 0_p \end{pmatrix} B' x_t = \frac{\partial G_i^j}{\partial \delta_{ij}} B'_{i+1} x_t \tag{2.7.15}$$

for $i = 1, \ldots, m - 1$, where

$$\frac{\partial G_i^j}{\partial \delta_{ij}} = \text{diag} \left\{ 0, \ldots, \frac{\partial g_{ij}}{\partial \delta_{ij}}, \ldots, 0 \right\} \tag{2.7.16}$$

for $j = 1, \ldots, p$. When $\delta_{ij} = \gamma_{ij}$,

$$\frac{\partial g_{ij}}{\partial \gamma_{ij}} = (g_{ij})^2 \exp\{-\gamma_{ij}(s_t - c_{ij})\} (s_t - c_{ij}) = (s_t - c_{ij}) g^j_t (1 - g^j_t), \tag{2.7.17}$$

and when $\delta_{ij} = c_{ij}$,

$$\frac{\partial g_{ij}}{\partial c_{ij}} = -(g_{ij})^2 \exp\{-\gamma_{ij}(s_t - c_{ij})\} \gamma_{ij} = -\gamma_{ij} g^j_t (1 - g^j_t). \tag{2.7.18}$$

Finally,

$$\frac{\partial \Psi' B' x_t}{\partial \Omega} = 0 \tag{2.7.19}$$

The dimension of the first-order derivative of $\Psi' B' x_t$ with respect to $\theta$ is $p \times [(kp + q)mp + 2(m-1)p]$. 

88
Figure 2.1: The size discrepancy plot for Design 1: $p = 2$, $k = 1$ and $T = 30$. Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.2: The size discrepancy plot for Design 2: $p = 2$, $k = 1$ and $T = 100$. Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.3: The size discrepancy plot for Design 3: $p = 2$, $k = 2$ and $T = 30$. Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.4: The size discrepancy plot for Design 4: $p = 2$, $k = 5$ and $T = 30$. Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.5: The size discrepancy plot for Design 5: $p = 5$, $k = 1$ and $T = 50$. Solid line: the Rao's test; dashed line: the Wilks's $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.6: The size discrepancy plot for Design 6: $p = 10$, $k = 1$ and $T = 50$. Solid line: the Rao's test; dashed line: the Wilks's $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.7: Size discrepancy plot for tests of no serial correlation: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.8: Size discrepancy plot for tests of no additive nonlinearity: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.9: Size discrepancy plot for tests of parameter constancy: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). Solid line: the Rao’s test; dashed line: the Wilks’s $\Lambda$ test; dotted line: the LM test; dashed-dotted line: the rescaled test. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Chapter 2. Linearity and misspecification tests

Figure 2.10: The size discrepancy plot for the LM test: \( p = 2, k = 1 \) and \( \rho = 0 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.11: The size discrepancy plot for the rescaled F test: \( p = 2, k = 1 \) and \( \rho = 0 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.12: The size discrepancy plot for Wilks’s test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.13: The size discrepancy plot for Rao’s test: $p = 2$, $k = 1$ and $\rho = 0$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.14: The size discrepancy plot for the LM test: $p = 2$, $k = 1$ and $\rho = 0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.15: The size discrepancy plot for the rescaled F test: $p = 2$, $k = 1$ and $\rho = 0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.16: The size discrepancy plot for Wilks’s test: $p = 2$, $k = 1$ and $\rho = 0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.17: The size discrepancy plot for Rao’s test: $p = 2$, $k = 1$ and $\rho = 0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.18: The size discrepancy plot for the LM test: $p = 2$, $k = 1$ and $\rho = -0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.19: The size discrepancy plot for the rescaled F test: $p = 2$, $k = 1$ and $\rho = -0.9$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Figure 2.20: The size discrepancy plot for Wilks’s test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Figure 2.21: The size discrepancy plot for Rao’s test: \( p = 2, k = 1 \) and \( \rho = -0.9 \). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).
Chapter 3

Testing Constancy of the Error Covariance Matrix using a Spectral Decomposition and a Parametric Alternative in Vector Models
Testing Constancy of the Error Covariance Matrix
using a Spectral Decomposition and a Parametric
Alternative in Vector Models

abstract

I consider multivariate (vector) time series models in which the error covariance matrix may be time-varying. I derive a test of constancy of the error covariance matrix against the alternative that the covariance matrix changes over time. I design a new family of Lagrange-multiplier tests against the alternative hypothesis that the innovations are time-varying according to several parametric specifications. I investigate the size and power properties of these tests and find them satisfactory.

3.1 Introduction

In univariate time series analysis, testing the adequacy of the estimated model has long been standard practice, see Box and Jenkins (1970). In vector models, most of the tests available in the literature, for checking the specification, have been designed to test the conditional mean. However, the error covariance matrix can also be nonconstant over time, even when the conditional mean is correctly specified, see Lütkepohl (2004) for a detailed discussion. Test exists for testing the constancy of the error variance in univariate models, whereas less has been done in multivariate models.

The most recent work for testing the constancy of the error covariance matrix can be found in Eklund and Teräsvirta (2007). They derived a family of test statistics against various types of misspecification making the use of the constant conditional correlation framework of Bollerslev (1990). One of the motivations of this chapter is to develop a new multivariate heteroskedasticity test as an alternative to the one proposed in Eklund and Teräsvirta (2007).

My test is based on a different decomposition of the error covariance matrix, namely the spectral decomposition. I develop a new family of tests which allow for various types of misspecification under this decomposition. The idea with the decomposition is to obtain tests against parsimoniously parameterised alternatives such that the test would still be powerful against many kinds of departure from parameter constancy. The spectral decomposition assumption has the advantage that it considerably decrease the dimension of the vector model compared to the case where all the diagonal and lower-diagonal elements of the covariance matrix can vary freely under the alternative hypothesis.

We will see that this test is extremely useful in the case when the practitioner is dealing with a structural vector model, whereas the constant conditional correlation assumption does not fit well in this case. Also it offers a new family of different specifictions for modelling the time-varying covariance matrix.
Following Eklund and Teräsvirta (2007), three types of alternatives to constancy are considered in this chapter. The first one may be viewed as a multivariate generalization of the heteroskedasticity test of White (1980), and the second one generalizes the test against autoregressive conditional heteroskedasticity of Engle (1982). The third variant of the test generalizes the univariate constant variance test of Medeiros and Veiga (2003), in which it is assumed that under the alternative hypothesis the variance changes smoothly over time. It can be seen that the joint constancy test for error covariance matrix against multivariate heteroskedasticity is very easy to implement and use. It has satisfactory size and power properties even in high-dimensional vector models.

The plan of the paper is as follows. The statistical model is introduced in Section 3.2. I make two assumptions which are employed in this chapter. The tests are developed based on these two assumptions. The Lagrange-multiplier type test statistic is derived in Section 3.3. In Section 3.4, I discuss different kinds of specification. The finite sample properties of the tests are investigated in Section 3.5. Section 3.6 concludes.

3.2 The model

Consider the following multivariate (vector) model:

\[ y_t = f(x_t) + u_t \]  \hspace{1cm} (3.2.1)

where \( y_t = (y_{1t}, ..., y_{pt})' \) is a \( p \times 1 \) vector of observable variables, \( u_t \) are serially uncorrelated errors with mean zero and the covariance matrix \( \Sigma_t \) which may be time-varying, \( f \) is a vector of functions, and \( x_t \) is a vector of variables which may contain lags of the dependent variable \( y_t \), the intercept, deterministic dummy variables and exogenous variables. The model (3.2.1) may be nonlinear.

The covariance \( \Sigma_t \) is a symmetric positive definite matrix, conditional on all the information available at time \( t \). I make the following assumption:

**Assumption 3.2.1.** (Spectral decomposition) The time-varying conditional covariance matrix \( \Sigma_t \) can be decomposed as follows:

\[ \Sigma_t = P \Lambda_t P' \]  \hspace{1cm} (3.2.2)

where the time-invariant matrix \( P \) satisfies \( PP' = I_p \), and \( \Lambda_t = \text{diag}(\lambda_{1t}, ..., \lambda_{pt}) \) whose elements are all positive.

Equation (3.2.2) defines the spectral decomposition of the error covariance matrix. Assumption 3.2.1 implies that the conditional error covariance is time-varying in the way that the eigenvectors remain constant through time and only the corresponding eigenvalues can vary.

Assumption 3.2.1 is different from the constant conditional correlation (CCC) decomposition in Bollerslev (1990). Under the CCC assumption, the contemporaneous correlation structure of the errors is assumed time-invariant, while under Assumption 3.2.1, both the correlation and
the variances of the error vector are time-varying. However, Assumption 3.2.1 implies that there exists a matrix (several linear combinations) such that left-multiplying the matrix to both sides of (3.2.1) can remove the contemporaneous correlation. Assumption 3.2.1 is applicable in the structural vector models to help identifying the shocks. We see that in that case, the vector of eigenvalues are simply the vector of variances of the structural model with identified shocks, and hence may be heteroskedastic. Compared to Assumption 3.2.1, although the CCC assumption implies a constant correlation structure, the correlation between errors cannot be removed if the variances are time-variant.

Under Assumption 3.2.1, the log-likelihood function for observation \( t = 1, ..., T \) based on vector Gaussian distributed errors is:

\[
\log L_t = c - \frac{1}{2} \log |\Sigma_t| - \frac{1}{2} u_t'\Sigma_t^{-1}u_t \\
= c - \frac{1}{2} \log |\Lambda_t| - \frac{1}{2} w_t'\Lambda_t^{-1}w_t \\
= c - \frac{1}{2} \sum_{i=1}^{p} (\log \lambda_{it} + w^2_{it}\lambda^{-1}_{it})
\]  

(3.2.3)

where \( w_t = P'u_t = (w_{1t}, ..., w_{pt})' \). When the error vector is not Gaussian, (3.2.3) is called quasi Gaussian log-likelihood function for observation \( t \). Let \( \varphi_i \) be the vector of parameters in \( \lambda_{it} \) for \( i = 1, ..., p \), and define \( \varphi = (\varphi'_1, ..., \varphi'_p)' \). Let \( \phi \) be the vector of the parameters in the conditional mean. Consequently, \( \theta = (\varphi', \phi')' \) is the vector of all parameters except the ones in \( P \). Under Assumption 3.2.1, the matrix \( P \) does not contribute to maximizing the log-likelihood function, but serves to identify \( \Lambda_t \) in \( \Sigma_t = P\Lambda_tP' \). Therefore \( \theta \) excludes \( P \).

Based on Assumption 3.2.1, I make the following assumption:

**Assumption 3.2.2.** The time-varying components \( \lambda_{it} = h_i(x_t), i = 1, ..., p \), where \( h_i(x_t) \) is a general positive function. The function \( h_i(x_t) \) is at least second-order differentiable almost everywhere. Furthermore, the argument \( x_t = \varphi'_iz_{it} \), where \( \varphi_i \) is a vector of parameters and \( z_{it} \) is a vector of variables.

Assuming \( h_i \) to be at least second-order differentiable ensures the existence of the corresponding information matrix. Assumption 3.2.2 allows for a wide variety of error covariance structures. The exponential function \( h_i(x_t) = \exp(x_t) \) is one possibility, which ensures the function is strictly positive-valued. Although the functional form of \( h_i \) is quite flexible, it does not play a role in deriving the test statistic. In the following, I denote \( h'_i \) the first-order derivative and \( h''_i \) the second-order derivative of the the function \( h_i \) with respect to \( x_t \). The definition of \( z_{it} \) depends on the alternative to parameter constancy. Several cases will be introduced later.

### 3.3 LM test statistic

Our focus is on testing the constancy of the whole covariance matrix when the alternative is characterized by Assumptions 3.2.1 and 3.2.2. The null hypothesis to be tested is thus:

\[
H_0 : \quad \lambda_{it} = \lambda_i, \quad i = 1, ..., p.
\]  

(3.3.1)
or, put differently,

\[ H_0 : \varphi = (\varphi_i, 0, ..., 0)' \] (3.3.2)

That is, the vector \( \varphi \) has only one non-zero element under the null hypothesis.

The tests to be considered here are Lagrange-multiplier tests. This family of tests has the advantage that there is no need to estimate the model under the alternative hypothesis. Consequently, I only have to estimate the model under the null hypothesis (3.3.1). In order to derive the test, the log-likelihood function of each observation in (3.2.3) is needed. I define the average score vector and the average information matrix of the quasi log-likelihood function as follows:

\[ s(\theta) = T^{-1} \sum_{t=1}^{T} \frac{\partial \log L_t}{\partial \theta} \] (3.3.3)

\[ I(\theta) = -T^{-1} \sum_{t=1}^{T} E \left[ \frac{\partial^2 \log L_t}{\partial \theta \partial \theta'} \right] \] (3.3.4)

where \( \log L_t \) has been defined in (3.2.3). Let \( \hat{\theta}, \hat{\phi}, \hat{\varphi}, \) and \( \hat{\varphi}_i, i = 1, ..., p, \) be the estimates of the parameters under the null hypothesis. Thus, I have the average score vector \( s(\hat{\theta}) \) and the average information matrix \( I(\hat{\theta}) \) evaluated under the null hypothesis. The LM test statistic takes the form

\[ LM = Ts(\hat{\theta})'I^{-1}(\hat{\theta})s(\hat{\theta}) \] (3.3.5)

and is asymptotically \( \chi^2 \) distributed, with the the degrees of freedom equal to the number of restrictions, when the null hypothesis is valid.

I define the corresponding blocks of the average score vector and of the average information matrix of the quasi log-likelihood function as follows:

\[ s_{\varphi}(\theta) = T^{-1} \sum_{t=1}^{T} \frac{\partial \log L_t}{\partial \varphi} \] (3.3.6)

\[ I_{\varphi}(\theta) = -T^{-1} \sum_{t=1}^{T} E \left[ \frac{\partial^2 \log L_t}{\partial \varphi \partial \varphi'} \right] \] (3.3.7)

Thus, under Assumption 3.2.2, the Lagrange-multiplier test (3.3.5) can be equivalently applied as follows:

\[ LM = Ts_{\varphi}(\hat{\theta})'I^{-1}(\hat{\theta})s_{\varphi}(\hat{\theta}) \] (3.3.8)

see Godfrey (1978), Breusch and Pagan (1978) and Breusch and Pagan (1980) for details. I have the following theorem:

**Theorem 3.3.1.** Under Assumption 3.2.1 and 3.2.2, the corresponding blocks of the average score vector and of the average information matrix of the quasi Gaussian log-likelihood based in (3.2.3) are

\[ s_{\varphi}(\theta) = (2T)^{-1} \sum_{t=1}^{T} \left[ \zeta_1 \tilde{g}_{1t} \tilde{z}_{1t}', ..., \zeta_p \tilde{g}_{pt} \tilde{z}_{pt}' \right]' \] (3.3.9)

\[ I_{\varphi,i}(\theta) = (2T)^{-1} \sum_{t=1}^{T} \tilde{z}_{it} \tilde{z}_{it}' \] (3.3.10)
where $\tilde{\zeta}_i = \tilde{h}' \tilde{\lambda}_i^{-1}$, $\tilde{g}_{it} = \tilde{w}_{it}^2 / \tilde{\lambda}_i - 1$, $h'_i$ is the scalar first-order derivative of the positive function $h_i$, and they are evaluated under the null hypothesis of constancy. The LM test statistic (3.3.5) has the following form:

$$LM = \frac{1}{2} \sum_{i=1}^{p} \left[ \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}'_{it} \right) \left( \sum_{t=1}^{T} \tilde{z}_{it} \tilde{z}'_{it} \right)^{-1} \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}_{it} \right) \right].$$

(3.3.11)

Proof. See Appendix A. □

Note that $\tilde{z}_{it}$ may contain the transformed error term $\tilde{w}_t$ estimated from the restricted model. Moreover, it is seen from (3.3.11) that the general positive function $h_i$ and its derivative have been canceled out as the argument of $h_i$ is a constant under $H_0$. There is thus no need to uniquely define the functional form of $h_i$ when setting up the test.

Consider the fact that $T^{-1} \sum_{t=1}^{T} \tilde{g}_{it}^2$ converges to 2 in probability under the null hypothesis and that the errors are Gaussian. Denote

$$R^2_i = \left( \sum_{t=1}^{T} \tilde{g}_{it}^2 \right)^{-1} \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}'_{it} \right) \left( \sum_{t=1}^{T} \tilde{z}_{it} \tilde{z}'_{it} \right)^{-1} \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}_{it} \right),$$

(3.3.12)

for $i = 1, \ldots, p$. Computing $R^2_i$ is quite easy. After obtaining the sequence $\{\tilde{g}_{it}\}_{t=1}^{T}$ for $i = 1, \ldots, p$, run a simple auxiliary regression of $\tilde{g}_{it}$ on $\tilde{z}_{it}$ and collect the residuals. Denote the $SSG_i$ the sum of squared $\tilde{g}_{it}$, and the $RSS_i$ the corresponding residual sum of squares in the auxiliary regression. It follows that

$$R^2_i = \frac{SSG_i - RSS_i}{SSG_i}$$

(3.3.13)

Thus, the LM statistic (3.3.5) can also be asymptotically evaluated by using the following statistic:

$$\sum_{i=1}^{p} T \frac{SSG_i - RSS_i}{SSG_i} = \sum_{i=1}^{p} TR^2_i$$

(3.3.14)

The test can be carried out as follows:

- Estimate the vector model (3.2.1) under the null hypothesis of constant covariances. Collect the residuals $\tilde{u}_t$, $t = 1, \ldots, T$. Compute the empirical covariance matrix $\tilde{\Sigma}$, and the eigenvalue decomposition $\tilde{\Sigma} = \tilde{P} \tilde{\Lambda} \tilde{P}'$, where $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p)$.

- Compute the transformed residuals $\tilde{w}_t = \tilde{P}' \tilde{u}_t$, and $\tilde{g}_{it} = \tilde{w}_{it}^2 / \tilde{\lambda}_i - 1$, for $t = 1, \ldots, T$, $i = 1, \ldots, p$.

- For each equation, regress $\tilde{g}_{it}$ on $\tilde{z}_{it}$ and compute the corresponding $TR^2_i$. Compute the LM test $\sum_{i=1}^{p} TR^2_i$.

In the next section, I will discuss different specifications of $\tilde{z}_{it}$. 
3.4 Specifications for heteroskedastic residuals

There are a number of possible specifications for heteroskedasticity in the errors. I will consider three useful covariance specifications against the homoskedasticity in the following. They have already been considered in Eklund and Teräsvirta (2007), but as already mentioned, the decomposition of $\Sigma_t$ is different from theirs.

The first time-varying variance specification (White specification), proposed in a single-equation case by White (1980) as an alternative to homoskedasticity, is obtained by defining:

$$\lambda_{it} = h_i (\sigma_i^2 + \delta_i \text{vech}(x_t x_t'))$$ (3.4.1)

where vech() represents the half-vectorization which collects the lower triangular elements of a symmetric matrix, $\delta_i$, $i = 1, ..., p$ are $p(p + 1)/2 \times 1$ parameter vectors and $x_t$ is defined as in (3.2.1). The null hypothesis of a constant covariance matrix in (3.3.1) is

$$H_0: \delta_i = 0, \ i = 1, ..., p.$$ (3.4.2)

The corresponding number of degrees of freedom of the LM test is $p^2(p + 1)/2$.

The second variance specification (ARCH specification) is obtained by defining

$$\lambda_{it} = h_i (\sigma_i^2 + \sum_{j=1}^{q} \alpha_{ij} w_{i,t-j}^2)$$ (3.4.3)

Note that we use the transformed error $w_{i,t-j}$ instead of $u_{i,t-j}$, because $\Lambda_t$ is the covariance matrix of $w_t$. The null hypothesis corresponding to (3.3.1) is

$$H_0: \alpha_{ij} = 0, \ i = 1, ..., p, \ j = 1, ..., q.$$ (3.4.4)

The corresponding number of degrees of freedom of the LM test is $pq$.

The third (smooth transition) specification is obtained by assuming $u_t$ to be a heteroskedastic error term with a smoothly changing covariance matrix, that is,

$$\Sigma_t = E_t(u_t u_t') = \Sigma_1 + G(s_t)\Sigma_2$$ (3.4.5)

where $\Sigma_1$ and $\Sigma_2$ are symmetric matrices, and $G(s_t)$ is a transition function whose value is controlled by the transition variable $s_t$.

Assume that the transition function is a real-valued, bounded, monotonically increasing and at least second-order differentiable function, e.g. a logistic function:

$$G(s_t; \gamma, c) = (1 + \exp(-\gamma(s_t - c)))^{-1}$$ (3.4.6)

where the parameter $\gamma > 0$ determines the smoothness of the transition, and $c$ is the location parameter. It is seen from (3.4.5) and (3.4.6) that the covariance matrix changes smoothly from $\Sigma_1$ to $\Sigma_1 + \Sigma_2$ as a function of $s_t$. Both $\Sigma_1$ and $\Sigma_1 + \Sigma_2$ must be positive definite matrices.

Following Assumption 3.2.1 and Equation (3.4.5), write $\Sigma_1 = P\Lambda_1 P'$ and $\Sigma_2 = P\Lambda_2 P'$. It is obvious that

$$\Sigma_t = P(\Lambda_1 + G(s_t)\Lambda_2)P' = P\Lambda_t P'$$ (3.4.7)
where

\[
\begin{align*}
\Lambda_1 &= \text{diag} (\lambda_{11}, \ldots, \lambda_{1p}) \\
\Lambda_2 &= \text{diag} (\lambda_{21}, \ldots, \lambda_{2p}) \\
\Lambda_t &= \text{diag} (\lambda_{1t}, \ldots, \lambda_{pt}) \\
\lambda_{it} &= \lambda_{1i} + G(s_t)\lambda_{2i} \\
\end{align*}
\]

for \( i = 1, \ldots, p \).

The null hypothesis under the specification (3.4.6), (3.4.7) and (3.4.8) is: \( H_0 : \gamma = 0 \). It is seen that under the null hypothesis \( G(s_t) = 1/2 \) and hence the parameters in \( \Sigma_2 \) are not identified. In order to solve this problem, the first-order Taylor expansion of (3.4.6) around \( \gamma = 0 \) can be employed here, see Luukkonen et al. (1988). This means writing

\[
\lambda_{it} = \lambda_{1i} + (as_t + b + r_t)\lambda_{2i} \approx \lambda_{0i}^* + \lambda_{1i}^* s_t,
\]

where \( a \) and \( b \) are constants, and \( r_t \) is the remainder. In this case, the null hypothesis is:

\[
H_0 : \lambda_{1i}^* = 0, \quad i = 1, \ldots, p.
\]

The corresponding number of degrees of freedom of the LM test is \( p \).

It is worth mentioning that the smooth transition specification can be easily extended by applying a higher-order Taylor expansion. This may often increase the power of the test. For a Taylor expansion of order \( N > 1 \), (3.4.9) can be extended to:

\[
\lambda_{it} \approx \sum_{n=0}^{N} \lambda_{ni}^* s_t^n.
\]

The null hypothesis is

\[
H_0 : \lambda_{ni}^* = 0, \quad i = 1, \ldots, p, \quad n = 1, \ldots, N.
\]

In this case, the number of degrees of freedom of the LM test is \( pN \). However, in the following, the focus will be on the first-order approximation to the logistic function (3.4.6).

### 3.5 Finite sample properties of the test

When investigating the properties of a classical test statistic, two aspects are of prime importance. First, I have to check whether the empirical size of the test (the probability of rejecting the null when it is true) is close to the nominal size (used to calculate the asymptotic critical values) at typical sample sizes. Given that empirical size is a reasonable approximation to the nominal size, I then have to investigate the empirical power of the test (the probability of rejecting the null when it is false) for a number of different alternative hypotheses.
In order to investigate the size and power properties of the test in finite samples, I conduct a series of Monte Carlo simulations. I consider the bivariate case $p = 2$, the trivariate case $p = 3$ and a high-dimensional case $p = 5$. The data generating process is a special case of (3.2.1):

$$y_{i,t} = 0.8y_{i,t-1} + u_{i,t}, \quad i = 1, \ldots, p$$  \hspace{1cm} (3.5.1)

This is a simple design in the sense that the variables in the VAR model are only linked through the covariance matrix. The finite sample sizes I investigate in the size experiments are $T = 100$ and $T = 500$. Actually the settings in (3.5.1) is exactly the same as the one in Eklund and Teräsvirta (2007), and this makes it easier to compare the size properties of the two tests under the null hypothesis of constant covariance matrix over time. Thus, I will not repeat their size experiments here.

Furthermore, I employ the size discrepancy and power plots recommended by Davidson and MacKinnon (1998). The number of replications of the Monte Carlo simulations is $N = 10000$.

For space reasons, only a fraction of the results are shown. The remaining ones, as well as the R programmes, are available upon request.

### 3.5.1 Size experiments

In investigating the finite sample size behaviour of the test statistics, I set $u_{i,t}$ either i.i.d. Gaussian or $t(5)$ distribution in the basic data generating process (3.5.1). In the bivariate case, the covariance matrix is

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$  \hspace{1cm} (3.5.2)

where $\rho = 0.9, 0$ and $-0.9$. Eklund and Teräsvirta (2007) used the same design for the bivariate case. In the trivariate case, The covariance matrix is

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$  \hspace{1cm} (3.5.3)

where $\rho = 0.9$ and $0$. For the high-dimensional case $p = 5$, I only report the results from $\rho = 0$ for space reasons. It can be seen in the following that from the case $p = 3$ the correlation may affect the size properties, and it will also do when $p = 5$. In the following, I conduct the three LM tests by setting

**White specification**: $z_{it} = (1, \text{vech}(x_t^2 x_t')')'$, where $x_t = y_{t-1}$;

**ARCH(5) specification**: $\tilde{z}_{it} = (1, \tilde{w}_{i,t-1}^2, \ldots, \tilde{w}_{i,t-5}^2)'$, where $\tilde{w}_{i,t-q}, q = 1, \ldots, 5$ are estimates of the transformed errors under the null hypothesis of constancy;

**Smooth transition specification**: $z_{it} = (1, \tau)'$, where $\tau = t/T$. 

109
In the smooth transition specification I choose \( s_t = \tau \) to be the transition variable. Note that one can choose other variables to be the transition variable. Here I just focus on testing whether the covariance matrix changes over time. Thus, I choose the normalized time \( \tau = t/T \).

I compare the size properties of the three LM tests when \( \rho = 0 \) in Figures 3.1 to 3.8. It is seen that the empirical distributions of all three tests converge towards their limiting distribution when \( T \) increases. The test against the White specification over-rejects in all the cases, especially when the errors are \( t(5) \), whereas the test against the ARCH(5) specification seems to under-reject. It is seen from Figures 3.5 to 3.8 that when the errors are \( t(5) \), the tests against the White specification and the ARCH(5) specification have greater size distortion than test against the smooth transition specification. The latter test performs well in almost all cases.

In order to see whether the correlation \( \rho \) plays a role, I report the corresponding size-discrepancy plots from Figure 3.9 to 3.20. Note that the sample sizes I used now are \( T = 25, 50, 100, 250, 500 \) and 1000. Figures from 3.9 to 3.14 depict the bivariate case, whereas Figures from 3.15 to 3.20 show the trivariate case. It is seen that the correlation does not play a role in the size properties in the bivariate case, but it may do so in trivariate case. It is worth noting that in Figure 3.15 the test against the White specification converges towards its limiting distribution a bit faster when there is no correlation in errors in finite sample case from \( T = 25 \) to 50. This is also the case when the errors are \( t(5) \), see Figure 3.18. The correlation does not seem to affect the size of the test against the ARCH(5) specification. This may be due to the fact that the ARCH specification (3.4.3) does not allow for any cross-effects between different equations. Moreover, the correlation does not affect the size of the test against the smooth transition specification. It is seen that this test is free from size distortion in almost all cases considered.

### 3.5.2 Power experiments

In power simulations we assume that the data generating process has \( h_i(\varphi' z_{it}) = \varphi' z_{it} \) for simplicity. I only consider the bivariate case \( p = 2 \) and the high-dimensional case \( p = 5 \). The data generating process is still (3.5.1), but now the covariance matrix will change over time. The errors are drawn from the corresponding conditional vector Gaussian distribution. I will conduct the following three kinds of power simulations.

**Power simulations under Assumption 3.2.1**

In this case, the covariance matrix of the errors will change under Assumption 3.2.1. That is,

\[
\Sigma_t = P\Lambda_t P'.
\]  

(3.5.4)

First, I consider the case that the covariance matrix changes once through time and the transition is threshold-like. The threshold point is at \( T/2 \), that is, \( c = 0.5 \). Figures 3.21 and 3.22 depict the rejection frequencies when \( p = 2 \), and Figures 3.23 and 3.24 depict the rejection frequencies when \( p = 5 \).
I also consider the special case that the covariance matrix changes once through time but the transition is rather smooth. The smooth function takes the form (3.4.6) where $\gamma = 12$ and $c = 0.5$. Let $s_t = t/T$, which ensures that the sample size $T$ does not matter the smoothness of the transition. Figures 3.29 and 3.30 depict the rejection frequencies when $p = 2$, and Figures 3.31 and 3.32 depict the rejection frequencies when $p = 5$. It is seen that the test against the smooth transition specification performs always the best in both the case in which the transition is threshold and the case in which the transition is rather smooth. This is not surprising because the DGP is just the smooth transition specification.

Next I investigate the power of the tests when $\lambda_{it}$ in the covariance matrix evolves through time using the ARCH specification (3.4.3). More specifically, I assume an ARCH(2) process for all eigenvalues with parameters $\sigma_i^2 = 1$, $\alpha_{i1} = 0.25$ and $\alpha_{i2} = 0.2$. Figures 3.45 and 3.46 depict the rejection frequencies when $p = 2$, and Figures 3.47 and 3.48 show them when $p = 5$. It is not surprising that the test against the ARCH specification outperforms the others.

Finally, I consider the case when $\lambda_{it}$ in the covariance matrix evolves through time using the White specification (3.4.1). The parameters are $\sigma_i^2 = 1$ and $\delta_i = (1, ..., 1)'$ is a $p(p + 1)/2 \times 1$ vector. Figures 3.49 and 3.50 depict the rejection frequencies when $p = 2$, and Figures 3.51 and 3.52 show them when $p = 5$. This time the test against the White specification is the best performer. It seems that the other tests have only trivial power in this case even in large samples.

**Power simulations when Assumption 3.2.1 is violated**

It is important to investigate the consequences of violating Assumption 3.2.1. Since Assumption 3.2.1 is very restrictive, one may argue that if the null hypothesis of constant covariance matrix is rejected, it would be difficult without any further investigation to distinguish between a rejection due to time-varying $\Lambda$, time-varying $P$ or a combination of the two. Here I investigate the case that

$$\Sigma_t = P_t \Lambda P_t'.$$  \hfill (3.5.5)

The data generating process takes the form (3.5.1). $\Lambda = \text{diag}(0.50, 0.40)$ for $p = 2$, and $\Lambda = \text{diag}(0.50, 0.40, 0.30, 0.25, 0.20)$ for $p = 5$.

I consider the case that $P$ changes once through time. The transition is threshold-like, and the threshold point is at $T/2$. The way to choose the matrices $P_1$ and $P_2$ are tricky. Let $U_i$, $i = 1, 2$, be a $p \times p$ matrix whose elements are a sample of independent draws from a standard Gaussian distribution. Let $P_i$ be the eigenvectors of $U_i U_i'$. Then, the rotation (ortonormal basis) $P_i$ are uniformly distributed over the set of all rotation matrices (ortonormal bases). Figures 3.25 and 3.26 depict the rejection frequencies when $p = 2$, and Figures 3.27 and 3.28 depict the rejection frequencies when $p = 5$.

The case that $P$ changes smoothly through time will also be considered. I use the way mentioned in the previous paragraph to pick $P_1$ and $P_2$. However, the conditional covariance
matrix at time $t$ should be computed as follows:

$$\Sigma_t = P_1 \Lambda P_1' (1 - G(\tau; \gamma, c)) + P_2 \Lambda P_2' G(\tau; \gamma, c) \quad (3.5.6)$$

where $G$ has been defined in (3.4.6), $\tau = t/T$, $\gamma = 12$ and $c = 0.5$. Figures 3.33 and 3.34 depict the rejection frequencies when $p = 2$, and Figures 3.35 and 3.36 do the same when $p = 5$.

The results show that in low-dimensional case all the tests from Assumption 3.2.1 have very little power, whereas in high-dimensional case they have some power. The test can detect the change in $\Lambda$, but it is not very sensitive to the change in the rotation matrix $P$.

**Power simulations under the constant conditional correlation assumption**

I also simulated the situation when the constant conditional correlation (CCC) assumption is satisfied. Under the CCC assumption, the time-varying covariance matrix can be decomposed as follows:

$$\Sigma_t = D_t Q D_t', \quad (3.5.7)$$

where

$$D_t = \text{diag}(\omega_{1t}^{1/2}, \ldots, \omega_{pt}^{1/2}) \quad (3.5.8)$$

is a diagonal matrix of error standard deviations, and $Q = [\rho_{ij}]$ is the corresponding correlation matrix. The value of $\rho_{ij}$ is chosen in following way. Let $U$ be a $p \times p$ matrix whose elements are a sample of independent draws from a standard Gaussian distribution, and denote $u_{ij}$ the element of $U_i U_j'$ where $i$ is the row number and $j$ is the column number. $\rho_{ij} = u_{ij} / \sqrt{u_{ii} u_{jj}}$.

I consider first that the error variances $\omega_{it}$ change once through time and the transition is threshold-like. The threshold point is as $T/2$, that is, $c = 0.5$. Figures 3.37 and 3.38 depict the rejection frequencies when $p = 2$, and Figures 3.39 and 3.40 show them when $p = 5$.

Second, I consider the case in which the error variances $\omega_{it}$ change smoothly through time from $\omega_{i1}$ to $\omega_{i2}$. The transition function takes the form (3.4.6). Let $s_t = t/T$, $\gamma = 12$ and $c = 0.5$. Figures 3.41 and 3.42 depict the rejection frequencies when $p = 2$, and Figures 3.43 and 3.44 do the same when $p = 5$.

It is seen that the tests derived from Assumption 3.2.1 have very satisfactory power, though the model is misspecified. It is surprising to see that the test against the smooth transition specification has the best performance not only in the threshold case but in the smooth transition case as well. It performs even better than the ET test in Eklund and Teräsvirta (2007) especially in the high-dimensional finite sample case, see Figures 3.39 and 3.43.

### 3.6 Concluding remarks

In this work, I develop a test of constancy of the error covariance matrix against the alternative that the covariance matrix changes over time. The test is based on the spectral decomposition
of the covariance matrix. This implies that the conditional error covariance is time-varying in the way that the eigenvectors remain constant through time and only the corresponding eigenvalues are time-varying. There exist linear combinations which make error vectors in the corresponding structural vector model contemporaneously uncorrelated but still heteroskedastic. I design a family of LM tests against the alternative hypothesis that the errors are time-varying and follow parametric specifications.

Three specifications are considered. They are: the White specification which generalizes the heteroskedasticity test of White (1980), the ARCH specification which generalizes the test against autoregressive conditional heteroskedasticity of Engle (1982) and the smooth transition specification which generalizes the test against smoothly changing variance of Medeiros and Veiga (2003). The test of constancy of the error covariance matrix is very easy to implement and use. From the simulation experiments it is seen that the test has satisfactory size and power properties even in high-dimensional vector models.
Appendix A: Proof for Theorem 3.3.1

Proof. Based in (3.2.3), assume that $\varphi = (\varphi_1', ..., \varphi_p')'$ and $\lambda_{it} = h_i(\varphi_i'z_{it})$, $i = 1, ..., p$. I have

$$\frac{\partial \log L_t}{\partial \varphi_i} = \frac{\partial \log L_t}{\partial \lambda_{it}} \frac{\partial \lambda_{it}}{\partial \varphi_i},$$

(3.6.1)

$$\frac{\partial^2 \log L_t}{\partial \varphi_i \partial \varphi'_i} = \frac{\partial^2 \log L_t}{\partial \lambda_{it}^2} + \frac{\partial \log L_t}{\partial \lambda_{it}} \frac{\partial \lambda_{it}}{\partial \varphi_i} \frac{\partial \lambda_{it}}{\partial \varphi_i'} + \frac{\partial^2 \lambda_{it}}{\partial \varphi_i \partial \varphi_i'},$$

(3.6.2)

$$\frac{\partial^2 \log L_t}{\partial \varphi_i \partial \varphi'_j} = 0 \text{ for } i \neq j.$$  

(3.6.3)

Furthermore, in (3.6.1), (3.6.2) and (3.6.3), I have

$$\frac{\partial \log L_t}{\partial \lambda_{it}} = \frac{1}{2\lambda_{it}} g_{it},$$

(3.6.4)

$$\frac{\partial^2 \log L_t}{\partial \lambda_{it}^2} = \frac{1}{2\lambda_{it}^2} \left( 1 - 2w_{it}^2 \lambda_{it}^{-1} \right),$$

(3.6.5)

$$\frac{\partial \lambda_{it}}{\partial \varphi_i} = h_i' z_{it},$$

(3.6.6)

$$\frac{\partial^2 \lambda_{it}}{\partial \varphi_i \partial \varphi'_i} = h_i'' z_{it} z_{it}',$$

(3.6.7)

where $g_{it} = w_{it}^2/\lambda_{it} - 1$, $h_i'$ and $h_i''$ are the scalar first-order and second-order derivatives of the positive function $h_i$, respectively.

The corresponding blocks of the average score vector and of the average information matrix of the quasi log-likelihood are defined to be:

$$s_\varphi(\theta) = T^{-1} \sum_{t=1}^{T} \frac{\partial \log L_t}{\partial \varphi}$$

(3.6.8)

$$I_\varphi(\theta) = -T^{-1} \sum_{t=1}^{T} E \left[ \frac{\partial^2 \log L_t}{\partial \varphi \partial \varphi'} \right].$$

(3.6.9)

It can be seen from (3.6.3) that the corresponding hession matrix $(\partial^2 \log L_t/\partial \varphi \partial \varphi')$ is block diagonal, and so is $I_\varphi(\theta)$ in (3.6.9).

From (3.6.1), (3.6.4) and (3.6.6), it is seen that:

$$\frac{\partial \log L_t}{\partial \varphi_i} = \left( \frac{h_i'}{2\lambda_{it}} \right) g_{it} z_{it}$$

(3.6.10)

Thus, under the null hypothesis of constant covariance over time, I have the average score vector:

$$s_\varphi(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} \left[ \left( \frac{\tilde{h}_1'}{2\tilde{\lambda}_1} \right) \tilde{g}_{it} \tilde{z}_1', ..., \left( \frac{\tilde{h}_p'}{2\tilde{\lambda}_p} \right) \tilde{g}_{it} \tilde{z}_p' \right]'$$

(3.6.11)

where $\tilde{g}_{it} = \tilde{w}_{it}^2/\tilde{\lambda}_i - 1$.  

114
From (3.6.2), and (3.6.4) to (3.6.7), I have that
\[
E \left[ \frac{\partial^2 \log L_t}{\partial \varphi_i \partial \varphi'_i} \right] = E \left[ \frac{\partial^2 \log L_t}{\partial \lambda_{it}^2} \right] + E \left[ \frac{\partial \log L_t}{\partial \varphi_i} \frac{\partial^2 \lambda_{it}}{\partial \varphi'_i} \right] + E \left[ \frac{1}{2} \left( \frac{h'_i}{\lambda_{it}} \right)^2 \left( 1 - 2w^2_{it} \lambda_{it}^{-1} \right) z_{it} z'_{it} \right] + E \left[ \frac{1}{2\lambda_{it}} g_{it} h''_{it} z_{it} z'_{it} \right] = -\frac{1}{2} \left( \frac{h'_i}{\lambda_{it}} \right)^2 E \left[ z_{it} z'_{it} \right], \tag{3.6.12}
\]
due to the fact that \( E \left[ 1 - 2w^2_{it} \lambda_{it}^{-1} \right] = -1 \) and \( E \left[ g_{it} \right] = 0 \).

Thus, under the null hypothesis of constant covariance over time, the diagonal block \( i \) of the average information matrix takes the form:
\[
I_{\varphi,i}(\tilde{\theta}) = \frac{1}{2T} \sum_{t=1}^{T} \left( \frac{\tilde{h}'_i}{\tilde{\lambda}_i} \right)^2 E \left[ \tilde{z}_{it} \tilde{z}'_{it} \right]. \tag{3.6.13}
\]

The LM test can be consistently estimated as follows:
\[
LM = \frac{1}{2} \sum_{i=1}^{p} \left[ \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}_{it} \right) \left( \sum_{t=1}^{T} \tilde{z}_{it} \tilde{z}'_{it} \right)^{-1} \left( \sum_{t=1}^{T} \tilde{g}_{it} \tilde{z}_{it} \right) \right], \tag{3.6.14}
\]
where \( (h'_i/\lambda_{it}) \) has been cancelled out. \( \Box \)
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.1: The size discrepancy plot: Gaussian errors $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.

Figure 3.2: The size discrepancy plot: Gaussian errors $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.
Figure 3.3: The size discrepancy plot: Gaussian errors \( p = 5, T = 100 \) and \( N = 10000 \). LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95\% confidence region.

Figure 3.4: The size discrepancy plot: Gaussian errors \( p = 5, T = 500 \) and \( N = 10000 \). LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95\% confidence region.
Figure 3.5: The size discrepancy plot: t(5) errors $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.

Figure 3.6: The size discrepancy plot: t(5) errors $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.
Figure 3.7: The size discrepancy plot: t(5) errors $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.

Figure 3.8: The size discrepancy plot: t(5) errors $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted). The grey area represents the 95% confidence region.
Figure 3.9: The size discrepancy plot of LM test against White specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. 

Chapter 3. Testing constancy of the error covariance matrix
Figure 3.10: The size discrepancy plot of LM test against ARCH specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. 

Chapter 3. Testing constancy of the error covariance matrix
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.11: The size discrepancy plot of LM test against smooth transition specification: Gaussian errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. 

122
Figure 3.12: The size discrepancy plot of LM test against White specification: $t(5)$ errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. 

Chapter 3. Testing constancy of the error covariance matrix
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.13: The size discrepancy plot of LM test against ARCH specification: $t(5)$ errors $p = 2$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, -0.9$. 

124
Figure 3.14: The size discrepancy plot of LM test against smooth transition specification: $t(5)$ errors $p = 2, T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0, 0.9$. 
Figure 3.15: The size discrepancy plot of LM test against White specification: Gaussian errors $p = 3$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0$. 

Chapter 3. Testing constancy of the error covariance matrix
Figure 3.16: The size discrepancy plot of LM test against ARCH specification: Gaussian errors \( p = 3, \) \( T = 25, 50, 100, 250, 500, 1000 \) from 1 to 6 and \( N = 10000. \) From top to bottom: \( \rho = 0.9, 0. \)
Figure 3.17: The size discrepancy plot of LM test against smooth transition specification: Gaussian errors $p = 3$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0.$
Figure 3.18: The size discrepancy plot of LM test against White specification: t(5) errors $p = 3$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0.129$. 

Chapter 3. Testing constancy of the error covariance matrix
Figure 3.19: The size discrepancy plot of LM test against ARCH specification: t(5) errors $p = 3$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0$. 

Chapter 3. Testing constancy of the error covariance matrix
Figure 3.20: The size discrepancy plot of LM test against smooth transition specification: t(5) errors $p = 3$, $T = 25, 50, 100, 250, 500, 1000$ from 1 to 6 and $N = 10000$. From top to bottom: $\rho = 0.9, 0.$
Figure 3.21: The power plot: $\Sigma_t = P \Lambda_t P'$ with threshold change at $T/2$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.22: The power plot: $\Sigma_t = P \Lambda_t P'$ with threshold change at $T/2$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.23: The power plot: $\Sigma_t = P\Lambda_t P'$ with threshold change at $T/2$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.24: The power plot: $\Sigma_t = P\Lambda_t P'$ with threshold change at $T/2$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.25: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.26: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.27: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dotted-dashed).

Figure 3.28: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with threshold change at $T/2$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dotted-dashed).
Figure 3.29: The power plot: $\Sigma_t = P \Lambda_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.30: The power plot: $\Sigma_t = P \Lambda_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.31: The power plot: $\Sigma_t = P\Lambda_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.32: The power plot: $\Sigma_t = P\Lambda_t P'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.33: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.34: The power plot: $\Sigma_t = P_t \Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.35: The power plot: $\Sigma_t = P_t\Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.36: The power plot: $\Sigma_t = P_t\Lambda P_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.37: The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Ter"asvirta (2007) (dot-dashed).

Figure 3.38: The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Ter"asvirta (2007) (dot-dashed).
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.39: The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.40: The power plot: $\Sigma_t = D_t Q D_t'$ with threshold change at $T/2$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Figure 3.41: The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.42: The power plot: $\Sigma_t = D_t Q D_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.43: The power plot: $\Sigma_t = D_tQD_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).

Figure 3.44: The power plot: $\Sigma_t = D_tQD_t'$ with smooth change at $T/2$, $\gamma = 12$, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed), LM test against White specification (dotted) and test in Eklund and Teräsvirta (2007) (dot-dashed).
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.45: The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

Figure 3.46: The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).
Chapter 3. Testing constancy of the error covariance matrix

Figure 3.47: The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

Figure 3.48: The power plot: $\Sigma_t = P\Lambda_t P'$ with ARCH specification, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).
Figure 3.49: The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 2$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

Figure 3.50: The power plot: $\Sigma_t = P\Lambda_t P'$ with White specification, $p = 2$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).
Figure 3.51: The power plot: $\Sigma_t = P \Lambda_t P'$ with White specification, $p = 5$, $T = 100$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).

Figure 3.52: The power plot: $\Sigma_t = P \Lambda_t P'$ with White specification, $p = 5$, $T = 500$ and $N = 10000$. LM test against smooth transition specification (solid), LM test against ARCH specification (dashed) and LM test against White specification (dotted).
Bibliography


Davies, R. B.: 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247–254.

Davies, R. B.: 1987, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**, 33–43.


Hansen, B. E.: 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.


Bibliography


Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregres-


R code

Specification

# LM tests
# output: 1. p-value, 2.test statistic, 3.degree of freedom

SCORETEST <- function(mE, mX, mZ, flag)
# a function which runs the auxiliary regression
# mE, mX, mZ must be matrices!
# returns the LM test statistic and the degree of freedom
{
  iT = dim(mE)[1]
  ip = dim(mE)[2]
  ix = dim(mX)[2]
  iz = dim(mZ)[2]
  iK = ix + iz
  iDF = iz * ip

  RSS0 = t(mE)%*%mE
  # RSS0
  mXX = cbind(mX, mZ)
  mU = svd(mXX)$u
  mR = mE - mU%*%t(mU)%*%mE
  RSS1 = t(mR)%*%mR
  # RSS1
  R0 = svd(RSS0)$d
  R1 = svd(RSS1)$d

  LM = list()
  FT = list()
  WK = list()
  RA = list()
if((flag%%2)==1){
# LM test
dTR = sum(diag(solve(RSS0)%*%RSS1))
test = iT*(ip-dTR)
LM$pval = 1-pchisq(test,df=iDF)
LM$test = test
LM$df = iDF
# rescale test
iDF1 = iDF
iDF2 = ip*(iT-iK)
test = LM$test * (iT-iK) / (iT*LM$df)
FT$pval = 1-pf(test,df1=iDF1,df2=iDF2)
FT$test = test
FT$df1 = iDF1
FT$df2 = iDF2
}
flag = flag%%2
if((flag%%2)==1){
# Wilks test
Lambda = sum(log(R1))-sum(log(R0))
Lambda = Lambda * ( (ip+iz+1)*.5 + ix - iT )
WK$pval = 1-pchisq(Lambda,df=iDF)
WK$test = Lambda
WK$df = iDF
}
flag = flag%%2
if((flag%%2)==1){
# Rao’s test
iN = iT-ix-(ip+iz+1)*.5
is = sqrt( (iz*iz*ip*ip-4) / (ip*ip+iz*iz-5) )
iDF1 = iDF
iDF2 = iN*is - iz*ip*.5 + 1
RAO = exp((sum(log(R0))-sum(log(R1)))/is)-1
RAO = RAO *iDF2 /iDF1
RA$pval = 1-pf(RAO,df1=iDF1,df2=iDF2)
RA$test = RAO
RA$df1 = iDF1
RA$df2 = iDF2
}
GESCORETEST <- function(mE, mX, lZ, flag)
# a function which runs the auxiliary regression
# mE, mX must be matrices!
# lZ[[ip]] a list of matrices
# returns the LM test statistic and the degree of freedom
{
  iT = dim(mE)[1]
  ip = dim(mE)[2]
  ix = dim(mX)[2]
  iz = rep(0, ip)
  for(pter in 1:ip){
    iz[pter] = dim(lZ[[pter]])[2]
  }
  iK = ix + iz

  RSS0 = t(mE)%*%mE
  R0 = svd(RSS0)$d
  # RSS0
  mR = NULL
  for(pter in 1:ip){
    mXX = cbind(mX, lZ[[pter]])
    mU = svd(mXX)$u
    vR = c(mE[,pter] - mU%*%t(mU)%*%mE[,pter])
    mR = cbind(mR,vR)
  }
  RSS1 = t(mR)%*%mR
  R1 = svd(RSS1)$d
  # RSS1

  iDF = sum(iz)

  LM = list()
  FT = list()
  WK = list()
  RA = list()

  return(list(LM = LM, FT = FT, WK = WK, RA = RA))
}
if((flag%%2)==1){
    # LM test
    dTR = sum(diag(solve(RSS0)%*%RSS1))
    test = iT*(ip-dTR)
    LM$pval = 1-pchisq(test,df=iDF)
    LM$test = test
    LM$df = iDF
    # rescale test
    iDF1 = iDF
    iDF2 = sum(iT-iK)
    test = LM$test * sum(iT) / (iT*LM$df*ip)
    FT$pval = 1-pf(test,df1=iDF1,df2=iDF2)
    FT$test = test
    FT$df1 = iDF1
    FT$df2 = iDF2
}
flag = flag%/%2
if((flag%%2)==1){
    # Wilks test
    Lambda = sum(log(R1))-sum(log(R0))
    Lambda = Lambda * ( (ip+sum(iz)/ip+1)*.5 + ix - iT )
    WK$pval = 1-pchisq(Lambda,df=iDF)
    WK$test = Lambda
    WK$df = iDF
}
flag = flag%/%2
if((flag%%2)==1){
    # Rao’s test
    iN = iT-ix-(ip+sum(iz)/ip+1)*.5
    is = sqrt( (sum(iz)**2-4) / (ip*ip*(sum(iz)/ip)**2-5) )
    iDF1 = iDF
    iDF2 = iN*is - sum(iz)*.5 + 1
    RAO = exp((sum(log(R0))-sum(log(R1)))/is)-1
    RAO = RAO *iDF2 /iDF1
    RA$pval = 1-pf(RAO,df1=iDF1,df2=iDF2)
    RA$test = RAO
    RA$df1 = iDF1
    RA$df2 = iDF2
}
LMTEST <- function(mY, mX, mZ)
# mY, mX, mZ must be matrices!
# returns the LM test statistic and the degree of freedom
{
  iT = dim(mY)[1]
  ip = dim(mY)[2]
  idF = dim(mZ)[2]*ip
  mE = mY - mX%*%solve(t(mX)%*%mX)%*%t(mX)%*%mY
  RSS0 = t(mE)%*%mE
  mXX = cbind(mX, mZ)
  mK = mE - mXX%*%solve(t(mXX)%*%mXX)%*%t(mXX)%*%mE
  RSS1 = t(mK)%*%mK
  dTR = sum(diag(solve(RSS0)%*%RSS1))
  LM = iT*(ip-dTR)
  pval = 1-pchisq(LM, df=idF)
  return( c(pval, LM, idF) )
}

FTEST <- function(LM, iT, iK, ip)
{
  idF1 = LM[3]
  idF2 = ip*(iT-iK)
  FT = LM[2]*(iT-iK)/(iT*LM[3])
  pval = 1-pf(FT, df1=idF1, df2=idF2)
  return( c(pval, FT, idF1, idF2) )
}

WILKS <- function(mY, mX, mZ)
# mY, mX, mZ must be matrices!
# returns the Wilks’ test statistic and the degree of freedom
{
  iT = dim(mY)[1]
  ip = dim(mY)[2]
  ix = dim(mX)[2]
  iz = dim(mZ)[2]
  idF = iz*ip
  }
mE = mY - mX%*%solve(t(mX)%*%mX)%*%t(mX)%*%mY
RSS0 = t(mE)%*%mE
mXX = cbind(mX, mZ)
mK = mE - mXX%*%solve(t(mXX)%*%mXX)%*%t(mXX)%*%mE
RSS1 = t(mK)%*%mK
RSS0 = svd(RSS0)$d
RSS1 = svd(RSS1)$d
Lambda = sum(log(RSS1))-sum(log(RSS0))
# Lambda = log(det(RSS1))-log(det(RSS0))
# Lambda = log(det(RSS1)/det(RSS0))
Lambda = Lambda * ( (ip+iz+1)*.5 + ix - iT )
pval = 1-pchisq(Lambda,df=iDF)
return( c(pval, Lambda, iDF) )

RAO <- function(mY, mX, mZ)
#mY, mX, mZ must be matrices!
#returns the Rao’s test statistic and the degree of freedom
{
  iT = dim(mY)[1]
  ip = dim(mY)[2]
  ix = dim(mX)[2]
  iz = dim(mZ)[2]
  iN = iT-ix-(ip+iz+1)*.5
  is = sqrt( (iz*iz*ip*ip-4) / (ip*ip+iz*iz-5) )
  iDF1 = iz*ip
  iDF2 = iN*is - iz*ip*.5 + 1
  mE = mY - mX%*%solve(t(mX)%*%mX)%*%t(mX)%*%mY
  RSS0 = t(mE)%*%mE
  mXX = cbind(mX, mZ)
  mK = mE - mXX%*%solve(t(mXX)%*%mXX)%*%t(mXX)%*%mE
  RSS1 = t(mK)%*%mK
  RSS0 = svd(RSS0)$d
  RSS1 = svd(RSS1)$d
  RAO = exp((sum(log(RSS0))-sum(log(RSS1)))/is)-1
  # RAO = (det(RSS0)/det(RSS1))**(1/is) - 1
  RAO = RAO *iDF2 /iDF1
  pval = 1-pf(RAO,df1=iDF1,df2=iDF2)
  return( c(pval, RAO, iDF1, iDF2) )
}
LINTEST <- function(mDat, ind, ik=0, ord=1, econ=1, flag=7)
# Linearity test
# input:
# mDat: (iT+ikk)*n data matrix including Yt, Dt, and St
# ind: index list, ind$Y, ind$D, ind$S
# ind$S is a 2-dim vec, and the second element represents the lag from 0(no lag)
# to k, such that the St can be a laged variable.
# the third element is whether to normalize
# ord: order
# econ: economy form or not
# ik: lag length
# flag from 1 to 7. That is from 001 to 111. The corresponding bit position
# represents the LM WILKS and RAO from right to left.
# output:
# LM, rescaled, Wilks and Rao
{
    # simple checking
    if(flag==0)
        return(list(success=0, msg="No test"))
    if( ik < 0 )
        return(list(success=0, msg="Invalid k"))
    if( is.null(ind$Y) )
        return(list(success=0, msg="No Yt"))
    if( is.null(ind$S) )
        return(list(success=0, msg="St shouldn't be NULL"))
    if( length(ind$S)!=3 )
        return(list(success=0, msg="Invalid dimension of St"))

    ikk = max(ik, ind$S[2])
    # ikk is defined to be the bigger one of the two lags
    iT = dim(mDat)[1] - ikk
    # notice that the first ikk elements will be discarded.
    ip = length(ind$Y)

    # Build mY
    mY = mDat[(1+ikk):(iT+ikk), ind$Y]
    dim(mY) = c(iT, ip)
    # data start from ikk+1 to ikk+iT (the original order in the raw data),
# iT elements in all.
# t = 1+ikk to iT+ikk hereafter.

# Build mX
mX = NULL
if( is.null(ind$D) ){
  iq = 0
}else{
  iq = length(ind$D)
  mX = mDat[(1+ikk):(iT+ikk),ind$D]
  dim(mX) = c(iT,iq)
}

# see mY
}
if( ik > 0 ){
  for(iter in ik:1){# left binding
    mX = cbind( mDat[(1+ikk-iter):(iT+ikk-iter),ind$Y], mX)
  }
}
if( is.null(mX) ){
  return(list(success=0,msg="No explanatory variables in the model!"))
}

# order of the x_t' = ( y_t-1' , ... , y_t-k' , D_t' ), left binding!

# Build vSt(mSt)
vSt = c(mDat[ (1+ikk-ind$S[2]):(iT+ikk-ind$S[2]) , ind$S[1] ])
if(ind$S[3] == 1) vSt = vSt/sd(vSt)

# build mZ
mZ = NULL
temp = mX
SS = vSt
if(econ==0){
  for(iter in 1:ord){
    for(jter in 1:(ik*ip+iq)){
      XS = mX[,jter]*SS
      if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ){
        temp = cbind(temp,XS)
      }
    }
  }
}
R code

SS = SS*SS
}

} else {
  for(jter in 1:(ik*ip+iq)) {
    XS = mX[,jter]*SS
    if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ) {
      temp = cbind(temp,XS)
      mZ = cbind(mZ,XS)
    }
  }
  if(ord>1) {
    XS = SS
    for(jter in 2:ord) {
      XS = XS * SS
      if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ) {
        temp = cbind(temp,XS)
        mZ = cbind(mZ,XS)
      }
    }
    if(is.null(mZ)) {
      return(list(success=0,msg="No LM test!"))
    }
  }

  RET = list()
  length(RET) = ip+1

  mU = svd(mX)$u
  mE = mY - mU%*%t(mU)%*%mY
  # joint linearity test
  RET[[ip+1]] = SCORETEST(mE=mE, mX=mX, mZ=mZ, flag=flag)

  for(iter in 1:ip) {
    temp = mE[,iter]
    dim(temp) = c(iT,1)
    # linearity tests equation by equation using the same information
      # as the joint test
    RET[[iter]] = SCORETEST(mE=temp, mX=mX, mZ=mZ, flag=flag)
  }

  return(RET)

GELINTEST <- function(mDat,ind,ik=0,ord=1,econ=1,flag=7)
# General Linearity test which allows different transition variables
# and subdim testing
# input:
# mDat: (iT+ikk)*n data matrix including Yt, Dt, and St
# ind: index list, ind$Y, ind$D, ind$S
# Notice that ind$S is a list of m-1 p*3 matrices (m=2 here such that
# the ind can be reused in the following estimation),
# and the col element represents the lag from 0(no lag) to k ,
# such that the St can be a laged variable. the third col is the norms
# for each row.
# ord: order
# econ: economy form or not
# ik: lag length
# flag from 1 to 7. That is from 001 to 111. The corresponding bit position
# represents the LM WILKS and RAO from right to left.
# output:
# LM, rescaled, Wilks and Rao
{
# simple checking
if(flag==0)
return(list(success=0,msg="No test"))
if( ik < 0 )
return(list(success=0,msg="Invalid k"))
if( is.null(ind$Y) )
return(list(success=0,msg="No Yt"))
if( is.null(ind$S) )
return(list(success=0,msg="St shouldn't be NULL"))
if( length(ind$S)!=1 )
return(list(success=0,msg="Invalid dimension of St"))

ip = length(ind$Y)
ikk = 0
if( length(ind$S) != 1 )
return(list(success=0,msg="Invalid dimension of St"))
if( !all(dim(ind$S[[1]])==c(ip,3)) )
return(list(success=0,msg="Invalid dimension of St"))

ikk = max(ikk,max( ind$S[[1]][,2] ))

# ikk is defined to be the biggest one of the ip+1 lags

itt = dim(mDat)[1] - ikk

# notice that the first ikk elements will be discarded.

# Build mY
mY = mDat[(1+ikk):(itt+ikk),ind$Y]
dim(mY) = c(itt,ip)
# data start from ikk+1 to ikk+itt (the original order in the raw data),
#   # itt elements in all.
# t = 1+ikk to itt+ikk hereafter.

# Build mX
mX = NULL
if( is.null(ind$D) ){
  iq = 0
} else{
  # mD: T by q
  iq = length(ind$D)
  mX = mDat[(1+ikk):(itt+ikk),ind$D]
dim(mX) = c(itt,iq)
  # see mY
}
if( ik > 0 ){
  for(iter in ik:1){# left binding
    mX = cbind( mDat[(1+ikk-iter):(itt+ikk-iter),ind$Y], mX)
  }
}
if( is.null(mX) ){
  return(list(success=0,msg="No explanatory variables in the model!"))
}
# order of the x_t' = ( y_{t-1}' , ... , y_{t-k}' , D_{t}' ), left binding!

mSt = NULL
for( iter in 1:ip ){
if(ind$S[[1]][iter,1] > 0){
  vSt = c(mDat[ (1+ikk-ind$S[[1]][iter,2]):(iT+ikk-ind$S[[1]][iter,2]) ,
    ind$S[[1]][iter,1] ])
  if(ind$S[[1]][iter,3] == 1) vSt = vSt/sd(vSt)
  mSt = cbind(mSt,vSt)
} else{
  mSt = cbind(mSt,rep(0,iT))
}
}
dim(mSt) = c(iT,ip)

# build lZ
lZ = list()
length(lZ) = ip
pick = NULL

for(peter in 1:ip){
  SS = c(mSt[,pter])
  if(all(SS==0)) next
  pick = c(pick,pter)
  mZ = NULL
  temp = mX
  if(econ==0){
    for(iter in 1:ord){
      for(jeter in 1:(ik*ip+iq)){
        XS = mX[,jeter]*SS
        if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ){
          temp = cbind(temp,XS)
          mZ = cbind(mZ,XS)
        }
      }
      SS = SS*SS
    }
  } else{
    for(jeter in 1:(ik*ip+iq)){
      XS = mX[,jeter]*SS
    }
  }
}
}
if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ){
    temp = cbind(temp,XS)
    mZ = cbind(mZ,XS)
}

if(ord>1){
    XS = SS
    for(jter in 2:ord){
        XS = XS * SS
        if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ){
            temp = cbind(temp,XS)
            mZ = cbind(mZ,XS)
        }
    }
    if(is.null(mZ)){
        return(list(success=0,msg="No LM test!"))
    }
}

lZ[[pter]] = mZ

RET = list()
length(RET) = ip+1

mU = svd(mX)$u
mE = mY - mU%*%t(mU)%*%mY

EE = mE[,pick]
dim(EE) = c(iT,length(pick))

# joint linearity test
RET[[ip+1]] = GESCORETEST(mE=EE, mX=mX, lZ=lZ[pick], flag=flag)

for(iter in 1:ip){
    if(is.null(lZ[[iter]])) next
    temp = mE[,iter]
    dim(temp) = c(iT,1)
    # linearity tests equation by equation using the same information
R code

# as the joint test
RET[[iter]] = SCORETEST(mE=temp, mX=mX, mZ=lZ[[iter]], flag=flag)
}

return( RET )

# Breusch-Pagan autocorrelation test for linear models
BPTEST <- function (mDat,ind,ik=0,iJ=5,flag=7)
{
if(flag==0)
return(list(success=0,msg="No test"))
if( ik < 0 )
return(list(success=0,msg="Invalid k"))
if( is.null(ind$Y) )
return(list(success=0,msg="No Yt"))
if( iJ<=0 )
return(list(success=0,msg="J must be positive"))

ip = length(ind$Y)

ikk = 0
ikk = max( ik, ikk )
# ikk is defined to be the biggest one of the ip+1 lags
iT = dim(mDat)[1] - ikk
# notice that the first ikk elements will be discarded.

# Build mY
mY = mDat[(1+ikk):(iT+ikk),ind$Y]
dim(mY) = c(iT,ip)
# data start from ikk+1 to ikk+iT (the original order in the raw data),
# iT elements in all.
# t = 1+ikk to iT+ikk hereafter.

# Build mX
mX = NULL
if( is.null(ind$D) ){

166
R code

iq = 0
} else{# mD: T by q
iq = length(ind$D)
mX = mDat[(1+ikk):(iT+ikk),ind$D]
dim(mX) = c(iT,iq)
# see mY
}

if( ik > 0 ){# left binding
for(iter in ik:1){
mX = cbind( mDat[(1+ikk-iter):(iT+ikk-iter),ind$Y], mX)
}
}

if( is.null(mX) ){return(list(success=0,msg="No explanatory variables in the model!"))
}

# order of the x_t' = ( y_t-1' , ... , y_t-k' , D_t' ), left binding!
mU = svd(mX)$u
mE = mY - mU%*%t(mU)%*%mY

temp = t(mE)
mZ = matrix(0,(iT-iJ),iJ*ip)
for(iter in 1:(iT-iJ)){
mZ[iter,] = c(temp[,,(iter+iJ-1):iter])
}

BP = SCORETEST(mE=mE[(iJ+1):iT,], mX=mX[(iJ+1):iT,], mZ=mZ, flag=flag)

return(BP)

}

Evaluation

# LM tests for evaluation

SCORETEST <- function(mE, mX, mZ, flag)
#mE, mX, mZ must be matrices!
#returns the LM test statistic and the degree of freedom
{

iT = dim(mE)[1]
ip = dim(mE)[2]
ix = dim(mX)[2]
iz = dim(mZ)[2]
iK = ix + iz
iDF = iz * ip

RSS0 = t(mE) %*% mE
mXX = cbind(mX, mZ)
mU = svd(mXX)$u
mR = mE - mU %*% t(mU) %*% mE
RSS1 = t(mR) %*% mR
R0 = svd(RSS0)$d
R1 = svd(RSS1)$d

LM = list()
FT = list()
WK = list()
RA = list()

if((flag%%2)==1){
# LM test
dTR = sum(diag(solve(RSS0) %*% RSS1))
test = iT*(ip-dTR)
LM$pval = 1-pchisq(test, df=iDF)
LM$test = test
LM$df = iDF
# rescale test
iDF1 = iDF
iDF2 = ip*(iT-iK)
test = LM$test * (iT-iK) / (iT*LM$df)
FT$pval = 1-pf(test, df1=iDF1, df2=iDF2)
FT$test = test
FT$df1 = iDF1
FT$df2 = iDF2
}
flag = flag%/%2
if((flag%%2)==1){
# Wilks test
Lambda = sum(log(R1))-sum(log(R0))
Lambda = Lambda * ( (ip+iz+1)*.5 + ix - iT )
WK$pval = 1-pchisq(Lambda,df=iDF)
WK$test = Lambda
WK$df = iDF
}
flag = flag%%2
if((flag%%2)==1){
# Rao’s test
iN = iT-ix-(ip+iz+1)*.5
is = sqrt( (iz*iz*ip*ip-4) / (ip*ip+iz*iz-5) )
iDF1 = iDF
iDF2 = iN*is - iz*ip*.5 + 1
RAO = exp((sum(log(R0))-sum(log(R1)))/is)-1
RAO = RAO *iDF2 /iDF1
RA$pval = 1-pf(RAO,df1=iDF1,df2=iDF2)
RA$test = RAO
RA$df1 = iDF1
RA$df2 = iDF2
}
return(list(LM = LM, FT = FT, WK = WK, RA = RA))
}

EXTEST <- function(mE, mX, mZ)
{
iT = dim(mE)[1]
ip = dim(mE)[2]
ix = dim(mX)[2]
iz = dim(mZ)[2]
iK = ix + iz
iDF = iz * ip

Omega = t(mE)%*%mE/iT
mU = svd(mX)$u
temp = t(mZ) %*% mZ - t(mZ) %*% mU %*% t(mU) %*% mZ
temp = t(mE) %*% mZ %*% solve(temp) %*% t(mZ) %*% mE %*% solve(Omega)
temp = sum(diag( temp ))

LM = list()
LM$pval = 1-pchisq(temp,df=iDF)
LM$test = temp
LM$df = iDF

return(LM)
}

DerGFunc <- function(Psit,BX,ip,im,ms,mgamma,mc)
# compute the current derivative of \( \Psi_t = \) 
# \( (I, G_t^{-1}(s_t|\gamma,c), \ldots) \)'
# input:
# Psit, mp by p, the current \( \Psi_t \)
# BX, mp by 1, vector, \( BX = t(mB)\%*\%Xt \)
# ms, p by m-1, vector of the current transition variable \( s_t \)
# mgamma, p by m-1
# mc, p by m-1
# output:
# \( d\Psi_t/ddelta, 2*p*p*(m-1) \) vector
{
    DPsit = Psit - Psit*Psit
    DPsit1 = DPsit \%*\% (ms-mc)
    DPsit2 = -DPsit \%*\% mgamma
    tDP = matrix(0,ip,(im*ip))
    RET = NULL
    for(iter in 1:(im-1)){
        for(jter in 1:ip){
            tDP[iter,(iter*ip+jter)] = DPsit1[(iter*ip+jter),iter]
            RET = c(RET, c(tDP\%*\%BX))
            tDP[iter,(iter*ip+jter)] = 0
            tDP[iter,(iter*ip+jter)] = DPsit2[(iter*ip+jter),iter]
            RET = c(RET, c(tDP\%*\%BX))
            tDP[iter,(iter*ip+jter)] = 0
        }
    }
    return(RET)
}

EVALTEST <- function(mDat,ind,im,ik,Est,iJ=1,SAN=NULL,ord=1,flag=7,fTest=c(1,1,1))
# input:
# mDat, (iT+ik)*n data matrix including Yt, Dt, and St
# ind: index list, ind$Y, ind$D, ind$S
# Notice that ind$S is a list of m-1 p*3 matrices, and the col
# element represents the lag from 0(no lag) to k,
# such that the St can be a laged variable. the third col is
# the norms for each row.
# im, number of regimes
# ik, lag length
# Est, a list consisting of the estimates
# beta_i, (pk+q) by p, parameters from 1 to m, phi included
# GPar, list of logistic function parameters
# SAN is a 3-dim vec, and the second element represents the lag
# from 0(no lag) to k,
# such that the St can be a laged variable.
# the third element is whether to normalize
# fTest: 1 for SC, 2 for AN, 3 for PC
# output:
# LM, rescaled, Wilks and Rao
{
  if( is.null(ind$Y) )
  return(list(success=0,msg="No Yt"))
  ip = length(ind$Y)

  ikk = 0
  if( im>1 ){
    if( is.null(ind$S) )
      return(list(success=0,msg="St shouldn't be NULL"))
    if( length(ind$S)!= im-1 )
      return(list(success=0,msg="Invalid dimension of St"))
    for( iter in 1:(im-1) ){
      if( !all(dim(ind$S[[iter]])==c(ip,3)) )
        return(list(success=0,msg="Invalid dimension of St"))
    }
    for( iter in 1:(im-1) ){
      ikk = max(ikk,max( ind$S[[iter]][,2] ))
    }
    if(!is.null(SAN)){
      ikk = max(ikk,max( SAN[2] ))
    }
  }
}
ikk = max( ik, ikk )
# ikk is defined to be the biggest one of the ip+1 lags
iT = dim(mDat)[1] - ikk
# notice that the first ikk elements will be discarded.

lBeta = Est$Beta

## PARAMETERS CHECKING
if( iT <= 0 )
  return(list(success=0,msg="Invalid T"))
if( ip <= 0 )
  return(list(success=0,msg="Invalid p"))
if( im <= 1 )
  return(list(success=0,msg="Invalid m"))
if( ik < 0 )
  return(list(success=0,msg="Invalid k"))

# Build mY
mY = mDat[(1+ikk):(iT+ikk),ind$Y]
dim(mY) = c(iT,ip)
# data start from ikk+1 to ikk+iT (the original order in
# the raw data), iT elements in all.
# t = 1+ikk to iT+ikk hereafter.

# Build mX
mX = NULL
if( is.null(ind$D) ){
  iq = 0
}else#{ mD: T by q
  iq = length(ind$D)
  mX = mDat[(1+ikk):(iT+ikk),ind$D]
  dim(mX) = c(iT,iq)
# see mY
}
if( ik > 0 ){
  for(iter in ik:1){# left binding
    mX = cbind( mDat[(1+ikk-iter):(iT+ikk-iter),ind$Y], mX)
  }
}

if( is.null(mX) ){
R code

return(list(success=0,msg="No explanatory variables in the model!"))
}
# order of the x_t' = ( y_t-1' , ... , y_t-k' , D_t' ), left binding!

# Build lSt ( consists of mSt's )
lSt = list()
for( mter in 1:(im-1) ){
mSt = NULL
for( iter in 1:ip ){
vSt = c(mDat[ (1+ikk-ind$S[[mter]][iter,2]):(iT+ikk-
        ind$S[[mter]][iter,2]) , ind$S[[mter]][iter,1] ])
if(ind$S[[mter]][iter,3] == 1) vSt = vSt/sd(vSt)
mSt = cbind(mSt,vSt)
}
dim(mSt) = c(iT,ip)
lSt[[mter]] = mSt
}

# construct mB = (beta_1, dbeta_2...), (kp+q) by mp
mB = matrix(0,(ip*ik+iq),im*ip)
for(pter in 1:ip){
    for(mter in 1:im){
        mB[,((mter-1)*ip+pter)] = lBeta[[pter]][,mter]
    }
}

# Build GPar
GPar = list(gamma=matrix(0,ip,im-1),c=matrix(0,ip,im-1))
for(iter in 1:ip){
    GPar$gamma[iter,] = Est$G[[iter]]
    GPar$c[iter,] = Est$C[[iter]]
}

eps = matrix(0,iT,ip)
Psi = array(0,dim=c(iT,im*ip,ip))
mK = NULL
for(iter in 1:iT){
    Xt = mX[iter,]
dim(Xt) = c((ip*ik+iq),1)
    ms = NULL

for(mter in 1:(im-1)) {
    ms = cbind(ms, c(lSt[[mter]][iter,]))
}
Psit = mGFunc(ip=ip, im=im, ms=ms, mgamma=GPar$gamma, mc=GPar$c)
Psit[, ,] = Psit
eps[iter,] = mY[iter,] - t(Psit) %*% t(mB) %*% Xt

temp = c(Psit)
dim(temp) = c(1, im*ip*ip)
Kt = c(Xt %*% temp)
Kt = c(Kt, DerGFunc(Psit=Psit, BX=t(mB) %*% Xt, ip=ip, im=im, ms=ms,
                      mgamma=GPar$gamma, mc=GPar$c))

mK = rbind(mK, Kt)
}

## serial correlation
SC = NULL
if((iT-iJ) >= (((ip*ik+iq)*im + 2*(im-1))*ip*ip) && fTest[1] == 1) {
    print("SC")
    temp = t(eps)
    mZ = matrix(0, (iT-iJ), iJ*ip)
    for(iter in 1:(iT-iJ)) {
        mZ[iter,] = c(temp[, (iter+iJ-1):iter])
    }
    mE = eps[(iJ+1):iT,]
    mXX = mK[(iJ+1):iT,]
    mU = svd(mXX)$u
    mE = mE - mU %*% t(mU) %*% mE
    SC = SCORETEST(mE=mE, mX=mXX, mZ=mZ, flag=flag)
}

## additional nonlinearity
AN = NULL
if(iT >= (((ip*ik+iq)*im + 2*(im-1))*ip*ip) && !is.null(SAN) && fTest[2] == 1) {
    print("AN")
    mZ = NULL
temp = mX
# Build SS
SS = c( mDat[ (1+ikk-SAN[2]):(iT+ikk-SAN[2]) , SAN[1] ] )
if(SAN[3] == 1) SS = SS/sd(SS)

for(iter in 1:ord){
  for(jter in 1:(ik*ip+iq)){
    XS = mX[,jter]*SS
    if( !any(colSums(temp-matrix(XS,iT,dim(temp)[2]))==0) ){
      temp = cbind(temp,XS)
mZ = cbind(mZ,XS)
    }
  }
  SS = SS*SS
}

mU = svd(mK)$u
mE = eps - mU%*%t(mU)%*%eps

AN = list()
length(AN) = ip+1
AN[[ip+1]] = SCORETEST(mE=mE, mX=mK, mZ=mZ, flag=flag)

for(iter in 1:ip){
  temp = mE[,iter]
dim(temp) = c(iT,1)
  # equation by equation using the same information
  # as the joint test
  AN[[iter]] = SCORETEST(mE=temp, mX=mK, mZ=mZ, flag=flag)
}

## parameter constancy
PC = NULL
if(iT>=(((ip*ik+iq)*im + 2*(im-1))*ip*ip) && fTest[3]==1){
  print("PC")
  mZ = NULL
  for(iter in 1:iT){
    temp = c(Psi[iter,,] %x% (mX[iter,]*iter))
mZ = rbind(mZ,temp)
  }
}
```r
mU = svd(mK)$u
mE = eps - mU%*%t(mU)%*%eps

PC = list()
length(PC) = ip+1
PC[[ip+1]] = SCORETEST(mE=mE, mX=mK, mZ=mZ, flag=flag)

for(iter in 1:ip){
temp = mE[,iter]
dim(temp) = c(iT,1)
  # equation by equation using the same information
  # as the joint test
  PC[[iter]] = SCORETEST(mE=temp, mX=mK, mZ=mZ, flag=flag)
}

# return(eps)
return(list(SC = SC, AN = AN, PC = PC))
}
```

---

Testing constancy of the error covariance matrix

# LM test against error covariance change
#

FLAG = list ( white = 1, arch = 2, garch = 4, smooth = 8 )

logistic <- function(vs, vg, vc)
# calculate the logistic function
{
  tmp = (vc-vs)*vg
tmp = 1/(exp(tmp)+1)
  return(tmp)
}

syminv <- function(mX)
# invert a symmetric matrix
{
176
R code

tmp = svd(mX)
return( tmp$v %*% diag(1/tmp$d) %*% t(tmp$u) )
}

AUXILIARY <- function(vU, mZ)
# run the auxiliary regression
#input:
# vU a vector of the residuals from the first regression
# mZ a matrix
#output:
# R^2
{
RSS0 = c(t(vU)%*%vU)
tmp = svd(mZ)$u
vE = vU - tmp%*%t(tmp)%*%vU
RSS1 = c(t(vE)%*%vE)
ret = (RSS0-RSS1)/RSS0

return(ret)
}

LMTEST <- function(mU,mX=NULL,iJ=1,vS=NULL,order=1,flag=FLAG$arch)
#input:
# mU T by p matrix containing the residuals from some model
# mX a matrix of X used in the White test, note that
# X may not contain the intercept
# iJ the lag used in the arch and garch
# vS a vector of transition variables used in the smooth transition test
# order of Taylor expansion, only for smooth transition
#output:
# LM tests
{
iT = dim(mU)[1]
ip = dim(mU)[2]

# construct mG matrix containing g_it
COV = t(mU)%*%mU/iT

R code

tmp = eigen(COV)
vL = tmp$values
mP = tmp$vectors

mW = mU%*%mP
mW2 = mW**2
mG = t(t(mW**2)/vL) - 1

# initialize the return value
LM = list( white=NULL, arch=NULL, garch=NULL, smooth=NULL )

# white specification
if( flag%%2==1 && !is.null(mX) ){  
    # construct vech(xx')
    mZ = NULL
    iq = dim(mX)[2]
    for(iter in 1:iT){
        vZ = c(mX[iter,])
        tmp = vZ%*%t(vZ)
        for(qter in 1:iq){
            vZ = c(vZ,tmp[qter:iq,qter])
        }
        mZ = rbind(mZ,vZ)
    }
    mZ = cbind(1,mZ)

    white = 0
    for(iter in 1:ip){
        # sum R^2
        white = white + AUXILIARY(vU=mG[,iter],mZ=mZ)
    }
    test = white * iT
    df = (dim(mZ)[2]-1)*ip
    LM$white = list(test=test,df=df,pval=1-pchisq(test,df=df))
}
flag = flag%/%2

# ARCH specification
if(flag%%2==1 & & iJ>0){
    df = 0
arch = 0
for(pter in 1:ip){
    vg = c(mG[(iJ+1):iT,pter])
    mZ = NULL
    for(jter in 1:iJ){
        mZ = cbind(mW2[jter:(iT-iJ+jter-1),pter],mZ)
    }
    mZ = cbind(1,mZ)
    arch = arch + AUXILIARY(vU=vg,mZ=mZ)
    df = df + dim(mZ)[2]-1
}
    test = arch * iT
    LM$arch = list(test=test, df=df,pval=1-pchisq(test,df=df))
flag = flag%/%2

    # GARCH specification
    if(flag%%2==1){
    }
    flag = flag%/%2

    # smooth transition specification
    if(flag%%2==1 && !is.null(vS)){
        # construct mZ for transition variables
        mZ = 1
        for(iter in 1:order){
            mZ = cbind(mZ,vS**iter)
        }
        smooth = 0
        for(iter in 1:ip){
            smooth = smooth + AUXILIARY(vU=mG[,iter],mZ=mZ)
        }
        test = smooth * iT
        df = ip*order
        LM$smooth = list(test=test,df=df,pval=1-pchisq(test,df=df))
    }
return (LM)
}

BETT <- function(mU)
{
iT = dim(mU)[1]
ip = dim(mU)[2]

# construct mG matrix containing g_it
COV = t(mU) %*% mU / iT

mD = diag(sqrt(diag(COV)))
iD = solve(mD)
mP = cor(mU)
iP = solve(mP)

tao = 1:iT / iT

dPdth = matrix(0, 2*ip + ip*(ip-1)/2, ip*ip)
rter = ip*2+1
for(jter in 1:(ip-1)){
    for(kter in (jter+1):ip){
        dPdth[rter,(jter-1)*ip+kter] = 1
        dPdth[rter,(kter-1)*ip+jter] = 1
        rter = rter+1
    }
}

tmp5 = dPdth %*% (.5*( iP%x%iP )) %*% t(dPdth)

vS = 0
mI = 0

for(iter in 1:iT){
    ve = mU[iter,]
dim(ve) = c(ip,1)
ee = ve%*%t(ve)
diDdth = matrix(0,2*ip+ip*(ip-1)/2,ip*ip)
for(jter in 1:ip){
  cter = (jter-1)*ip+1
  rter = (jter-1)*2+1
  diDdth[rter,cter] = -.5*mD[jter,jter]**3
  rter = rter+1
  diDdth[rter,cter] = -.5*mD[jter,jter]**3*tao[iter]
}

tmp1 = diDdth %*% c( mD - .5*ee%*%iD%*%iP - .5*iP%*%iD%*%ee )
tmp2 = dPdth %*% c( iP%*%iD%*%ee%*%iD%*%iP - iP )*.5
vS = vS + tmp1 + tmp2

tmp3 = diDdth %*% (mD%x%mD + .5*(iP%x%COV + COV%x%iP)) %*% t(diDdth)
tmp4 = .5*( mD%x%iP + iP%x%mD )
tmp4 = diDdth %*% tmp4 %*% t(dPdth) + dPdth %*% tmp4 %*% t(diDdth)

mI = mI + tmp3 - tmp4
}
vS = vS/iT
mI = mI/iT + tmp5

test = iT*c(t(vS)%*%syminv(mI)%*%vS)
df = ip

return(list(test=test,df=df,pval=1-pchisq(test,df=df)))
}
PhD Theses since 1 July 2011

2011-4  Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics

2011-5  Christian Bach: The Game of Risk

2011-6  Stefan Holst Bache: Quantile Regression: Three Econometric Studies

2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management

2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization

2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders

2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade

2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting

2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping

2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing

2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing

2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies

2012-6 Miriam Wüst: Essays on early investments in child health

2012-7 Yukai Yang: Modelling Nonlinear Vector Economic Time Series