View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DIAL UCLouvain

Digital access to libraries

"Linearity and misspecification tests for vector smooth transition regression models"

Terasvirta, Timo ; Yang, Yukai

Abstract

In this paper, we derive Lagrange multiplier and Lagrange multiplier type specification and misspecification tests for vector smooth transition models. We report results from simulation studies in which the size and power properties of the proposed tests in small samples are considered. The results show that these asymptotic tests generally suffer from size distortion. We find that Wilks's Λ and Rao's F statistic both have satisfactory size properties and can be recommended for empirical use. Bootstrapping the standard asymptotic LM statistic offers another solution to the problem.

Document type : *Document de travail (Working Paper)*

Référence bibliographique

Terasvirta, Timo ; Yang, Yukai. *Linearity and misspecification tests for vector smooth transition regression models.* CORE Discussion Papers ; 2014/61 (2014) 38 pages

Available at:
http://hdl.handle.net/2078.1/152767

[Downloaded 2019/04/19 at 09:05:16]

2014/61

■

Linearity and misspecification tests for vector smooth transition regression models

Timo Teräsvirta and Yukai Yang

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE
Voie du Roman Pays 34, L1.03.01
B-1348 Louvain-la-Neuve, Belgium.
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
E-mail: corestat-library@uclouvain.be
http://www.uclouvain.be/en-44508.html

CORE DISCUSSION PAPER
2014/61

Linearity and misspecification tests for vector smooth transition regression models

Timo Teräsvirta[*] and Yukai Yang[**]

[*]CREATES, Aarhus University[1]
[**]CORE, Université catholique de Louvain and CREATES, Aarhus University[2]

First version: February 2014, revised version: November 2014

Abstract

In this paper, we derive Lagrange multiplier and Lagrange multiplier type specification and misspecification tests for vector smooth transition models. We report results from simulation studies in which the size and power properties of the proposed tests in small samples are considered. The results show that these asymptotic tests generally suffer from size distortion. We find that Wilks's Λ and Rao's F statistic both have satisfactory size properties and can be recommended for empirical use. Bootstrapping the standard asymptotic LM statistic offers another solution to the problem.

Keywords: Vector STAR models; Linearity test; Misspecification test; Vector nonlinear time series; Serial correlation; Parameter constancy; Residual nonlinearity test.
JEL Classification: C12, C32, C52

Acknowledgements: This research has been supported by CREATES, funded by the Danish National Research Foundation, Grant Number DNFR78. Material from this paper has been presented at the Second International Symposium in Computational Economics and Finance, Tunis, March 2012, Nonlinear and Asymmetric Models in Applied Economics, Paris, April 2012, Theory and Applications Workshop, Mannheim, December 2012, Applicable Semiparametrics Conference, Humboldt- Universität zu Berlin, and Macromodels 2013, Warsaw, both in October 2013, and in seminars at Bank of Finland, Helsinki, and DIW Berlin. Comments from participants of these occasions are gratefully acknowledged. We also wish to thank Rickard Sandberg and Dick van Dijk for many constructive remarks. Responsibility for any errors or shortcomings in this work remains ours.

[1]Timo Teräsvirta, email: tterasvirta@creates.au.dk
[2]Yukai Yang, email: yukai.yang@uclouvain.be

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 logistic vector smooth transition autoregressive (LVSTAR) or regression (LVSTR) models. Furthermore, we derive various misspecification tests for estimated LVSTAR or LVSTR 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 tests 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 LVSTR 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 LVSTR 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 for all equations, but the set of transition variables is known. This means that it is known which variable belongs to which smooth transition equation. In some applications, there may exist underlying theory propositions determining these transition variables and thus justifying this type of test.

When the LVSTR 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), Edgerton and Shukur (1999) and Shukur and Edgerton (2002). Edgerton and Shukur (1999) 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), the test was the functional form specification test, RESET by Ramsey (1969), and the conclusion was

similar. 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, 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 designed and applied in the LVSTAR or LVSTR modelling framework of Teräsvirta and Yang (2014). For a recent survey of Vector Threshold Autoregressive (VTAR) and LVSTAR models, see Hubrich and Teräsvirta (2013).

The plan of the paper is as follows. In Section 2, we first introduce the LVSTAR model developed in Teräsvirta and Yang (2014). We then develop linearity tests for two cases. First, this is done when under the alternative a single transition variable controls transitions. Second, tests are derived for a situation in which each equation has its own transition function and transition variable. In Section 3, we discuss the size distortion problem and propose test statistics that alleviate it. In Section 4, several misspecification tests for model evaluation are derived. In Section 5, we carry out simulation experiments to investigate the size properties of the tests and report the results. Section 6 concludes.

2 The joint linearity tests

2.1 The logistic vector smooth transition model

In Teräsvirta and Yang (2014), we define the Logistic Vector Smooth Transition AutoRegressive (LVSTAR) model with k lags as follows:

$$\mathbf{y}_t = \{\sum_{i=1}^{m}(\mathbf{G}_t^{i-1} - \mathbf{G}_t^i)\mathbf{F}_i'\}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\mathbf{y}_t$ is a $p\times 1$ column vector, $\mathbf{x}_t = (\mathbf{y}_{t-1}', ..., \mathbf{y}_{t-k}', \mathbf{d}_t')'$ is a $(kp+q)\times 1$ vector in which $\mathbf{d}_t$ is a $q \times 1$ vector of intercept and any exogenous variables. $\mathbf{F}_i$ is a $(kp+q)\times p$ matrix of linear parameters: $\mathbf{F}_i = (\mathbf{A}_{i1}', ..., \mathbf{A}_{ik}', \boldsymbol{\Phi}_i')'$, where each $\mathbf{A}_{ij}$, $i = 1, ..., m$, $j = 1, ..., k$, is a $p \times p$ matrix, and each $\boldsymbol{\Phi}_i$, $i = 1, ..., m$, is a $q \times p$ matrix. The $p \times 1$ error vector $\boldsymbol{\varepsilon}_t$ is assumed *i.i.d.* $\mathcal{N}(0, \boldsymbol{\Omega})$, where the covariance matrix $\boldsymbol{\Omega}$ is positive definite.

$\mathbf{G}_t^i$ is a $p \times p$ diagonal matrix of transition functions which takes the form:

$$\mathbf{G}_t^i = \text{diag}\left\{\ g(s_{1t}|\gamma_{i1}, c_{i1}),\ \ ...,\ \ g(s_{pt}|\gamma_{ip}, c_{ip})\ \right\}, \tag{2.2}$$

for $i = 1, ..., m-1$, and $\mathbf{G}_t^0 = \mathbf{I}_p$, $\mathbf{G}_t^m = \mathbf{0}$. The diagonal elements of $\mathbf{G}_t^i$ in (2.2) are logistic functions of their transition variables:

$$g(s_{jt}|\gamma_{ij}, c_{ij}) = (1 + \exp\{-\gamma_{ij}\,(s_{jt} - c_{ij})\})^{-1},\ \gamma_{ij} > 0, \tag{2.3}$$

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

We consider the LVSTAR model (2.1) using the following reparametrization:

$$\mathbf{y}_t = (\mathbf{B}_1' + \mathbf{G}_t^1\mathbf{B}_2' + ... + \mathbf{G}_t^{m-1}\mathbf{B}_m')\mathbf{x}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\Psi}_t'\mathbf{B}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{2.4}$$

where $\boldsymbol{\Psi}_t = \left(\mathbf{I}_p, \mathbf{G}_t^1, ..., \mathbf{G}_t^{m-1}\right)'$ is an $mp \times p$ full rank matrix. Furthermore, $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_m)$ is a $(kp+q)\times mp$ matrix, where $\mathbf{B}_1 = \mathbf{F}_1$, and $\mathbf{B}_i = \mathbf{F}_i - \mathbf{F}_{i-1}$, $i = 2, ..., m$.

As $\boldsymbol{\varepsilon}_t \sim\ i.i.d.\,\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, the conditional log-likelihood function takes the form

$$\begin{aligned} \log L(\theta) \;=\; & -(Tp/2)\log 2\pi - (T/2)\log|\boldsymbol{\Omega}| \\ & -(1/2)\sum_{t=1}^{T}(\mathbf{y}_t - \boldsymbol{\Psi}_t'\mathbf{B}'\mathbf{x}_t)'\,\boldsymbol{\Omega}^{-1}\,(\mathbf{y}_t - \boldsymbol{\Psi}_t'\mathbf{B}'\mathbf{x}_t)\,. \end{aligned} \tag{2.5}$$

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

Moreover, we introduce the following regularity conditions given by Feigin (1976):

Assumption 2.1. *The log-likelihood function* $\log L(\theta)$ *is second-order differentiable with respect to the parameters in* θ*.*

Assumption 2.2. *Both the first-order and second-order derivatives of the log-likelihood function* (2.5) *with respect to the parameters* $\theta \in \boldsymbol{\Theta}$ *are* L_T*-integrable.*

Assumption 2.3. *We assume that the following limiting information exists*

$$\lim_{T\to\infty} T^{-1}E_\theta(\mathbf{i}_T(\theta)) = \mathbf{i}(\theta) < \infty \tag{2.6}$$

where $\mathbf{i}_T(\theta) = -\partial^2 \log L/\partial\theta^2$*.*

Assumptions 2.1-3 guarantee the existence of the moments of the score vector and the information matrix, and are needed in proving asymptotic normality of the score.

2.2 Testing linearity against LVSTAR with single transition variable

Throughout the paper, we only consider the case in which the diagonal elements of the transition matrix $\mathbf{G}_t^i$ 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 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 only a single transition function for the whole system.

Consider the p-dimensional single transition logistic VSTAR model in (2.4):

$$\mathbf{y}_t = \mathbf{B}_1'\mathbf{x}_t + \mathbf{G}_t\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \tag{2.7}$$

where the sequence $\{\boldsymbol{\varepsilon}_t\}$ is $i.i.d.$ $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. The null hypothesis of linearity can be written as $H_0 : \gamma_j = 0$, $j = 1, ..., p$. When the null holds, $\mathbf{G}_t \equiv (1/2)\mathbf{I}_p$ and (2.7) becomes linear, while the location parameters c_j in the logistic functions and the parameters in the linear combination $\mathbf{B}_1 + (1/2)\mathbf{B}_2$ 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 an n-order Taylor approximation around $\gamma_j = 0$ as proposed in Luukkonen et al. (1988). This gives

$$g(s_t|\gamma_j, c_j) = \sum_{i=0}^{n} a_{j,n-i} s_t^{n-i} + r_{jt}$$

where $a_{j,0}, ..., a_{j,n}$ are the coefficients, and r_{jt} is the remainder term of the Taylor expansion. We rewrite $\mathbf{G}_t$ in (2.7) as follows:

$$\begin{aligned} \mathbf{G}_t &= \operatorname{diag}\left\{ \sum_{i=0}^{n} a_{1,n-i} s_t^{n-i} + r_{1t},\ ...,\ \sum_{i=0}^{n} a_{p,n-i} s_t^{n-i} + r_{pt} \right\} \\ &= \sum_{i=0}^{n} \mathbf{A}_{n-i} s_t^{n-i} + \mathbf{R}_t \end{aligned} \tag{2.8}$$

where $\mathbf{A}_{n-i} = \operatorname{diag}(a_{1,n-i}, ..., a_{p,n-i})$, and $\mathbf{R}_t = \operatorname{diag}(r_{1t}, ..., r_{pt})$

Inserting (2.8) into (2.7) yields:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{B}_1'\mathbf{x}_t + \left(\sum_{i=0}^{n}\mathbf{A}_{n-i}s_t^{n-i} + \mathbf{R}_t\right)\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ &= (\mathbf{B}_1' + \mathbf{A}_0\mathbf{B}_2')\,\mathbf{x}_t + \sum_{i=1}^{n}\mathbf{A}_i\mathbf{B}_2'\mathbf{x}_t s_t^i + \mathbf{R}_t\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\Theta}_0'\mathbf{x}_t + \sum_{i=1}^{n}\boldsymbol{\Theta}_i'\mathbf{x}_t s_t^i + \boldsymbol{\varepsilon}_t^* \qquad (2.9)\end{aligned}$$

where $\boldsymbol{\Theta}_0 = \mathbf{B}_1 + \mathbf{B}_2\mathbf{A}_0$, $\boldsymbol{\Theta}_i = \mathbf{B}_2\mathbf{A}_i$, and $\boldsymbol{\varepsilon}_t^* = \mathbf{R}_t\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t$. The null hypothesis implies $\mathbf{A}_i = \mathbf{0}$ for $i = 1, ..., n$, $\mathbf{A}_0 = \frac{1}{2}\mathbf{I}_p$, and $\mathbf{R}_t = \mathbf{0}$ in (2.8). Thus, under the null hypothesis, the auxiliary VAR model (2.9) is linear, with $\boldsymbol{\Theta}_0 = \mathbf{B}_1 + \frac{1}{2}\mathbf{B}_2$ and $\boldsymbol{\Theta}_1 = ... = \boldsymbol{\Theta}_n = \mathbf{0}$. The linearity hypothesis is therefore $H_0 : \boldsymbol{\Theta}_1 = ... = \boldsymbol{\Theta}_n = \mathbf{0}$ in (2.9). Moreover, due to the fact that $\mathbf{R}_t = \mathbf{0}$ under the null hypothesis, the error term $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t$. Since the Lagrange multiplier test only requires estimating the model under the null hypothesis, the remainder term does not affect the normality of the errors or the standard asymptotic inference.

Denoting $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_T)'$, $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T)'$, $\mathbf{E}^* = (\boldsymbol{\varepsilon}_1^*, ..., \boldsymbol{\varepsilon}_T^*)'$, $\boldsymbol{\Theta}_n = (\boldsymbol{\Theta}_1', ..., \boldsymbol{\Theta}_n')'$, and

$$\mathbf{Z}_n = \begin{bmatrix} \mathbf{x}_1's_1 & \mathbf{x}_1's_1^2 & \dots & \mathbf{x}_1's_1^n \\ \mathbf{x}_2's_2 & \mathbf{x}_2's_2^2 & \dots & \mathbf{x}_2's_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_T's_T & \mathbf{x}_T's_T^2 & \dots & \mathbf{x}_T's_T^n \end{bmatrix},$$

(2.9) can be rewritten into matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Theta}_0 + \mathbf{Z}_n\boldsymbol{\Theta}_n + \mathbf{E}^*. \qquad (2.10)$$

The null hypothesis is $H_0 : \boldsymbol{\Theta}_n = \mathbf{0}$. The subscript in $\mathbf{Z}$ and $\boldsymbol{\Theta}$ indicates the order of the corresponding Taylor expansion of the transition function.

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

$$\frac{\partial \log L(\tilde{\theta})}{\partial \boldsymbol{\Theta}_n} = \mathbf{Z}_n'\left(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\Theta}}_0\right)\tilde{\boldsymbol{\Omega}}^{-1} \qquad (2.11)$$

where $\tilde{\boldsymbol{\Theta}}_0$ and $\tilde{\boldsymbol{\Omega}}$ are estimates from the restricted model (under the null hypothesis). Under regularity conditions and assuming that $E\mathbf{Z}_n'\mathbf{Z}_n$ exists, the score approaches a matrix variate normal distribution with zero mean and variance $\mathbf{Z}_n'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z}_n \otimes \boldsymbol{\Omega}^{-1}$, where $\mathbf{P}_x \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the limiting projection matrix as $T \to \infty$.

Theorem 2.4. *Consider the model* (2.7) *and its approximation* (2.10)*. The LM test statistic for testing the null hypothesis* $H_0 : \gamma_j = 0$*,* $j = 1, ..., p$ *in* (2.7)*, or equivalently,*

$H_0 : \mathbf{\Theta}_n = \mathbf{0}$ *in* (2.10)*, equals:*

$$LM_n = \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{\Theta}}_0)'\mathbf{Z}_n \left[\mathbf{Z}_n'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z}_n\right]^{-1} \mathbf{Z}_n'(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{\Theta}}_0)\}. \tag{2.12}$$

Under the null hypothesis, the test statistic has an asymptotic χ^2 *distribution with* $np(kp+q)$ *degrees of freedom.*

Proof. See Appendix A. □

The explanatory variable vector $\mathbf{x}_t$ may contain the intercept, seasonal dummies, trend and other deterministic terms. In the case that $\mathbf{x}_t$ 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 $\mathbf{Z}_n$ to avoid perfect collinearity. The number of degrees of freedom equals the number of restrictions, i.e., p multiplied by the column dimension of $\mathbf{Z}_1$.

The test can also be performed as follows:

1. Estimate the restricted model: regress $\mathbf{Y}$ on $\mathbf{X}$. Collect the residuals $\tilde{\mathbf{E}} = (\mathbf{I}_T - \mathbf{P}_x)\mathbf{Y}$, and the matrix residual sum of squares $\mathbf{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}}$.

2. Run an auxiliary regression of $\tilde{\mathbf{E}}$ on $(\mathbf{X}, \mathbf{Z}_n)$. Collect the residuals $\tilde{\mathbf{\Xi}}$, and compute the matrix residual sum of squares $\mathbf{RSS}_1 = \tilde{\mathbf{\Xi}}'\tilde{\mathbf{\Xi}}$.

3. Compute the test statistic

$$\begin{aligned} LM_n &= T\,\text{tr}\{\mathbf{RSS}_0^{-1}(\mathbf{RSS}_0 - \mathbf{RSS}_1)\} \\ &= T(p - \text{tr}\{\mathbf{RSS}_0^{-1}\mathbf{RSS}_1\}). \end{aligned} \tag{2.13}$$

The joint test statistic defined in Theorem 2.4 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.

The choice of the order of the Taylor expansion n is somewhat arbitrary. A higher order will increase the column dimension of $\mathbf{Z}_n$. But then, rejecting the null hypothesis may become easier, since a higher order often increases the power of the test. On the other hand, a lower order, for example $n = 1$, normally leads to a test with better size properties, because it uses fewer parameters than a higher-order test. However, it may suffer from the problem, pointed out by Luukkonen et al. (1988), that in the situations where $s_t = y_{t-d,j}$ for some $1 \leq d \leq k$, $1 \leq j \leq p$ (the self-exiting STAR model), and only the intercept fluctuates across regimes, the LM_1 statistic only has trivial power against the alternative. They suggested to solve this problem by choosing $n = 3$ instead of $n = 1$.

Although the high-order test statistic LM_n, for example $n = 3$, is mainly designed for the particular case where only the intercept differs across regimes, it can be used for more

general alternatives. A parsimonious version of the LM_n 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 LM_1 with regressors from s_t^2 up to s_t^n. Thus, we have the parsimonious n-order LM test, which is denoted as LM_n^e

$$LM_n^e = \mathrm{tr}\{\tilde{\mathbf{\Omega}}^{-1}(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{\Theta}}_0)'\mathbf{Z}_n^e \left[\mathbf{Z}_n^{e\prime}(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z}_n^{e\prime}\right]^{-1} \mathbf{Z}_n^{e\prime}(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{\Theta}}_0)\},$$

where

$$\mathbf{Z}_n^e = \begin{bmatrix} \mathbf{x}_1' s_1 & s_1^2 & \dots & s_1^n \\ \mathbf{x}_2' s_2 & s_2^2 & \dots & s_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_T' s_T & s_T^2 & \dots & s_T^n \end{bmatrix}.$$

The test statistic has an asymptotic χ^2 distribution and its number of degrees of freedom under the null is $p(kp + q + n - 1)$. The parsimonious test LM_n^e reduces the column dimension of $\mathbf{Z}_n$, and is likely to have better finite-sample properties than LM_n.

2.3 Testing linearity against LVSTAR with different transition variables

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

When the theory behind the model does not lead to a single potential transition variable, linearity may be tested using a set of candidate variables. This is done by testing against the alternative models defined by the candidate variables one by one. The simple rule, discussed in Teräsvirta (1994) for the univariate STAR case, suggests selecting the alternative with the transition variable that leads to the strongest rejection, measured by the p-value, of the null hypothesis. This is, however, just a guideline. It may well be that several transition variables lead to strong rejections with p-values of the same magnitude. In that case it is advisable to fit all of these alternative models to the data and leave the decision of selecting the final model to the model evaluation stage. It may turn out that a linear combination of some of the candidate variables (preferably with known weights) would make a reasonable transition variable. Model misspecification tests useful in evaluating LVSTAR models are discussed in Section 4.

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

$$\mathbf{y}_t = \mathbf{\Theta}_0'\mathbf{x}_t + \sum_{i=1}^{n} \mathbf{S}_t^i \mathbf{\Theta}_i'\mathbf{x}_t + \boldsymbol{\varepsilon}_t^*, \tag{2.14}$$

where $\mathbf{S}_t = \text{diag}(s_{1t},\ s_{2t},\ ...,\ s_{pt})$, $\boldsymbol{\Theta}_i$ for $i = 0, ..., n$, and $\boldsymbol{\varepsilon}_t^*$ are defined as in (2.9). This nests the case where $s_{kt} = s_{lt}$ for $k \neq l$.

The null hypothesis of linearity, that is, $\gamma_{ij} = 0$ for all $i = 1, ..., m-1$ and $j = 1, ..., p$, can be equivalently stated as $H_0 : \boldsymbol{\Theta}_1 = ... = \boldsymbol{\Theta}_n = \mathbf{0}$. Then we have an LM test similar to that in (2.12), with

$$\mathbf{Z}_n^s = \begin{pmatrix} \text{vec}(\mathbf{S}_1 \otimes \mathbf{x}_1)' & \text{vec}(\mathbf{S}_1^2 \otimes \mathbf{x}_1)' & \dots & \text{vec}(\mathbf{S}_1^n \otimes \mathbf{x}_1)' \\ \vdots & \vdots & \ddots & \vdots \\ \text{vec}(\mathbf{S}_T \otimes \mathbf{x}_T)' & \text{vec}(\mathbf{S}_T^2 \otimes \mathbf{x}_T)' & \dots & \text{vec}(\mathbf{S}_T^n \otimes \mathbf{x}_T)' \end{pmatrix}. \tag{2.15}$$

One can carry out this test by applying the testing procedure in the previous section, replacing $\mathbf{Z}_n$ by $\mathbf{Z}_n^s$.

The test suffers from the drawback that the dimension of the null hypothesis grows very fast, which has an adverse effect on the empirical size of the test. In practice we set $n = 1$ and $\mathbf{Z}_1^s = (\text{vec}(\mathbf{S}_1 \otimes \mathbf{x}_1), ..., \text{vec}(\mathbf{S}_T \otimes \mathbf{x}_T))'$ in most cases. The size distortion caused by a large-dimensional null hypothesis will be discussed in Section 3.

As already mentioned, some of the transition variables can still be the same, that is, $s_{kt} = s_{lt}$ for some $k \neq l$. This implies that the matrix $\mathbf{Z}_1^s$ has reduced rank and some of its columns must be removed to avoid collinearity. In the following, we propose a parsimonious "algorithm" for carrying out the test:

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

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

3. Compute the test statistic in (2.13).

Unfortunately, the limiting null distribution of this LM test 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 that is valid assuming that the errors are independent and $E\mathbf{Z}_n^{s\prime}\mathbf{Z}_n^s < \infty$. This works as follows. First, estimate the restricted model under the null hypothesis. Then generate $\mathbf{y}_t$ using $\mathbf{x}_t$, the estimates of the parameters from the restricted model and the bootstrapped residuals, and compute the test statistic using the parsimonious algorithm suggested above. Repeat the sampling procedure and the computation of the test statistic N times. This gives N values of the test statistics generated under the null hypothesis. After ordering them, select the relevant quantile to be the critical value for a given significance level.

Table 1: The empirical size of the bootstrapping test.

	$\rho = 0.9$			$\rho = 0$			$\rho = -0.9$		
Nom. size	1%	5%	10%	1%	5%	10%	1%	5%	10%
$T = 50$	1.35%	5.90%	10.90%	0.90%	5.00%	9.65%	1.10%	4.50%	8.50%
$T = 100$	1.05%	5.15%	9.85%	1.25%	4.95%	10%	0.9%	5.35%	10.8%
$T = 200$	1.00%	5.25%	10.40%	0.80%	4.65%	9.55%	0.9%	4.95%	9.20%
$T = 500$	1.05%	5.60%	10.95%	1.35%	5.60%	9.85%	1.40%	5.25%	9.55%

To see how this works, we conduct a simple simulation-based size experiment. We simulate the following bivariate data generating process:

$$y_{it} = 0.8y_{i,t-1} + \varepsilon_{it} \tag{2.16}$$

where $i = 1, 2$, and let $T = 50$, 100, 200, 500. The covariance matrix of the errors is

$$\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{2.17}$$

where $\rho = 0.9$, 0, or -0.9. We choose $N = 10000$ and repeat the bootstrapped test 2000 times. The results are given in Table 1. We see that they are quite satisfactory, already for $T = 50$.

3 Linearity test statistic with improved size

In small samples, the test introduced in Section 2.2 is 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 $\mathbf{y}_t$, is high compared to T. 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.

Bartlett and Bartlett-type corrections have been widely used as a possible remedy to the size problem of LM-type tests. We first mention the Laitinen-Meisner 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

$$F = \frac{(pT - K)}{G \cdot pT} LM \tag{3.1}$$

where LM represents any of the three tests LM_1, LM_3 and LM_3^e 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) showed that the Laitinen-Meisner correction is likely to overcorrect the size. We consider 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 3.1. *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:*

$$\mathbf{W}_1 \sim \mathcal{W}_p\Big(\mathbf{\Omega}, \mathrm{cd}(\mathbf{Z})\Big) \quad \mathbf{W}_2 \sim \mathcal{W}_p\Big(\mathbf{\Omega}, T - \mathrm{cd}(\mathbf{X}) - \mathrm{cd}(\mathbf{Z})\Big) \tag{3.2}$$

where $\mathrm{cd}(\cdot)$ *is the column dimension of a matrix.*

Proof. See Appendix C. □

Matrix $\mathbf{Z}$ in (3.2) can be any of the three matrices $\mathbf{Z}_1$, $\mathbf{Z}_3$ and $\mathbf{Z}_3^e$, and $\mathbf{\Omega}$ 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 χ^2-distributed, which implies an F-test.

We define Wilks's Λ-distribution as follows:

Definition 3.2. *When* $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, m)$ *and* $\mathbf{B} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ *are independent,* $\mathbf{\Sigma}$ *is a* $p \times p$ *positive definite matrix,* $m \geq p$,

$$\Lambda = |\mathbf{A}|/|\mathbf{A} + \mathbf{B}| = |\mathbf{I}_p + \mathbf{A}^{-1}\mathbf{B}|^{-1} \sim \mathcal{L}(p, m, n) \tag{3.3}$$

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 $\mathbf{\Sigma}$. Wilks's Λ statistic has the following form:

$$\Lambda = |\mathbf{W}_2|/|\mathbf{W}_2 + \mathbf{W}_1| = |\mathbf{RSS}_1|/|\mathbf{RSS}_0|. \tag{3.4}$$

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

$$\lambda = \Big(\frac{1}{2}(p + \mathrm{cd}(\mathbf{Z}) + 1) + \mathrm{cd}(\mathbf{X}) - T\Big) \log \Lambda \sim \chi^2_{\mathrm{cd}(\mathbf{Z})p}, \tag{3.5}$$

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 2.2 but computing the value of the test statistic defined in (3.4) and (3.5) instead of step 3.

Rao (1951; 1965, Section 8c.5) defined 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:

$$F_{Rao} = \left[\left(\frac{|\mathbf{RSS}_0|}{|\mathbf{RSS}_1|} \right)^{1/s} - 1 \right] \frac{Ns - (1/2)\mathrm{cd}(\mathbf{Z})p + 1}{\mathrm{cd}(\mathbf{Z})p} \tag{3.6}$$

where

$$s = \left(\frac{\mathrm{cd}^2(\mathbf{Z})p^2 - 4}{p^2 + \mathrm{cd}^2(\mathbf{Z}) - 5} \right)^{1/2}, \; N = T - \mathrm{cd}(\mathbf{X}) - (1/2)\left(p + \mathrm{cd}(\mathbf{Z}) + 1\right).$$

The corresponding degrees of freedom of (3.6) are $\mathrm{cd}(\mathbf{Z})p$ and $Ns - (1/2)\mathrm{cd}(\mathbf{Z})p + 1$. Similarly to Wilks's Λ, the test can be carried out by performing steps 1 and 2 outlined in the algorithm in Section 2.2, and then computing the value (3.6). When $p = 1$ and $s = 1$, (3.6) becomes the F-type LM test.

4 Evaluation tests

The evaluation stage of the LVSTAR modelling strategy designed and applied in Teräsvirta and Yang (2014) makes use of misspecification tests of the estimated LVSTAR 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 Λ and Rao's F statistic considered in Section 3, can be applied even here to alleviate the size distortion problem.

4.1 Serial correlation in the error process

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

$$\mathbf{y}_t = \{\sum_{i=1}^{m} (\mathbf{G}_t^{i-1} - \mathbf{G}_t^i)\mathbf{F}_i'\}\mathbf{x}_t + \mathbf{u}_t = \mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t + \mathbf{u}_t, \tag{4.1}$$

where

$$\mathbf{u}_t = \sum_{i=1}^{J} \mathbf{P}_i'\mathbf{u}_{t-i} + \boldsymbol{\varepsilon}_t = \mathbf{P}(L)\mathbf{u}_t + \boldsymbol{\varepsilon}_t. \tag{4.2}$$

In (4.2), $\mathbf{P}_i$ is a $p \times p$ matrix, $\mathbf{P}(L) = \sum_{i=1}^{J} \mathbf{P}_i' L^i$, L is the lag operator, J is the lag length, and $\boldsymbol{\varepsilon}_t \sim i.i.d.\mathcal{N}(0, \boldsymbol{\Omega})$ is a $p \times 1$ vector. We assume that the roots of the polynomial $|\mathbf{I}_p - \sum_{i=1}^{J} \mathbf{P}_i' z^i|$ lie outside the unit circle. Furthermore, we assume that the sequence of $\{\mathbf{y}_t\}$ is stationary and ergodic such that the parameters can be estimated consistently under the null hypothesis of no serial correlation $\{\mathbf{u}_t\}$, that is, when

$$H_0 : \mathbf{P}_1 = \mathbf{P}_2 = ... = \mathbf{P}_J = \mathbf{0}$$

holds. This is a high-level assumption, as general asymptotic theory for LVSTAR models does not exist, for stability (but not consistency) results, see Saikkonen (2008). Left-multiplying (4.1) by $\mathbf{I}_p - \mathbf{P}(L)$ yields

$$\begin{aligned} \mathbf{y}_t &= \mathbf{P}(L)(\mathbf{y}_t - \boldsymbol{\Psi}_t' \mathbf{B}' \mathbf{x}_t) + \boldsymbol{\Psi}_t' \mathbf{B}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ &= \mathbf{P}' \mathbf{z}_t + \boldsymbol{\Psi}_t' \mathbf{B}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{4.3}$$

see Godfrey (1988, Section 4.4), where

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_J \end{bmatrix}, \quad \mathbf{z}_t = \begin{bmatrix} \mathbf{y}_{t-1} - \boldsymbol{\Psi}_{t-1}' \mathbf{B}' \mathbf{x}_{t-1} \\ \mathbf{y}_{t-2} - \boldsymbol{\Psi}_{t-2}' \mathbf{B}' \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{y}_{t-J} - \boldsymbol{\Psi}_{t-J}' \mathbf{B}' \mathbf{x}_{t-J} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \\ \vdots \\ \mathbf{u}_{t-J} \end{bmatrix}.$$

We have the following log-likelihood function:

$$\begin{aligned} \log L &= -((T-J)p/2) \log 2\pi - ((T-J)/2) \log |\boldsymbol{\Omega}| \\ &\quad -(1/2) \sum_{t=J+1}^{T} (\mathbf{y}_t - \mathbf{P}' \mathbf{z}_t - \boldsymbol{\Psi}_t' \mathbf{B}' \mathbf{x}_t)' \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \mathbf{P}' \mathbf{z}_t - \boldsymbol{\Psi}_t' \mathbf{B}' \mathbf{x}_t) \\ &= -((T-J)p/2) \log 2\pi - ((T-J)/2) \log |\boldsymbol{\Omega}| \\ &\quad -(1/2) \sum_{t=J+1}^{T} (\mathbf{u}_t - \mathbf{P}' \mathbf{z}_t)' \boldsymbol{\Omega}^{-1} (\mathbf{u}_t - \mathbf{P}' \mathbf{z}_t). \end{aligned}$$

The Lagrange multiplier test is based on the score evaluated under the null hypothesis:

$$\frac{\partial \log L(\tilde{\theta})}{\partial \mathbf{P}} = \sum_{t=J+1}^{T} \left[\tilde{\mathbf{z}}_t \tilde{\mathbf{u}}_t' \tilde{\boldsymbol{\Omega}}^{-1} \right] = \tilde{\mathbf{Z}}' \tilde{\mathbf{U}} \tilde{\boldsymbol{\Omega}}^{-1}, \tag{4.4}$$

where

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \tilde{\mathbf{z}}_{J+1}' \\ \tilde{\mathbf{z}}_{J+2}' \\ \vdots \\ \tilde{\mathbf{z}}_T' \end{bmatrix}, \tilde{\mathbf{U}} = \begin{bmatrix} \tilde{\mathbf{u}}_{J+1}' \\ \tilde{\mathbf{u}}_{J+2}' \\ \vdots \\ \tilde{\mathbf{u}}_T' \end{bmatrix},$$

and $\tilde{\mathbf{z}}_t$, $\tilde{\mathbf{u}}_t$ and $\tilde{\boldsymbol{\Omega}}$ are estimates under the null hypothesis. Notice that the subscripts of $\tilde{\mathbf{Z}}$ and $\tilde{\mathbf{U}}$ begin from $J+1$, because usually $\tilde{\mathbf{u}}_0$, $\tilde{\mathbf{u}}_{-1}$, ..., $\tilde{\mathbf{u}}_{1-J}$ are not available. We have the following theorem:

Theorem 4.1. *Consider the model* (4.1) *and assume that the parameter estimates are consistent. Under the null hypothesis* $\mathbf{P}_1 = \mathbf{P}_2 = ... = \mathbf{P}_J = \mathbf{0}$*, the LM test statistic*

$$LM = \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{U}}'\tilde{\mathbf{Z}}[\tilde{\mathbf{Z}}'(\mathbf{I}_{T-J} - \tilde{\mathbf{K}}(\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}')\tilde{\mathbf{Z}}]^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{U}}\}, \tag{4.5}$$

where

$$\tilde{\mathbf{K}} = \begin{bmatrix} \text{vec}[\partial(\tilde{\mathbf{\Psi}}'_{J+1}\tilde{\mathbf{B}}'\mathbf{x}_{J+1})/\partial\theta]' \\ \text{vec}[\partial(\tilde{\mathbf{\Psi}}'_{J+2}\tilde{\mathbf{B}}'\mathbf{x}_{J+2})/\partial\theta]' \\ \vdots \\ \text{vec}[\partial(\tilde{\mathbf{\Psi}}'_{T}\tilde{\mathbf{B}}'\mathbf{x}_{T})/\partial\theta]' \end{bmatrix}. \tag{4.6}$$

has an asymptotic χ^2 *distribution with* Jp^2 *degrees of freedom.*

Proof. See Appendix B. □

The vectorised first order derivatives of $\mathbf{\Psi}'_t\mathbf{B}'\mathbf{x}_t$ w.r.t. parameters θ can be easily found in both univariate and multivariate cases, see Eitrheim and Teräsvirta (1996). We summarize them in Appendix D. However, in the multivariate case, the column dimension of the $\tilde{\mathbf{K}}$ matrix is $[(pk+q)m + 2(m-1)]p^2$, which grows very rapidly as a function of m and p. It is seen that $T - J \geq [(pk+q)m + 2(m-1)]p^2$ is a necessary condition for $\tilde{\mathbf{K}}'\tilde{\mathbf{K}}$ to have full rank, and consequently, the existence of the LM statistic. If either p or m or both are large, inverting $\tilde{\mathbf{K}}'\tilde{\mathbf{K}}$ requires care. The matrix may be near-singular and the inversion slow. This is bound to restrict the dimension of the null hypothesis, in practice the lag length J, in small and moderate samples, unless p, the dimension of the model, is sufficiently low.

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.2 by replacing $\mathbf{X}$ by $\tilde{\mathbf{K}}$ and $\mathbf{Z}_1$ by $\tilde{\mathbf{Z}}$. It contains the following steps:

1. Estimate the LVSTAR 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 $\mathbf{RSS}_0 = \tilde{\mathbf{U}}'\tilde{\mathbf{U}}$, where $\tilde{\mathbf{U}} = (\tilde{\mathbf{u}}_{J+1}, ..., \tilde{\mathbf{u}}_T)'$.

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

3. Compute the test statistic

$$\begin{aligned} LM &= (T-J)\,\text{tr}\{\mathbf{RSS}_0^{-1}(\mathbf{RSS}_0 - \mathbf{RSS}_1)\} \\ &= (T-J)(p - \text{tr}\{\mathbf{RSS}_0^{-1}\mathbf{RSS}_1\}). \end{aligned} \tag{4.7}$$

After obtaining $\mathbf{RSS}_0$ and $\mathbf{RSS}_1$, we can apply the tests with improved size suggested in Section 3. Nevertheless, there may be positive size distortion even in the tests with

improved empirical size. This is due to the fact that in practice $\tilde{\mathbf{U}}$ may not be completely orthogonal to the gradient matrix $\tilde{\mathbf{K}}$. This is because the LM statistic obtained from the two-step auxiliary regression is equivalent to (4.5) only when $\tilde{\mathbf{U}}'\tilde{\mathbf{K}} = \mathbf{0}$. To remedy the situation in univariate case, Eitrheim and Teräsvirta (1996) suggested to replace $\tilde{\mathbf{U}}$ by its orthogonal part to the space spanned by $\tilde{\mathbf{K}}$, i.e. to use $\tilde{\mathbf{V}} = (\mathbf{I} - \tilde{\mathbf{K}}(\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}')\tilde{\mathbf{U}}$. In the multivariate case, following Eitrheim and Teräsvirta (1996) leads to the following procedure:

1. Estimate the LVSTAR model under the null hypothesis of no serial correlation. Choose a lag length J of the serial correlation in residuals. Regress the residuals $\tilde{\mathbf{U}}$ on $\tilde{\mathbf{K}}$ from $t = J+1$ to T. Collect the residuals $\tilde{\mathbf{V}}$ and compute the matrix residual sum of squares $\mathbf{RSS}_0 = \tilde{\mathbf{V}}'\tilde{\mathbf{V}}$.

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

3. Compute the value of the test statistic (4.7).

It may be noted that this test as well as the subsequent ones is also valid when the estimated model is a VTAR model. The reason is that the threshold parameter in these models is estimated super consistently and can be assumed known in the tests. The test of no error autocorrelation thus becomes analogous to the corresponding test in linear VAR model, see Strikholm and Teräsvirta (2006) for discussion. A similar argument is valid for the test of no additive nonlinearity and parameter constancy.

4.2 Additive nonlinearity

We shall now consider the alternative hypothesis that after fitting a LVSTAR 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 exist. First, one may accept the alternative and estimate a VSTAR model with two transitions. Since the reasons for a rejection usually remain 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 LVSTAR model

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}_1'\mathbf{x}_t + \mathbf{G}_t^1\mathbf{B}_2'\mathbf{x}_t +, ..., +\mathbf{G}_t^{m-1}\mathbf{B}_m'\mathbf{x}_t + \boldsymbol{\varepsilon}_t + \mathbf{G}_t^m\mathbf{B}_{m+1}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ &= \mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t + \mathbf{G}_t^m\mathbf{B}_{m+1}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{4.8}$$

where $\boldsymbol{\varepsilon}_t$ is *i.i.d.* $N(\mathbf{0}, \boldsymbol{\Omega})$. We estimate the m-regime LVSTAR model without the additive nonlinear term $\mathbf{G}_t^m \mathbf{B}'_{m+1} \mathbf{x}_t$ and test $H_0 : \gamma_{mj} = 0$, $j = 1, ..., p$, against the alternative H_1 : at least one $\gamma_{mj} > 0$.

In order to carry out the test, 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 the theory behind the model 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, should the system have a single transition variable, a joint test of no additive nonlinearity test would be appropriate. We shall now develop such a test. As in Section 2, 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 2.2, 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

$$\mathbf{y}_t = \boldsymbol{\Psi}'_t \mathbf{B}' \mathbf{x}_t + \boldsymbol{\Theta}'_1 \mathbf{x}_t s_t + \boldsymbol{\varepsilon}^*_t, \tag{4.9}$$

or the third-order one, leading to

$$\mathbf{y}_t = \boldsymbol{\Psi}'_t \mathbf{B}' \mathbf{x}_t + \boldsymbol{\Theta}'_1 \mathbf{x}_t s_t + \boldsymbol{\Theta}'_2 \mathbf{x}_t s_t^2 + \boldsymbol{\Theta}'_3 \mathbf{x}_t s_t^3 + \boldsymbol{\varepsilon}^*_t. \tag{4.10}$$

The error vector $\boldsymbol{\varepsilon}^*_t$ contains the remainder term of the Taylor expansion. The corresponding null hypotheses are:

$$H_0 : \boldsymbol{\Theta}_1 = \mathbf{0} \tag{4.11}$$

for (4.9), and

$$H_0 : \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}_3 = \mathbf{0} \tag{4.12}$$

for (4.10). Let again

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_T \end{bmatrix}, \quad \mathbf{Z}_N = \begin{bmatrix} \mathbf{x}'_1 s_1 & \cdots & \mathbf{x}'_1 s_1^N \\ \mathbf{x}'_2 s_2 & \cdots & \mathbf{x}'_2 s_2^N \\ \vdots & & \vdots \\ \mathbf{x}'_T s_T & \cdots & \mathbf{x}'_T s_T^N \end{bmatrix},$$

Without loss of generality, we only consider testing the null hypothesis (4.11) in equation (4.9). The corresponding score evaluated under the null hypothesis is

$$\frac{\partial \log L(\tilde{\theta})}{\partial \boldsymbol{\Theta}_1} = \sum_{t=1}^{T} \left[\mathbf{x}_t s_t \left(\mathbf{y}_t - \tilde{\boldsymbol{\Psi}}'_t \tilde{\mathbf{B}}' \mathbf{x}_t \right)' \tilde{\boldsymbol{\Omega}}^{-1} \right],$$

where $\tilde{\mathbf{\Psi}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{\Omega}}$ are estimates under the null hypothesis. Denote $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \tilde{\mathbf{\Psi}}_t'\tilde{\mathbf{B}}'\mathbf{x}_t$, and $\tilde{\mathbf{E}} = (\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, ..., \tilde{\boldsymbol{\varepsilon}}_T)'$. Following Luukkonen et al. (1988), the LM-type statistic for testing (4.9) or (4.10) can be written as

$$LM = \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{E}}'\mathbf{Z}_N[\mathbf{Z}_N'(\mathbf{I}_T - \tilde{\mathbf{K}}(\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}')\mathbf{Z}_N]^{-1}\mathbf{Z}_N'\tilde{\mathbf{E}}\}, \tag{4.13}$$

where $\tilde{\mathbf{K}}$ has been defined in (4.6) 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 $\tilde{\mathbf{K}}'\tilde{\mathbf{K}}$ to have full rank, and consequently, for the existence of the LM test. We have the following theorem:

Theorem 4.2. *Consider the model in* (4.8) *and assume that the parameter estimators of the null model are consistent. Under the null hypothesis* (4.11) *or* (4.12)*, the LM test statistic*

$$LM = \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{E}}'\mathbf{Z}_N[\mathbf{Z}_N'(\mathbf{I}_T - \tilde{\mathbf{K}}(\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}')\mathbf{Z}_N]^{-1}\mathbf{Z}_N'\tilde{\mathbf{E}}\} \tag{4.14}$$

has an asymptotic χ^2 distribution with $pN(kp + q)$ degrees of freedom.

Proof. See Appendix B. □

Note that if $\mathbf{G}_t^1 = ... = \mathbf{G}_t^{m-1} = \mathbf{0}$ in (4.8), the test collapses into the linearity test discussed in Section 3.

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.2 by replacing $\mathbf{X}$ by $\tilde{\mathbf{K}}$. However, it also suffers from the empirical size distortion problem due to the non-orthogonality between $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{K}}$. Taking this into account, we have the following algorithm:

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

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

3. Compute the test statistic (2.13).

4.3 Parameter constancy

One of the central assumptions of the linear VAR as well the LVSTAR 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 LVSTAR

model. As in the univariate case, considered in Eitrheim and Teräsvirta (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, $\tau = t/T$. Consider the following LVSTAR model:

$$\mathbf{y}_t = \mathbf{\Psi}_t' \mathbf{B}(\tau)' \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{4.15}$$

where $\mathbf{B}(\tau) = \mathbf{B}_a + \mathbf{B}_b \lambda(\tau|\gamma, c)$ 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

$$\lambda_1(\tau|\gamma, c) = (1 + \exp(-\gamma(\tau - c)))^{-1} - 1/2, \tag{4.16}$$

$$\lambda_2(\tau|\gamma, c) = (1 + \exp(-\gamma(\tau - c_1)(\tau - c_2)))^{-1} - 1/2, \tag{4.17}$$

$$\lambda_3(\tau|\gamma, c) = \left(1 + \exp\left(-\gamma \prod_{j=1}^{3} (\tau - c_j)\right)\right)^{-1} - 1/2. \tag{4.18}$$

The null hypothesis of parameter constancy is $H_0 : \lambda_j(\tau|\gamma, c) = 0$, $j = 1$, 2, or 3. Function (4.16) postulates a smooth monotonic parameters change and function (4.17) a nonmonotonic change that is symmetric around $(c_1 + c_2)/2$. Function (4.18) 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 $\gamma \to \infty$ in (4.16), the alternative becomes a single structural break. When the same occurs in (4.17), one obtains a special case of a double break if $c_1 \neq c_2$, whereas (4.18) implies a triple break but only two extreme regimes if $c_1 \neq c_2 \neq c_3$.

For notational simplicity, we only consider the function (4.16) here. The first order Taylor expansion of (4.16) becomes

$$\lambda_1(\tau|\gamma, c) = (\gamma/4)(\tau - c) + r_t, \tag{4.19}$$

where r_t is the remainder. The model (4.15) is approximated and reparameterised as follows:

$$\mathbf{y}_t = \mathbf{\Psi}_t' \mathbf{B}_a' \mathbf{x}_t + \mathbf{\Psi}_t' \mathbf{B}_b' \mathbf{x}_t \tau + \boldsymbol{\varepsilon}_t^*, \tag{4.20}$$

where $\boldsymbol{\varepsilon}_t^*$ also contains the remainder r_t from the Taylor expansion (4.19). Note, however, that under H_0, $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t$. The new null hypothesis is

$$H_0 : \mathbf{B}_b = \mathbf{0}. \tag{4.21}$$

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

$$\frac{\partial \log L(\tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}_b} = \sum_{t=1}^{T} \left[\mathbf{x}_t \tau \cdot \left(\mathbf{y}_t - \tilde{\mathbf{\Psi}}_t' \tilde{\mathbf{B}}_a' \mathbf{x}_t \right)' \tilde{\mathbf{\Omega}}^{-1} \tilde{\mathbf{\Psi}}_t' \right],$$

where $\tilde{\boldsymbol{\Psi}}$, $\tilde{\mathbf{B}}_a$, and $\tilde{\boldsymbol{\Omega}}$ are estimates under the null. Denote $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \tilde{\boldsymbol{\Psi}}_t'\tilde{\mathbf{B}}_a'\mathbf{x}_t$, and $\tilde{\mathbf{E}} = (\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, ..., \tilde{\boldsymbol{\varepsilon}}_T)'$. Let

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \text{vec}(\tilde{\boldsymbol{\Psi}}_1 \otimes \mathbf{x}_1\tau_1)' \\ \text{vec}(\tilde{\boldsymbol{\Psi}}_2 \otimes \mathbf{x}_2\tau_2)' \\ \vdots \\ \text{vec}(\tilde{\boldsymbol{\Psi}}_T \otimes \mathbf{x}_T\tau_T)' \end{bmatrix}, \quad \tilde{\mathbf{K}} = \begin{bmatrix} \text{vec}[(\partial\tilde{\boldsymbol{\Psi}}_1'\tilde{\mathbf{B}}_a'\mathbf{x}_1/\partial\theta)]' \\ \text{vec}[(\partial\tilde{\boldsymbol{\Psi}}_2'\tilde{\mathbf{B}}_a'\mathbf{x}_2/\partial\theta)]' \\ \vdots \\ \text{vec}[(\partial\tilde{\boldsymbol{\Psi}}_T'\tilde{\mathbf{B}}_a'\mathbf{x}_T/\partial\theta)]' \end{bmatrix},$$

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

Theorem 4.3. *Consider the model in* (4.15) *with* $\lambda_1(\tau|\gamma, c)$ *and assume that the parameter estimators of the null model are consistent. Under the null hypothesis* $\lambda_1(\tau) = 0$ *expressed as in* (4.21)*, the LM-type statistic*

$$LM = \text{tr}\{\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{Z}}[\tilde{\mathbf{Z}}'(\mathbf{I}_T - \tilde{\mathbf{K}}(\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}')\tilde{\mathbf{Z}}]^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{E}}\} \tag{4.22}$$

has an asymptotic χ^2 *distribution with* $mp^2(kp + q)$ *degrees of freedom.*

Proof. See Appendix B. □

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

The test can also be performed using the two-step auxiliary regression algorithm given in Section 2.2 by replacing $\mathbf{X}$ by $\tilde{\mathbf{K}}$ and $\mathbf{Z}_1$ by $\tilde{\mathbf{Z}}$. In order to alleviate the empirical size distortion problem due to the non-orthogonality between $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{K}}$, we propose the following algorithm:

1. Estimate the LVSTAR model under the null hypothesis of constant parameters over time. Regress the residuals $\tilde{\mathbf{E}}$ on $\tilde{\mathbf{K}}$. Collect the residuals $\tilde{\mathbf{V}}$, and compute the matrix residual sum of squares $\mathbf{RSS}_0 = \tilde{\mathbf{V}}'\tilde{\mathbf{V}}$.

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

3. Compute the test statistic (2.13).

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 (4.15) is linear, the test collapses into the corresponding parameter constancy test in a linear VAR model, see He et al. (2009).

4.4 Curse of dimensionality

All three tests introduced in previous sections suffer from the "curse of dimensionality". This happens when the combination of the number of equations p, the number of regimes m, and the number of lags k are sufficiently large. The necessary conditions for the existence of these tests are: $T - J \geq \text{cd}(\tilde{\mathbf{K}})$ for the test of no serial correlation test and $T \geq \text{cd}(\tilde{\mathbf{K}})$ for the other two joint tests, where $\text{cd}(\tilde{\mathbf{K}})$ is the column dimension of the matrix $\tilde{\mathbf{K}}$. The closer $\text{cd}(\tilde{\mathbf{K}})$ 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.

5 Simulation study

5.1 *P*-value plot and *p*-value discrepancy plot

As already noted, the LM or LM-type test statistic has the advantage that estimation of the alternative model is avoided. This makes the statistic relatively easy to simulate. In reporting results, we make use of the graphical methods by Davidson and MacKinnon (1998). These authors suggested p-value and p-value discrepancy plots for the purpose. Consider a Monte Carlo experiment in which N realizations of some test statistic τ 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 τ_j, $j = 1, ..., N$, in the nominal distribution, which can be the asymptotic null distribution of τ, or an approximation to the finite sample null distribution of τ.

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

$$\hat{F}(x_i) = \frac{1}{N} \sum_{j=1}^{N} I(p_j \leq x_i), \tag{5.1}$$

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 τ. We construct a discrete grid $\{x_i\}_{i=1}^{M}$ in advance in order to cover a reasonable sub interval of the (0,1) interval. In this work we focus on the (0, 0.2) interval, because it contains the most commonly applied significance levels and set

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

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 scatterplot 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 statistic tends to be rather conservative, however, and as such may sometimes mislead us to think that there is no under- or over-rejection when x_i is close to zero, although the truth is different.

Instead, we employ a 95% two-sided asymptotic normal confidence band assuming for the nominal distribution of the test statistic. Under this assumption, for each $j = 1, ..., N$, 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 distribution of the p-value discrepancy $\hat{F}(x_i) - x_i$ can be approximated by a normal distribution with zero mean and variance $N^{-1}x_i(1 - x_i)$.

5.2 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 statistic 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 between four different types of test statistics: the LM test in Theorem 2.4, the rescaled LM test in (3.1), Wilks's Λ-test in (3.5) and Rao's F-test in (3.6). 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 with *i.i.d.* errors. 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.

5.2.1 Linearity tests

The basic DGP is a p-dimensional vector autoregressive time series $\mathbf{y}_t$ 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 model is time-varying. The true DGP is as follows:

$$y_{i,t} = \sum_{j=1}^{k} \rho^i y_{i,t-j} + \varepsilon_{i,t}, \tag{5.2}$$

where $\rho = 0.4$. Different combinations of the values below will be considered:

$$p \in \{2, 5, 10\}, \quad k \in \{1, 2, 5\}, \quad T \in \{30, 50, 100\}. \tag{5.3}$$

We investigate three cases: the bivariate case ($p = 2$), a high-dimensional one ($p = 5$) and an extremely high-dimensional case ($p = 10$). The lag length k also matters, because it determines the column dimension of the matrix $\mathbf{Z}$, or in other words, the degrees of freedom of the nominal distribution. The following designs will be studied and the results plotted:

$$\begin{array}{lllll}
\text{Design 1:} & p = 2, & k = 1 & \text{and} & T = 30; \\
\text{Design 2:} & p = 2, & k = 1 & \text{and} & T = 100; \\
\text{Design 3:} & p = 2, & k = 2 & \text{and} & T = 30; \\
\text{Design 4:} & p = 2, & k = 5 & \text{and} & T = 30; \\
\text{Design 5:} & p = 5, & k = 1 & \text{and} & T = 50; \\
\text{Design 6:} & p = 10, & k = 1 & \text{and} & T = 50.
\end{array}$$

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

As for Design 1, both the LM test and the rescaled test, denoted by F in all graphs, are size-distorted. The former one over-rejects, whereas the latter under-rejects. This agrees with previous results. The empirical size of Wilks's Λ 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 Λ and Rao's F-test have a very similar performance. The LM test now works well, which accords with the

theoretical result that the LM statistic converges in distribution to the χ^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 3 and 4, respectively. Both designs are bivariate and $T = 30$. The lag lengths are different, two and five.

In comparing the results 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 Λ, 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 it doubles to ten in Design 6, other things equal. Again, while the behaviour of Wilks's Λ 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. This accords with the results in Edgerton and Shukur (1999) who considered testing autocorrelation in a linear vector system.

5.2.2 Misspecification tests

The basic DGP is a p-dimensional vector autoregressive time series $\mathbf{y}_t$ of lag order k with one additive logistic nonlinear component, i.e., LVSTAR. 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 $\mathbf{c} = \mathbf{0}$ to be the location parameter vector, and set $\boldsymbol{\gamma} = \mathbf{1}$. The data generating process is as follows:

$$y_{i,t} = \left(\sum_{j=1}^{k} \rho_1^i y_{i,t-j} \right) (1 - g(s_{it})) + \left(\sum_{j=1}^{k} \rho_2^i y_{i,t-j} \right) g(s_{it}) + \varepsilon_{i,t}, \tag{5.4}$$

where $\rho_1 = 0.4$, $\rho_2 = 0.2$.

We generate 5000 realizations from our DGP. For each realization, we estimate the LVSTAR model (5.4). We compute the residual vector for each realization, and following the three algorithms in Section 4, 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 Λ 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 7. As can be expected, the LM test over-rejects, whereas the rescaled test under-rejects. The

empirical size of Wilks's Λ 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 both 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 8 shows the p-value discrepancy plots for test of no additive nonlinearity. Both Wilks's Λ 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 9 shows the p-value discrepancy plots for the test of parameter constancy. For $T = 200$, Wilks's Λ 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 $\mathbf{Z}$ and $\mathbf{K}$. The column dimensions of the matrices $\mathbf{Z}$ and $\mathbf{K}$ affect the empirical size of the corresponding test statistic. In the three tests, $\mathbf{K}$ has the same column dimension, whereas those of $\mathbf{Z}$ 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 $\mathbf{Z}$ of the parameter constancy test is the squared size of a Kronecker product, and the test thus has the worst small sample performance. For certain sample size T, the LM test and the rescaled test of the three misspecification tests perform differently. This can be explained by different column dimension of the matrix $\mathbf{Z}$. However, Rao's F-test and the Wilks's Λ test are still performing well.

If the errors of the LVSTAR model are not iid, even Rao's F-test can be size-distorted. In Teräsvirta and Yang (2014), the authors corrected the size of the LM test using the wild bootstrap, proposed originally by Wu (1986), because the errors in their applications were likely to contain conditional heteroskedasticity. This is a valid procedure if the appropriate conditions, see Gonçalves and Kilian (2004), are satisfied when the null hypothesis holds. We do not simulate this variant of the LM test here because the wild bootstrap effectively corrects the size of the original LM test.

6 Concluding remarks

In this paper, we propose Lagrange-multiplier type linearity and misspecification tests in the LVSTR 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 LVSTR 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 are of interest because they are affected by the dimension of the model. We report the results of 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's Λ statistic and Rao's F statistic that have satisfying size properties are recommended for empirical use. Nevertheless, the size of the LM test can be corrected by an appropriate bootstrapped version. If the errors are suspected to contain conditional heteroskedasticity, the bootstrap to be used is the wild bootstrap.

References

Anderson, T. W.: 1958, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Bartlett, M. S.: 1954, A note on multiplying factors for various χ^2 approximations, *Journal of the Royal Statistical Society. Series B* **16**, 296–298.

Bera, A. K., Byron, R. P. and Jarque, C. M.: 1981, Further evidence on asymptotic tests for homogeneity and symmetry in large demand systems, *Economics Letters* **8**, 101–105.

Breusch, T. S. and Pagan, A. R.: 1980, The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies* **47**, 239–253.

Camacho, M.: 2004, Vector smooth transition regression models for US GDP and the composite index of leading indicators, *Journal of Forecasting* **23**, 173–196.

Davidson, R. and MacKinnon, J. G.: 1998, Graphical methods for investigating the size and power of hypothesis tests, *The Manchester School of Economic & Social Studies* **66**, 1–26.

Edgerton, D. and Shukur, G.: 1999, Testing autocorrelation in a system perspective, *Econometric Reviews* **18**, 343–386.

Eitrheim, Ø. and Teräsvirta, T.: 1996, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–75.

Feigin, P. D.: 1976, Maximum likelihood estimation for continuous-time stochastic processes, *Advances in Applied Probability* **8**, 712–736.

Godfrey, L. G.: 1988, *Misspecification tests in econometrics*, Cambridge University Press, Cambridge.

Gonçalves, S. and Kilian, L.: 2004, Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**(1), 89–120.

He, C., Teräsvirta, T. and González, A.: 2009, Testing parameter constancy in vector autoregressive models against continuous change, *Econometric Reviews* **28**, 225–245.

Hubrich, K. and Teräsvirta, T.: 2013, Thresholds and smooth transitions in vector autoregressive models, *Advances in Econometrics* **32**, 273–326.

Laitinen, K.: 1978, Why is demand homogeneity so often rejected?, *Economics Letters* **1**, 187–191.

Lütkepohl, H.: 2004, Vector autoregressive and vector error correction models, *in Applied Time Series Econometrics, H. Lütkepohl and M. Krätzig, eds.* pp. 86–158. Cambridge, Cambridge University Press.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Mardia, K. V., Kent, J. T. and Bibby, J. M.: 1979, *Multivariate Analysis*, Academic Press.

Meisner, J. F.: 1979, The sad fate of the asymptotic Slutsky symmetry test for large systems, *Economics Letters* **2**, 231–233.

Ramsey, J. B.: 1969, Test for specification error in classical linear least squares regression analysis, *Journal of the Royal Statistical Society, Series B* **31**, 350–371.

Rao, C. R.: 1951, An asymptotic expansion of the distribution of Wilks' criterion, *Bulletin of the International Statistical Institute* **33**, 177–180.

Rao, C. R.: 1965, *Linear Statistical Inference and Its Applications*, 1st edn, New York: Wiley.

Saikkonen, P.: 2008, Stability of regime switching error correction models under linear cointegration, *Econometric Theory* **24**, 294–318.

Saikkonen, P. and Luukkonen, R.: 1988, Lagrange multiplier tests for testing nonlinearities in time series models, *Scandinavian Journal of Statistics* **15**(1), 55–68.

Shukur, G. and Edgerton, D.: 2002, The small sample properties of the reset test as applied to systems of equations, *Journal of Statistical Computation and Simulation* **72**, 909–924.

Strikholm, B. and Teräsvirta, T.: 2006, A sequential procedure for determining the number of regimes in a threshold autoregressive model, *Econometrics Journal* **9**, 472–491.

Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T.: 1998, Modeling economic relationships with smooth transition regressions, *in* A. Ullah and D. E. Giles (eds), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford, Oxford University Press.

Teräsvirta, T. and Yang, Y.: 2014, Specification, Estimation and Evaluation of Vector Smooth Transition Autoregressive Models with Applications, *CREATES Research Paper, Aarhus University* .

Tsay, R. S.: 1998, Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* **93**, 1188–1202.

Wu, C. F. J.: 1986, Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* **14**(4), 1261–1295.

A Proof of Theorem 2.4

Proof. The Lagrange multiplier test under the null is derived from the score matrix

$$\frac{\partial \log L(\tilde{\theta})}{\partial \mathbf{\Theta}_1} = \sum_{t=1}^{T} \left\{ \mathbf{x}_t s_t \left(\mathbf{y}_t - \tilde{\mathbf{\Theta}}_0' \mathbf{x}_t \right)' \tilde{\mathbf{\Omega}}^{-1} \right\} = \mathbf{Z}' \left(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}_1 \right) \tilde{\mathbf{\Omega}}^{-1}, \tag{A.1}$$

where

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

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

To see this, we write (A.1) as follows

$$\begin{aligned} \mathbf{Q} &\equiv \frac{\partial \log L(\tilde{\theta})}{\partial \mathbf{\Theta}_1} = \mathbf{Z}' \left(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}_1 \right) \tilde{\mathbf{\Omega}}^{-1} \\ &= \mathbf{Z}'(\mathbf{Y} - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})\tilde{\mathbf{\Omega}}^{-1} \\ &= \mathbf{Z}'(\mathbf{I} - \mathbf{P}_x)(\mathbf{X}\mathbf{B}_1 + \mathbf{E})\tilde{\mathbf{\Omega}}^{-1} \\ &= \mathbf{Z}'(\mathbf{I} - \mathbf{P}_x)\mathbf{E}\tilde{\mathbf{\Omega}}^{-1}. \end{aligned}$$

Under the null hypothesis, $\mathbf{Y} = \mathbf{X}\mathbf{B}_1 + \mathbf{E}$, where $\mathbf{E} = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_T)'$ and $\text{vec}(\mathbf{E}')$ follows a $\mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \mathbf{\Omega})$ distribution. Under the null hypothesis, $\tilde{\mathbf{\Omega}}$ will converge to $\mathbf{\Omega}$ in probability. Set

$$\mathbf{S} = \left(\mathbf{Z}'(\mathbf{I} - \mathbf{P}_x)\mathbf{Z} \right)^{-\frac{1}{2}} \mathbf{Q}, \tilde{\mathbf{\Omega}}^{\frac{1}{2}}$$

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

$$LM = \text{tr}\{\mathbf{S}'\mathbf{S}\} = \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}_1)'\mathbf{Z}\left[\mathbf{Z}'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z}\right]^{-1}\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}_1)\},$$

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

B LM test statistic against an additive component

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

$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{s}_t|\mathbf{\Theta}_0) + \mathbf{A}_t'\mathbf{\Theta}_1'\mathbf{z}_t + \boldsymbol{\varepsilon}_t \tag{B.1}$$

where $\mathbf{f}$ is a vector of linear or nonlinear functions, in which $\mathbf{x}_t$ and $\mathbf{s}_t$ are vectors of independent variables, $\mathbf{s}_t$ may be referred to as the vector of transition variables, $\mathbf{A}_t$ is a time-varying matrix which contains some nuisance parameters, and $\mathbf{z}_t$ is a vector of independent variables in the additive component. In many cases, $\mathbf{z}_t = \mathbf{x}_t$, see, for example, the joint test of linearity against the LVSTAR alternative in Section 2, but $\mathbf{z}_t \neq \mathbf{x}_t$ is allowed as well. The test of no error serial correlation test in Section 4.1 serves as an example.

The corresponding set of parameters in the model (B.1) is $\theta = \{\mathbf{\Theta}_0, \mathbf{\Theta}_1, \mathbf{\Omega}\}$ where $\mathbf{\Theta}_0$ is a parameter matrix in $\mathbf{f}$, $\mathbf{\Theta}_1$ is a parameter matrix in the additive component, and $\mathbf{\Omega}$ is the positive definite covariance matrix of the errors. The existence of the additive component $\mathbf{A}_t'\mathbf{\Theta}_1'\mathbf{z}_t$ is going to be tested, and then the null hypothesis is $H_0 : \mathbf{\Theta}_1 = \mathbf{0}$.

B.1 The case when $\mathbf{A}_t = \mathbf{I}_p$

Consider the special case when $\mathbf{A}_t = \mathbf{I}_p$, for example, the error serial correlation test in Section 4.1, the joint test of no additive nonlinearity in Section 4.2, and the joint test of linearity against the LVSTAR alternative with a single transition variable in Section 2. The corresponding block of the score matrix takes the form

$$\frac{\partial \log L(\tilde{\theta})}{\partial \mathbf{\Theta}_1} = \sum_{t=1}^{T} \left\{\mathbf{z}_t\tilde{\boldsymbol{\varepsilon}}_t'\tilde{\mathbf{\Omega}}^{-1}\right\} = \mathbf{Z}'\tilde{\mathbf{E}}\tilde{\mathbf{\Omega}}^{-1}, \tag{B.2}$$

where $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{f}(\mathbf{x}_t, \mathbf{s}_t|\tilde{\mathbf{\Theta}}_0)$, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T)'$ and $\tilde{\mathbf{E}} = (\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, ..., \tilde{\boldsymbol{\varepsilon}}_T)'$. The tilde means estimates under the null hypothesis.

The vectorized LM test statistic is

$$LM = \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z})'\left((\mathbf{Z}'(\mathbf{I} - \mathbf{P}_\mathbf{K})\mathbf{Z}) \otimes \tilde{\mathbf{\Omega}}\right)^{-1}\text{vec}(\tilde{\mathbf{E}}\mathbf{Z})', \tag{B.3}$$

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

$$\mathbf{K} = \begin{bmatrix} \text{vec}\left(\partial \mathbf{f}(\mathbf{x}_1, \mathbf{s}_1|\tilde{\boldsymbol{\Theta}}_0)/\partial \boldsymbol{\Theta}_0\right)' \\ \text{vec}\left(\partial \mathbf{f}(\mathbf{x}_2, \mathbf{s}_2|\tilde{\boldsymbol{\Theta}}_0)/\partial \boldsymbol{\Theta}_0\right)' \\ \vdots \\ \text{vec}\left(\partial \mathbf{f}(\mathbf{x}_T, \mathbf{s}_T|\tilde{\boldsymbol{\Theta}}_0)/\partial \boldsymbol{\Theta}_0\right)' \end{bmatrix}. \tag{B.4}$$

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 2.1-3, the vectorized score matrix is asymptotically normally distributed with $p\,\text{cd}(\mathbf{Z})$ degrees of freedom, i.e. the number of elements in $\text{vec}(\tilde{\mathbf{E}}'\mathbf{Z})$, or the number of restrictions. See for example Breusch and Pagan (1980).

The statistic (B.3) can be written as follows:

$$\begin{aligned} LM &= \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z})' \left((\mathbf{Z}'(\mathbf{I} - \mathbf{P_K})\mathbf{Z}) \otimes \tilde{\boldsymbol{\Omega}}\right)^{-1} \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z}) \\ &= \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z})' \left((\mathbf{Z}'(\mathbf{I} - \mathbf{P_K})\mathbf{Z})^{-1} \otimes \tilde{\boldsymbol{\Omega}}^{-1}\right) \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z}) \\ &= \text{vec}(\tilde{\mathbf{E}}'\mathbf{Z})' \text{vec}\left(\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{E}}'\mathbf{Z}\,(\mathbf{Z}'(\mathbf{I} - \mathbf{P_K})\mathbf{Z})^{-1}\right) \\ &= \text{tr}\left\{\mathbf{Z}'\tilde{\mathbf{E}}\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{E}}'\mathbf{Z}\,(\mathbf{Z}'(\mathbf{I} - \mathbf{P_K})\mathbf{Z})^{-1}\right\} \\ &= \text{tr}\left\{\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{E}}'\mathbf{Z}\,(\mathbf{Z}'(\mathbf{I} - \mathbf{P_K})\mathbf{Z})^{-1}\,\mathbf{Z}'\tilde{\mathbf{E}}\right\}. \end{aligned} \tag{B.5}$$

Note that (B.5) avoids vectorization and Kronecker products. Furthermore, the value of (B.5) can be obtained by applying the following auxiliary regression:

1. Estimate the restricted model under the null hypothesis. Collect the residuals $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{f}(\mathbf{x}_t, \mathbf{s}_t|\tilde{\boldsymbol{\Theta}}_0)$ and form $\tilde{\mathbf{E}}$. Compute the matrix residual sum of squares $\mathbf{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}}$.

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

3. Compute the test statistic

$$\begin{aligned} LM &= T\,\text{tr}\{\mathbf{RSS}_0^{-1}(\mathbf{RSS}_0 - \mathbf{RSS}_1)\} \\ &= T(p - \text{tr}\{\mathbf{RSS}_0^{-1}\mathbf{RSS}_1\}). \end{aligned} \tag{B.6}$$

B.2 The case when $\mathbf{A}_t \neq \mathbf{I}_p$

Consider the case when $\mathbf{A}_t \neq \mathbf{I}_p$, for example the joint test of linearity against the LVSTAR alternative with different transition variables in Section 2 in which $\mathbf{A}_t = \mathbf{S}_t$, and the test

of parameter constancy in Section 4.3 in which $\mathbf{A}_t = \tilde{\mathbf{\Psi}}_t$. The corresponding block of the score matrix takes the form

$$\frac{\partial \log L(\tilde{\theta})}{\partial \mathbf{\Theta}_1} = \sum_{t=1}^{T} \left\{ \mathbf{z}_t \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\mathbf{\Omega}}^{-1} \mathbf{A}_t' \right\}, \tag{B.7}$$

where $\tilde{\boldsymbol{\varepsilon}}_t$ has been defined in (B.2).

The LM statistic (B.3) is still valid for testing $\mathbf{\Theta}_1 = \mathbf{0}$, when

$$\mathbf{Z} = \left(\text{vec}(\mathbf{A}_1 \otimes \mathbf{z}_1), \text{vec}(\mathbf{A}_2 \otimes \mathbf{z}_2), ..., \text{vec}(\mathbf{A}_T \otimes \mathbf{z}_T)\right)'. \tag{B.8}$$

To see this, write

$$\begin{aligned} \mathbf{A}_t' \mathbf{\Theta}_1' \mathbf{z}_t &= \text{vec}(\mathbf{z}_t' \mathbf{\Theta}_1 \mathbf{A}_t) = (\mathbf{A}_t \otimes \mathbf{z}_t)' \text{vec}(\mathbf{\Theta}_1) = \text{vec}\left(\text{vec}(\mathbf{\Theta}_1)' (\mathbf{A}_t \otimes \mathbf{z}_t)\right) \\ &= \left(\mathbf{I}_p \otimes \text{vec}(\mathbf{\Theta}_1)\right)' \text{vec}\left(\mathbf{A}_t \otimes \mathbf{z}_t\right) = \breve{\mathbf{\Theta}}_1' \breve{\mathbf{z}}_t, \end{aligned} \tag{B.9}$$

where $\breve{\mathbf{\Theta}}_1 = \mathbf{I}_p \otimes \text{vec}(\mathbf{\Theta}_1)$ and $\breve{\mathbf{z}}_t = \text{vec}\left(\mathbf{A}_t \otimes \mathbf{z}_t\right)$. Note that $\breve{\mathbf{\Theta}}_1$ contains the same elements as $\mathbf{\Theta}_1$, the remaining ones being equal to zero. The corresponding number of degrees of freedom should thus be equal to the number of nonzero parameters in $\breve{\mathbf{\Theta}}_1$, that is, the number of parameters in $\mathbf{\Theta}_1$, as only these parameters can vary freely.

The null hypothesis can be rewritten as $H_0 : \breve{\mathbf{\Theta}}_1 = \mathbf{0}$. The corresponding block of the score matrix is (B.2), with $\mathbf{Z}$ is defined as in (B.8).

Suppose that $\mathbf{A}_t$ is an $a \times p$ matrix, $\breve{\mathbf{\Theta}}_1$ is a $b \times a$ matrix and $\mathbf{z}_t$ is a $b \times 1$ vector. Moreover, write $\mathbf{A}_t = (\mathbf{a}_{1t}, ..., \mathbf{a}_{pt})$, where $\mathbf{a}_{jt}$, $j = 1, ..., p$, is an $a \times 1$ vector. We have

$$\breve{\mathbf{\Theta}}_1' \breve{\mathbf{z}}_t = \begin{bmatrix} \text{vec}(\mathbf{\Theta}_1)' (\mathbf{a}_{1t} \otimes \mathbf{z}_t) \\ \text{vec}(\mathbf{\Theta}_1)' (\mathbf{a}_{2t} \otimes \mathbf{z}_t) \\ \vdots \\ \text{vec}(\mathbf{\Theta}_1)' (\mathbf{a}_{pt} \otimes \mathbf{z}_t) \end{bmatrix}_{p \times 1} \quad \text{and} \quad \mathbf{Z}_j = \begin{bmatrix} (\mathbf{a}_{j1} \otimes \mathbf{z}_1)' \\ (\mathbf{a}_{j2} \otimes \mathbf{z}_2)' \\ \vdots \\ (\mathbf{a}_{jT} \otimes \mathbf{z}_T)' \end{bmatrix}. \tag{B.10}$$

If $\mathbf{A}_t = \mathbf{S}_t$, where $\mathbf{S}_t$ is a diagonal matrix of p transition variables $s_{1t}, ..., s_{pt}$, the statistic (B.3) is used for testing linearity against the LVSTAR model with these transition variables (s_{jt} for the j^{th} equation). A special case of this is $s_{1t} = ... = s_{pt}$ or $\mathbf{A}_t = s_t \mathbf{I}_p$, i.e., system has a single transition variable. Then (B.10) simplifies to

$$\breve{\mathbf{\Theta}}_1' \breve{\mathbf{z}}_t = \begin{bmatrix} \boldsymbol{\theta}_1' \cdot (\mathbf{z}_t s_{1t}) \\ \boldsymbol{\theta}_2' \cdot (\mathbf{z}_t s_{2t}) \\ \vdots \\ \boldsymbol{\theta}_p' \cdot (\mathbf{z}_t s_{pt}) \end{bmatrix}_{p \times 1} \quad \text{and} \quad \mathbf{Z}_j = \begin{bmatrix} (\mathbf{z}_1 s_{j1})' \\ (\mathbf{z}_2 s_{j2})' \\ \vdots \\ (\mathbf{z}_T s_{jT})' \end{bmatrix}, \tag{B.11}$$

where $\boldsymbol{\theta}_j$, $j = 1, ..., p$, is a $b \times 1$ column vector.

We can still use the auxiliary regression to compute the value of (B.5). However, from (B.10) we see that the auxiliary regression should be carried out equation by equation. This leads to the following procedure:

1. Estimate the restricted model under the null hypothesis. Collect the residuals $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{f}(\mathbf{x}_t, \mathbf{s}_t|\tilde{\boldsymbol{\Theta}}_0)$ and form $\tilde{\mathbf{E}}$. Compute the matrix residual sum of squares $\mathbf{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}}$.

2. Run an auxiliary regression for each column of $\tilde{\mathbf{E}}$, that is, regress column j of $\tilde{\mathbf{E}}$, $j = 1, ..., p$, on $(\mathbf{K}, \mathbf{Z}_j)$ where $\mathbf{Z}_j$ defined in (B.10) or in (B.11) if $\mathbf{A}_t = \mathbf{S}_t$. Collect the residuals $\tilde{\boldsymbol{\Xi}}$, and form the matrix residual sum of squares $\mathbf{RSS}_1 = \tilde{\boldsymbol{\Xi}}'\tilde{\boldsymbol{\Xi}}$.

3. Compute the test statistic in (B.6).

C Proof of Theorem 3.1

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

$$\frac{\partial \log L(\tilde{\theta})}{\partial \boldsymbol{\Theta}_1} = \mathbf{Z}' \left(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}_1 \right) \tilde{\boldsymbol{\Omega}}^{-1}. \tag{C.1}$$

Use of the auxiliary regression approach for computing the test statistic produces two residual sums of squares, $\mathbf{RSS}_0$ and $\mathbf{RSS}_1$. The first one, $\mathbf{RSS}_0$, is the residual sum of squares matrix from the restricted regression, i.e., $\mathbf{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}}$, $\tilde{\mathbf{E}} = (\mathbf{I} - \mathbf{P}_x)\mathbf{Y}$, where $\mathbf{P}_x$ is the projection matrix of $\mathbf{X}$. Notice that under the null hypothesis, $\mathbf{Y} = \mathbf{X}\mathbf{B}_1 + \mathbf{E}$, where $\mathbf{E} = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_T)'$, and $\text{vec}(\mathbf{E}') \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Omega})$. Consequently, $\tilde{\mathbf{E}} = (\mathbf{I} - \mathbf{P}_x)\mathbf{Y} = (\mathbf{I} - \mathbf{P}_x)\mathbf{E}$.

$\mathbf{RSS}_1$ is the residual sum of squares matrix from the auxiliary regression, i.e., $\mathbf{RSS}_1 = \tilde{\boldsymbol{\Xi}}'\tilde{\boldsymbol{\Xi}}$ with $\tilde{\boldsymbol{\Xi}} = (\mathbf{I} - \mathbf{P}_{xz})\tilde{\mathbf{E}}$, where $\mathbf{P}_{xz}$ is the projection matrix of the matrix $[\ \mathbf{X}, \mathbf{Z}\ ]$, i.e.,

$$\mathbf{P}_{xz} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}.$$

Let the $p \times p$ matrix $\mathbf{W}_1 = \mathbf{RSS}_0 - \mathbf{RSS}_1$ and the $p \times p$ matrix $\mathbf{W}_2 = \mathbf{RSS}_1$.

For $\mathbf{W}_1$, it follows that

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{RSS}_0 - \mathbf{RSS}_1 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}} - \tilde{\boldsymbol{\Xi}}'\tilde{\boldsymbol{\Xi}} \\ &= \tilde{\mathbf{E}}'\mathbf{P}_{xz}\tilde{\mathbf{E}} = \tilde{\mathbf{E}}'\mathbf{Z}(\mathbf{Z}'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{E}} \\ &= \mathbf{E}'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z}(\mathbf{Z}'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I}_T - \mathbf{P}_x)\mathbf{E}. \end{aligned}$$

Let $\mathbf{I}_T - \mathbf{P}_x = \mathbf{R}\mathbf{R}'$, where $\mathbf{R} \perp \mathbf{X}$ and $\mathbf{R}'\mathbf{R} = \mathbf{I}_{T-\text{cd}(\mathbf{X})}$. Then

$$\mathbf{W}_1 = \mathbf{E}'\mathbf{R}\mathbf{R}'\mathbf{Z}(\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{E}.$$

Set $\mathbf{V}_1 = \mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{E}$. So $\mathbf{V}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z} \otimes \boldsymbol{\Omega})$. It is seen that $\mathbf{W}_1$ follows a Wishart distribution generated by $\mathbf{V}_1$:

$$\mathbf{W}_1 = \mathbf{V}_1'(\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z})^{-1}\mathbf{V}_1 \sim \mathcal{W}_p(\boldsymbol{\Omega}, \text{cd}(\mathbf{Z}))$$

For $\mathbf{W}_2$, we obtain

$$\begin{aligned}\mathbf{W}_2 &= \mathbf{RSS}_1 = \tilde{\mathbf{\Xi}}'\tilde{\mathbf{\Xi}} = \tilde{\mathbf{E}}'(\mathbf{I}-\mathbf{P}_{xz})\tilde{\mathbf{E}} = \tilde{\mathbf{E}}'\tilde{\mathbf{E}} - \tilde{\mathbf{E}}'\mathbf{P}_{xz}\tilde{\mathbf{E}} \\ &= \tilde{\mathbf{E}}'\tilde{\mathbf{E}} - \tilde{\mathbf{E}}'\mathbf{Z}(\mathbf{Z}'(\mathbf{I}-\mathbf{P}_x)\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{E}} \\ &= \mathbf{E}'(\mathbf{I}-\mathbf{P}_x)(\mathbf{I}-\mathbf{Z}(\mathbf{Z}'(\mathbf{I}-\mathbf{P}_x)\mathbf{Z})^{-1}\mathbf{Z}')(\mathbf{I}-\mathbf{P}_x)\mathbf{E} \\ &= \mathbf{E}'\mathbf{R}\mathbf{R}'(\mathbf{I}-\mathbf{Z}(\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{R}\mathbf{R}'\mathbf{E} \\ &= \mathbf{E}'\mathbf{R}(\mathbf{I}-\mathbf{R}'\mathbf{Z}(\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R})\mathbf{R}'\mathbf{E}.\end{aligned}$$

We have $\mathbf{I}_{T-\mathrm{cd}(\mathbf{X})}-\mathbf{R}'\mathbf{Z}(\mathbf{Z}'\mathbf{R}\mathbf{R}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R} = \mathbf{Q}\mathbf{Q}'$, where $\mathbf{Q} \perp \mathbf{R}'\mathbf{Z}$ and $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_{T-\mathrm{cd}(\mathbf{X})-\mathrm{cd}(\mathbf{Z})}$. Using this, $\mathbf{W}_2 = \mathbf{E}'\mathbf{R}\mathbf{Q}\mathbf{Q}'\mathbf{R}'\mathbf{E}$. Set $\mathbf{V}_2 = \mathbf{Q}'\mathbf{R}'\mathbf{E}$, so we have $\mathbf{V}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\otimes\mathbf{\Omega})$. It is seen that $\mathbf{W}_2$ follows a Wishart distribution generated by $\mathbf{V}_2$:

$$\mathbf{W}_2 = \mathbf{V}_2'\mathbf{V}_2 \sim \mathcal{W}_p(\mathbf{\Omega}, T-\mathrm{cd}(\mathbf{X})-\mathrm{cd}(\mathbf{Z})).$$

Stacking the columns of $\mathbf{V}_1$ and $\mathbf{V}_2$ yields the random matrix

$$\mathbf{U} = \begin{pmatrix}\mathbf{V}_1 \\ \mathbf{V}_2\end{pmatrix} = \begin{pmatrix}\mathbf{Z}'\mathbf{R} \\ \mathbf{Q}'\end{pmatrix}\mathbf{R}'\mathbf{E}$$

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

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

because $\mathbf{Q} \perp \mathbf{R}'\mathbf{Z}$. We conclude that $\mathbf{V}_1$ and $\mathbf{V}_2$ are uncorrelated, and independent due to normality. It follows that $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent as desired. □

D The first-order partial derivatives of $\mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t$

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

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

$$\frac{\partial\mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t}{\partial b_{ij}} = \mathbf{\Psi}_t'\mathbf{H}_{ij}'\mathbf{x}_t, \tag{D.1}$$

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

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

$$\frac{\partial\mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t}{\partial\delta_{ij}} = \left(\mathbf{0}_p\ ,\ \ldots\ ,\ \frac{\partial\mathbf{G}_t^i}{\partial\delta_{ij}}\ ,\ \ldots\mathbf{0}_p\right)\mathbf{B}'\mathbf{x}_t = \frac{\partial\mathbf{G}_t^i}{\partial\delta_{ij}}\mathbf{B}_{i+1}'\mathbf{x}_t, \tag{D.2}$$

for $i = 1, ..., m-1$, where

$$\frac{\partial \mathbf{G}_t^i}{\partial \delta_{ij}} \;=\; \text{diag}\left\{ \; 0 \; , \; ... \; , \; \frac{\partial g_t^{ij}}{\partial \delta_{ij}} \; , ... \; , \; 0 \; \right\}, \tag{D.3}$$

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

$$\frac{\partial g_t^{ij}}{\partial \gamma_{ij}} = (g_t^{ij})^2 \exp\{-\gamma_{ij}(s_t - c_{ij})\}(s_t - c_{ij}) = (s_t - c_{ij}) \; g_t^{ij}(1 - g_t^{ij}), \tag{D.4}$$

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

$$\frac{\partial g_t^{ij}}{\partial c_{ij}} = -(g_t^{ij})^2 \exp\{-\gamma_{ij}(s_t - c_{ij})\}\gamma_{ij} = -\gamma_{ij} \; g_t^{ij}(1 - g_t^{ij}). \tag{D.5}$$

Finally,

$$\frac{\partial \mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t}{\partial \mathbf{\Omega}} = \mathbf{0}. \tag{D.6}$$

The dimension of the first-order derivative of $\mathbf{\Psi}_t'\mathbf{B}'\mathbf{x}_t$ with respect to θ is $p \times [(kp+q)mp + 2(m-1)p]$.

E Figures

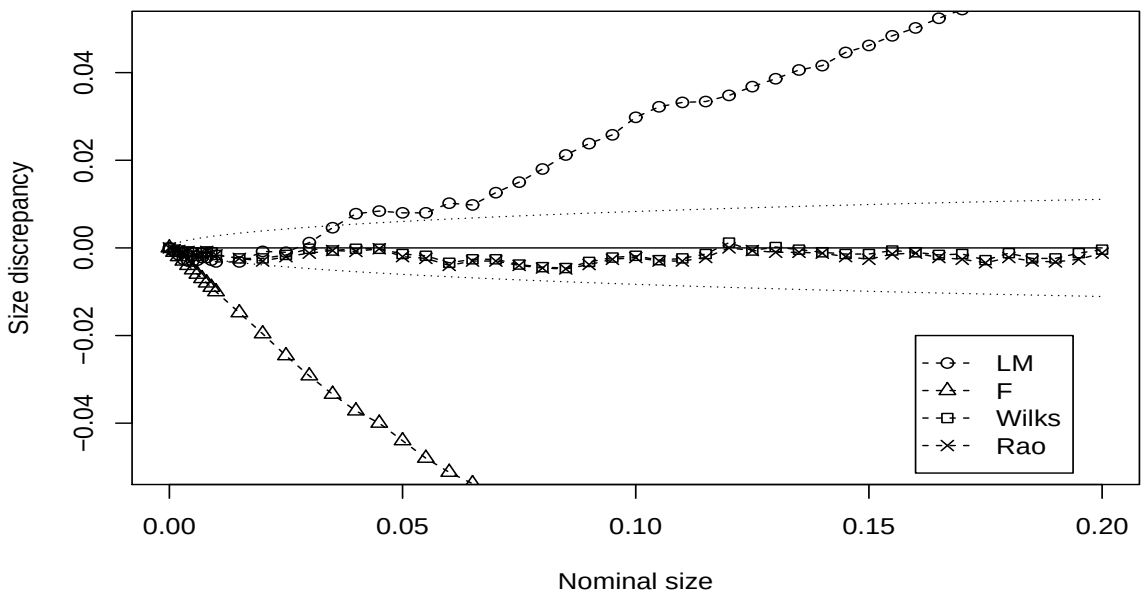

Figure 1: The size discrepancy plot for Design 1: $p = 2$, $k = 1$ and $T = 30$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

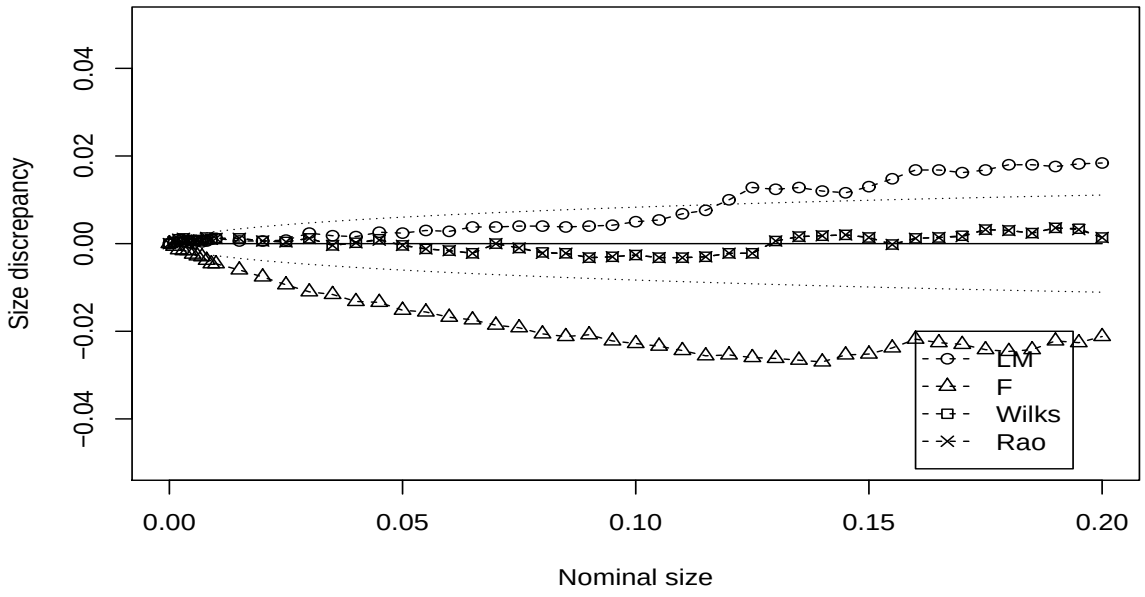

Figure 2: The size discrepancy plot for Design 2: $p = 2$, $k = 1$ and $T = 100$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

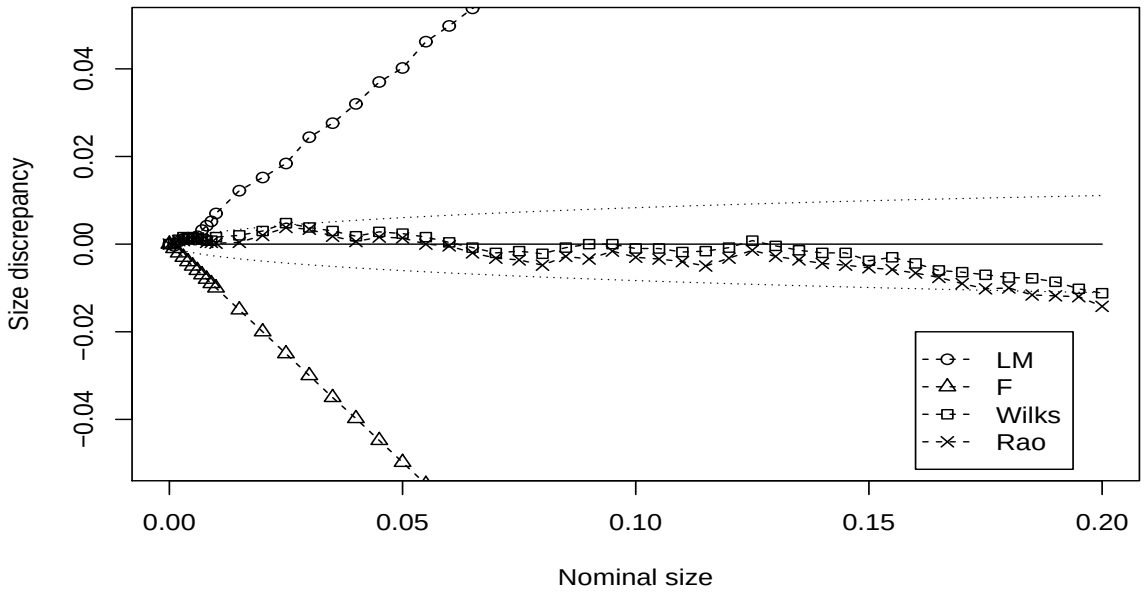

Figure 3: The size discrepancy plot for Design 3: $p = 2$, $k = 2$ and $T = 30$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

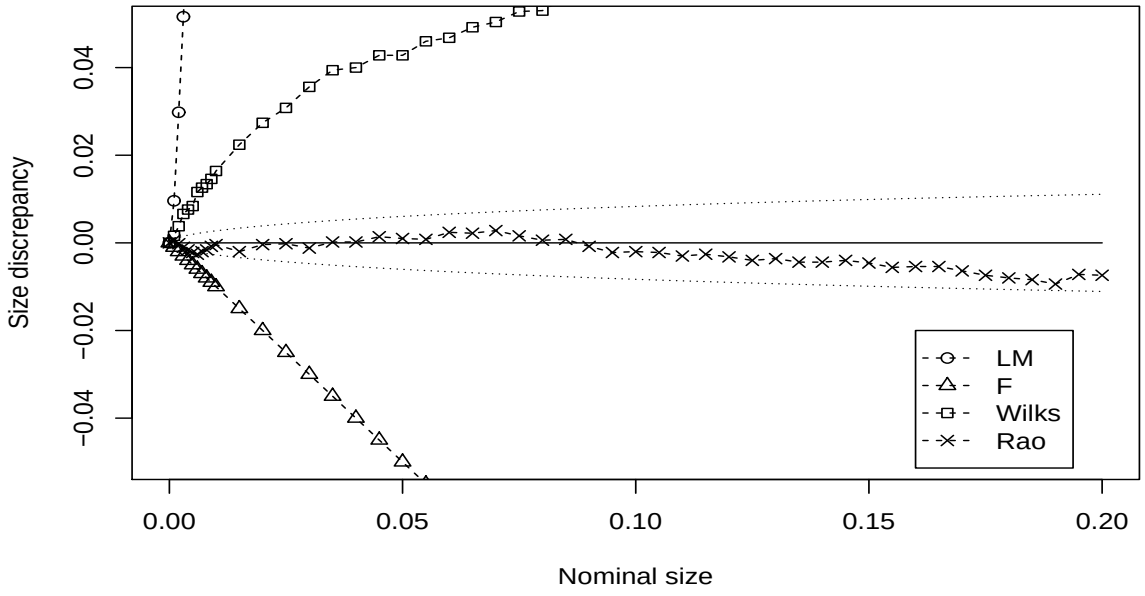

Figure 4: The size discrepancy plot for Design 4: $p = 2$, $k = 5$ and $T = 30$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

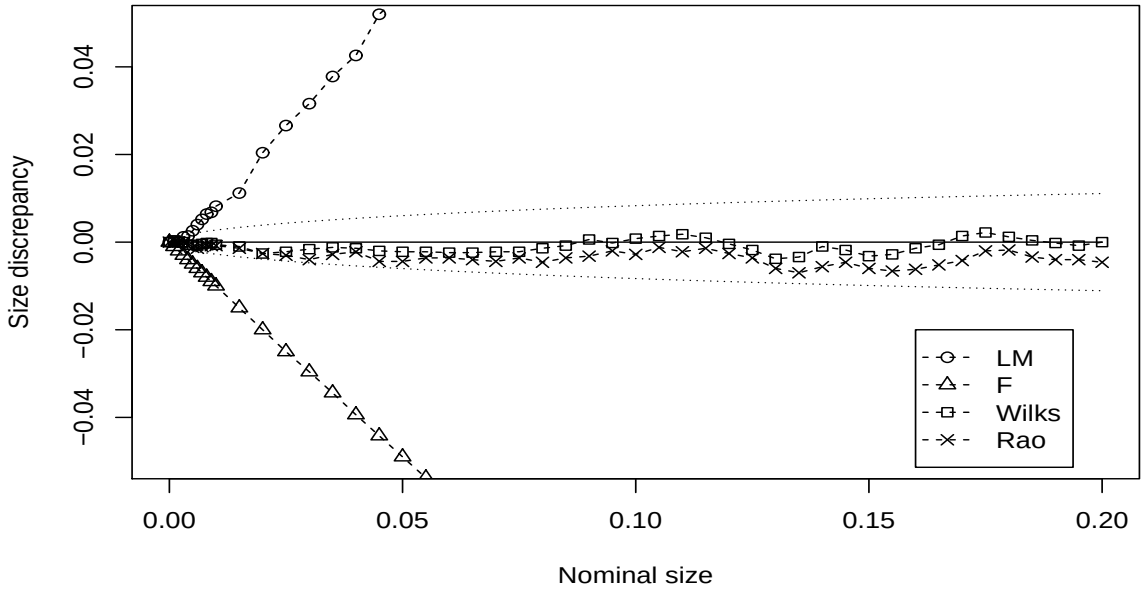

Figure 5: The size discrepancy plot for Design 5: $p = 5$, $k = 1$ and $T = 50$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

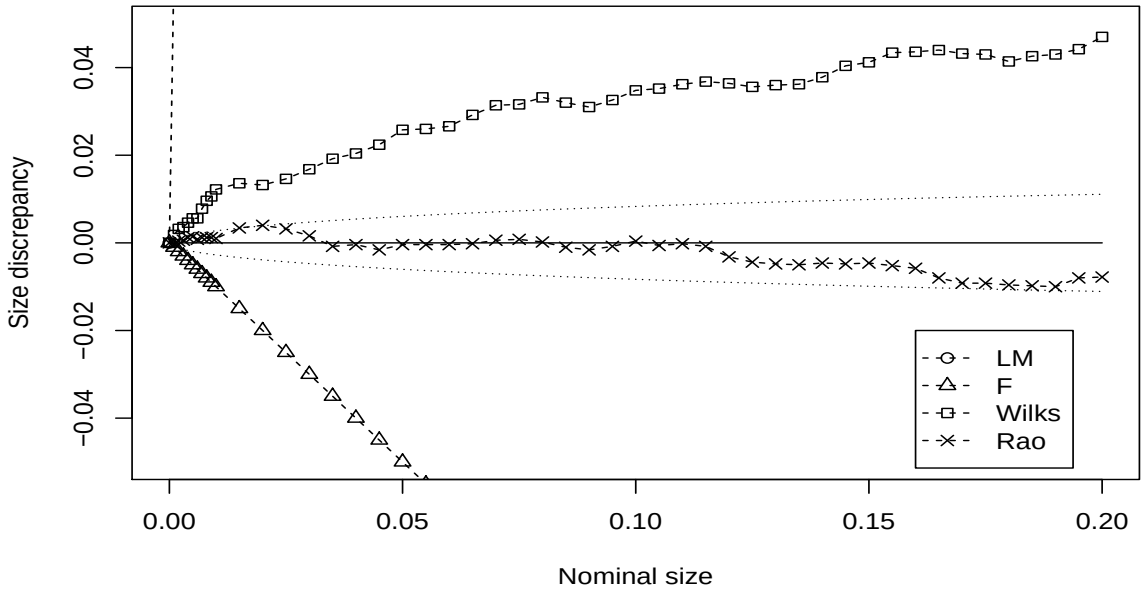

Figure 6: The size discrepancy plot for Design 6: $p = 10$, $k = 1$ and $T = 50$. The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

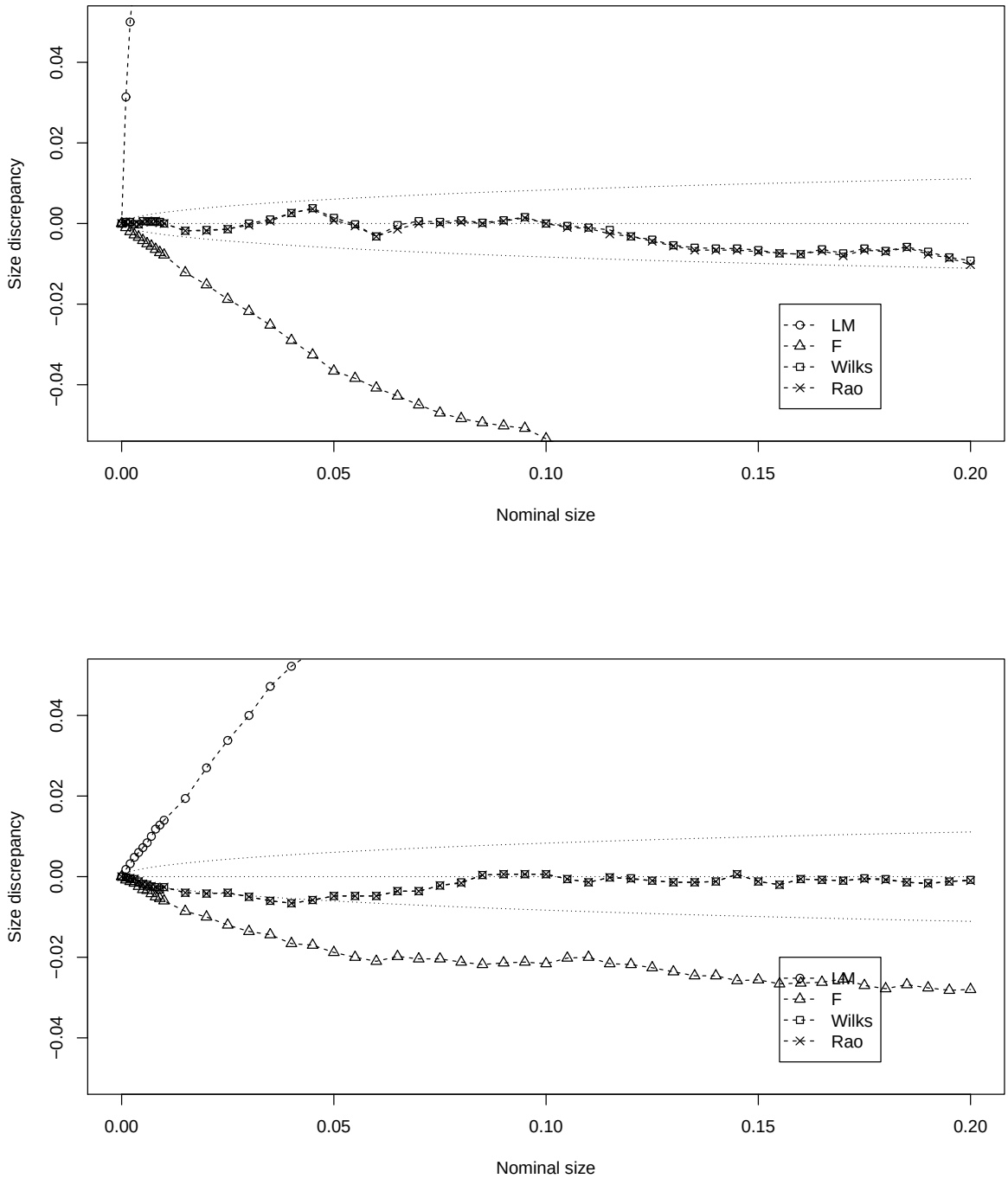

Figure 7: Size discrepancy plot for tests of no serial correlation: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

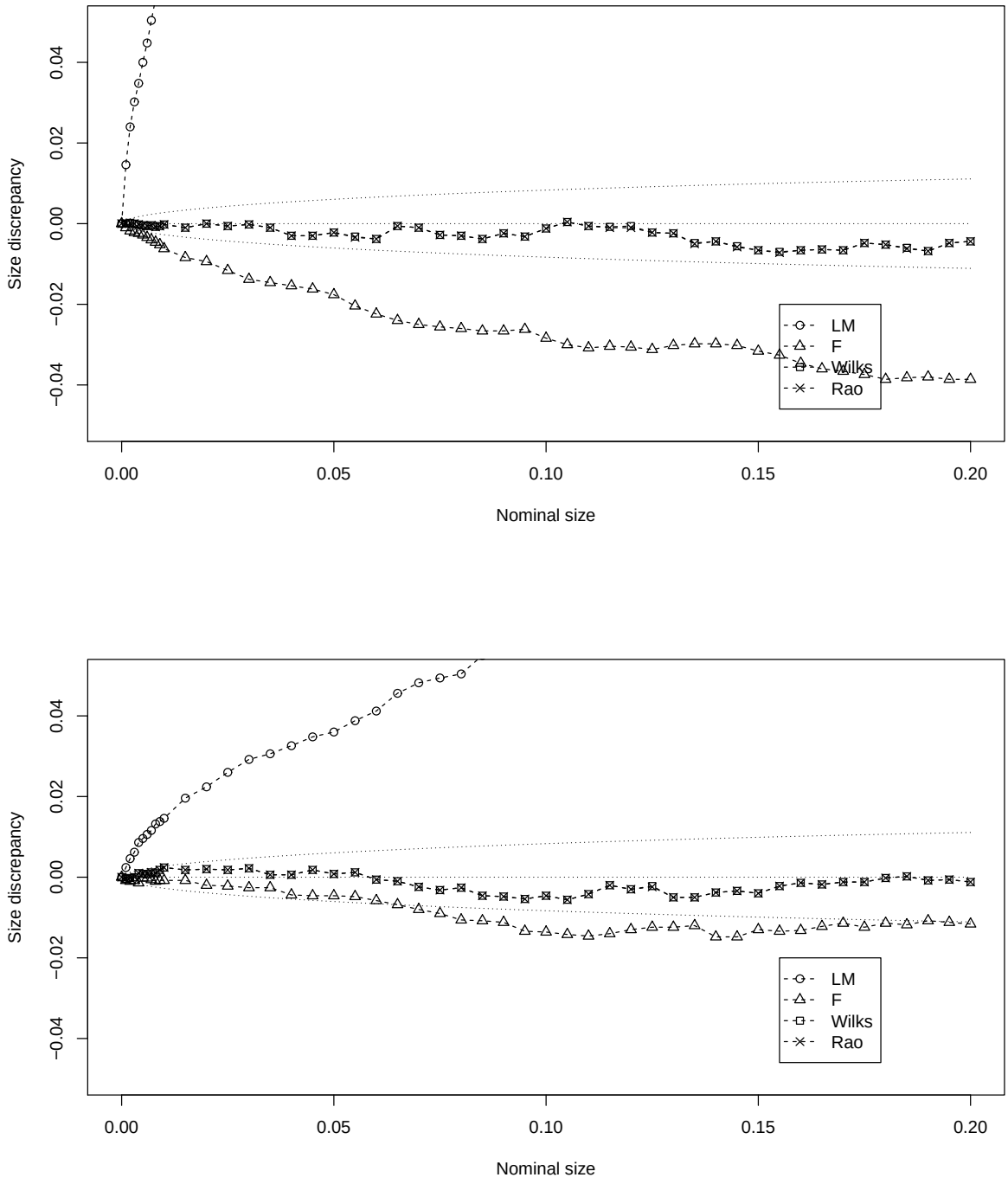

Figure 8: Size discrepancy plot for tests of no additive nonlinearity: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

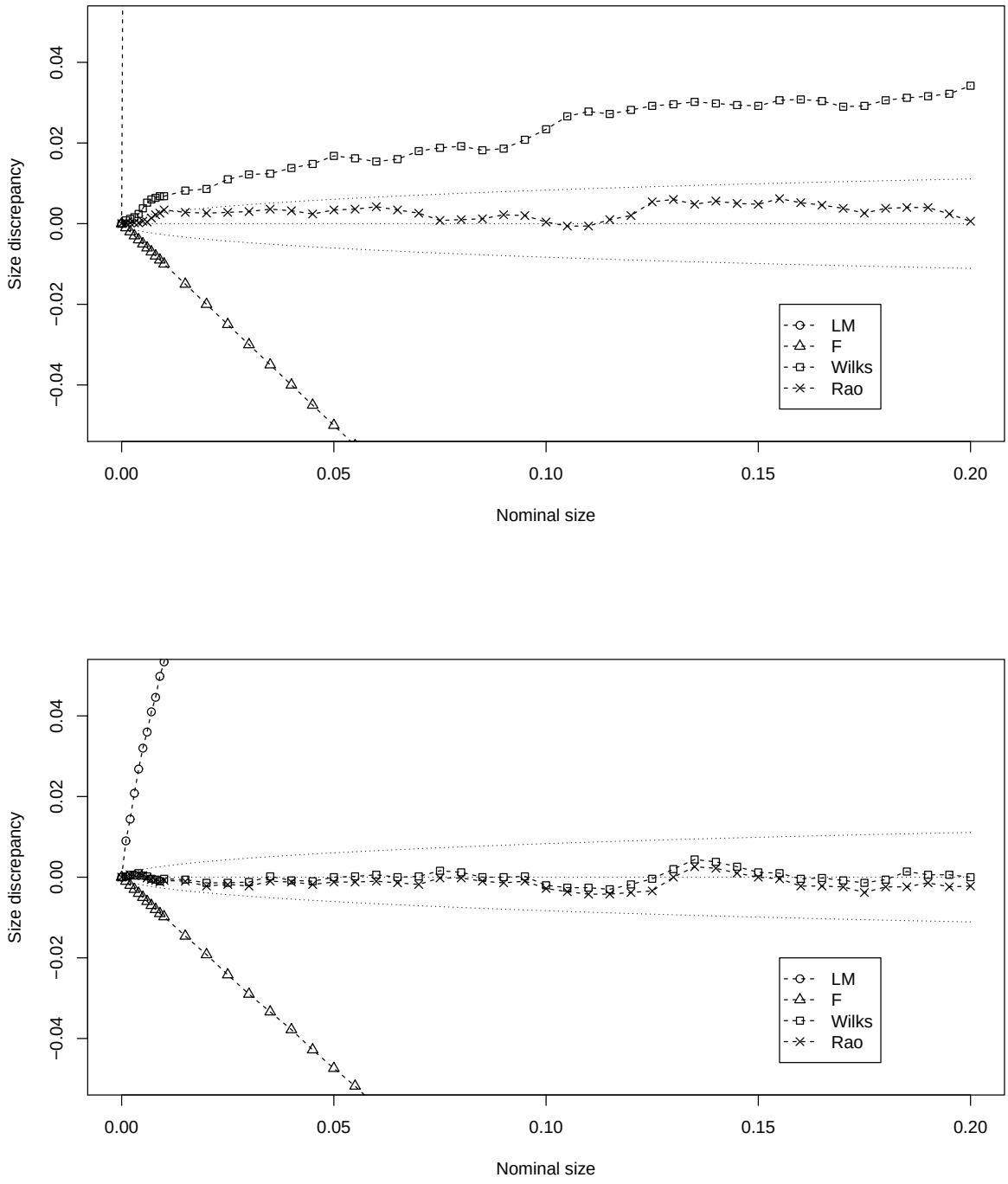

Figure 9: Size discrepancy plot for tests of parameter constancy: $p = 2$, $k = 2$ and $T = 200$ (Top); $T = 500$ (Bottom). The dotted lines represent the upper 95% confidence bound (top), zero line (middle) and the lower 95% confidence bound (bottom).

Recent titles

CORE Discussion Papers

2014/19 Paul BELLEFLAMME and Martin PEITZ. Digital piracy: an update.
2014/20 Eva-Maria SCHOLZ. Licensing to vertically related markets.
2014/21 N. Baris VARDAR. Optimal energy transition and taxation of non-renewable resources.
2014/22 Benoît DECERF. Income poverty measures with relative poverty lines.
2014/23 Antoine DEDRY, Harun ONDER and Pierre PESTIEAU. Aging, social security design and capital accumulation.
2014/24 Biung-Ghi JU and Juan D. MORENO-TERNERO. Fair allocation of disputed properties.
2014/25 Nguyen Thang DAO. From agriculture to manufacture: How does geography matter ?
2014/26 Xavier Y. WAUTHY. From Bertrand to Cournot via Kreps and Scheinkman: a hazardous journey.
2014/27 Gustavo BERGANTIÑOS and Juan MORENO-TERNERO. The axiomatic approach to the problem of sharing the revenue from bundled pricing.
2014/28 Jean HINDRIKS and Yukihiro NISHIMURA. International tax leadership among asymmetric countries.
2014/29 Jean HINDRIKS and Yukihiro NISHIMURA. A note on equilibrium leadership in tax competition models.
2014/30 Olivier BOS and Tom TRUYTS. Auctions with prestige motives.
2014/31 Juan D. MORENO-TERNERO and Lars P. ØSTERDAL . Normative foundations for equity-sensitive population health evaluation functions.
2014/32 P. Jean-Jacques HERINGS, Ana MAULEON and Vincent VANNETELBOSCH. Stability of networks under Level-*K* farsightedness.
2014/33 Lionel ARTIGE, Laurent CAVENAILE and Pierre PESTIEAU. The macroeconomics of PAYG pension schemes in an aging society.
2014/34 Tanguy KEGELART and Mathieu VAN VYVE. A conic optimization approach for SKU rationalization.
2014/35 Ulrike KORNEK, Kei LESSMANN and Henry TULKENS. Transferable and non transferable utility implementations of coalitional stability in integrated assessment models.
2014/36 Ibrahim ABADA, Andreas EHRENMANN and Yves SMEERS. Endogenizing long-term contracts in gas market models.
2014/37 Julio DAVILA. Output externalities on total factor productivity.
2014/38 Diane PIERRET. Systemic risk and the solvency-liquidity nexus of banks.
2014/39 Paul BELLEFLAMME and Julien JACQMIN. An economic appraisal of MOOC platforms: business models and impacts on higher education.
2014/40 Marie-Louise LEROUX, Pierre PESTIEAU and Grégory PONTHIERE. Longévité différentielle et redistribution: enjeux théoriques et empiriques.
2014/41 Chiara CANTA, Pierre PESTIEAU and Emmanuel THIBAULT. Long term care and capital accumulation: the impact of the State, the market and the family.
2014/42 Gilles GRANDJEAN, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Whom are you talking with ? An experiment on credibility and communication structure.
2014/43 Julio DAVILA. The rationality of expectations formation.
2014/44 Florian MAYNERIS, Sandra PONCET and Tao ZHANG. The cleaning effect of minimum wages. Minimum wages, firm dynamics and aggregate productivity in China.
2014/45 Thierry BRECHET, Natali HRITONENKOVA and Yuri YATSENKO. Domestic environmental policy and international cooperation for global commons.
2014/46 Mathieu PARENTI, Philip USHCHEV and Jacques-François THISSE. Toward a theory of monopolistic competition.
2014/47 Takatoshi TABUCHI, Jacques-François THISSE and Xiwei ZHU. Does technological progress affect the location of economic activity?
2014/48 Paul CASTANEDA DOWER, Victor GINSBURGH and Shlomo WEBER. Colonial legacy, linguistic disenfranchisement and the civil conflict in Sri Lanka.

Recent titles

CORE Discussion Papers - continued

2014/49 Victor GINSBURGH, Jacques MELITZ and Farid TOUBAL. Foreign language learnings: An econometric analysis.
2014/50 Koen DECANCQ and Dirk NEUMANN. Does the choice of well-being measure matter empirically? An illustration with German data.
2014/51 François MANIQUET. Social ordering functions.
2014/52 Ivar EKELAND and Maurice QUEYRANNE. Optimal pits and optimal transportation.
2014/53 Luc BAUWENS, Manuela BRAIONE and Giuseppe STORTI. Forecasting comparison of long term component dynamic models for realized covariance matrices.
2014/54 François MANIQUET and Philippe MONGIN. Judgment aggregation theory can entail new social choice results.
2014/55 Pasquale AVELLA, Maurizio BOCCIA and Laurence A. WOLSEY. Single-period cutting planes for inventory routing problems.
2014/56 Jean-Pierre FLORENS and Sébastien VAN BELLEGEM. Instrumental variable estimation in functional linear models.
2014/57 Abdelrahaman ALY and Mathieu VAN VYVE. Securely solving classical networks flow problems.
2014/58 Henry TULKENS. Internal vs. core coalitional stability in the environmental externality game: A reconciliation.
2014/59 Manuela BRAIONE and Nicolas K. SCHOLTES. Construction of Value-at-Risk forecasts under different distributional assumptions within a BEKK framework.
2014/60 Jörg BREITUNG and Christian M. HAFNER. A simple model for now-casting volatility series.
2014/61 Timo TERASVIRTA and Yukai YANG. Linearity and misspecification tests for vector smooth transition regression models.

Books

V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.
M. FUJITA and J.F. THISSE (2013), *Economics of agglomeration: cities, industrial location, and globalization.* (2nd edition). Cambridge University Press.
J. HINDRIKS and G.D. MYLES (2013). *Intermediate public economics.* (2nd edition). MIT Press.
J. HINDRIKS, G.D. MYLES and N. HASHIMZADE (2013). *Solutions manual to accompany intermediate public economics.* (2nd edition). MIT Press.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.