



WORKING PAPER

9/2013

RESIDUAL-BASED INFERENCE FOR COMMON
NONLINEAR FEATURES

DAO LI
STATISTICS

ISSN 1403-0586

Residual-based Inference for Common Nonlinear Features

Dao Li*

First version: December 2012, Second verion: April 2013,

This version: August 12, 2013

Abstract

This paper investigates common nonlinear features (CNFs) in multivariate nonlinear autoregressive models via testing the residuals. A Wald-type test is proposed, and it is asymptotically Chi-square distributed. Simulation evidence is given to examine the finite-sample properties of the proposed test and, furthermore, to provide a bootstrap version of the test. In addition to the empirical size and power, a specification of the reduced-rank matrix coefficient is studied to measure the departure from the null of CNFs. As the model moves further from containing CNFs, the power of the test increases substantially. Bootstrap critical values are used to improve the empirical size and power, especially when the test has size distortions. To perform a bootstrap version of the test, an algorithm is also provided for the estimation using nonlinear reduced-rank regression (NRRR).

KEYWORDS: Common features, Nonlinearity, Residual-based test, Bootstrap test, Reduced-rank regression.

1. INTRODUCTION

We study a special class of vector smooth-transition autoregressive (VSTAR) models that contain common nonlinear features (CNFs) and examine whether the nonlinearity is removed by taking a linear combination of a group of nonlinear time series. Li and He (2013) has studied CNFs in VSTAR models via examining population residuals; therefore, the classical inference is valid. This paper studies

*Dao Li, Department of Statistics, Örebro University School of Business, Örebro, Sweden. The previous results have been presented at the 6th International Conference on CFE, Ovideo, December 2012.

CNFs in VSTAR models via examining the estimated residuals, concentrating on the inferences based on the residuals. When a residual-based diagnostic is used, one of the key factors in determining the asymptotic distribution of the test, which is constructed from a set of the parameters on which we want to make inferences, is substituting estimates for another set of the (nuisance) parameters. Therefore, the inferential statistic is generally different from the statistic when the nuisance parameters are known. One of the popular applications is the residual-based tests for cointegration provided by Phillips and Ouliaris (1990). Residual-based inference in the cointegration case has been developed by applying one of the standard unit root tests to the estimated residuals from the cointegrating regression model. However, the asymptotic distributions of these statistics for testing the unit root of the original time series are not applicable.

This paper intends to apply the residual-based inference to derive the limiting distribution of the estimators of the parameters in the residual regression for testing CNFs. Pierce (1982) provides a general and easily used result for the limiting effect of substituting estimators for nuisance parameters. Related work includes Randles (1982), Pierce and Schafer (1986) and Pagan (1986). Applications that might use these results include, for example, goodness-of-fit tests and residual-based testing. However, such results have been largely neglected until the 21st century. Vella and Verbeek (1999) have studied two-step estimation using related inferences for panel models and censored data. Fu et al. (2002) is one of these applications in the catalog of goodness-of-fit tests. Pena and Slate (2006) provided a procedure that only relies on the standardized residual vector to implement and test global validation of the model assumptions.

The issue of CNFs is related to previous studies of common features e.g. Engle and Kozicki (1993) and Anderson and Vahid (1998). Engle and Kozicki (1993) provided a general procedure for detecting common features similar to those used in cointegration by Engle and Granger (1987). This idea is straightforward to apply to common nonlinear features, but estimation and testing procedures must be developed for nonlinear models. Anderson and Vahid (1998) studied common

nonlinearities in a similar way in multivariate nonlinear models. Our definition of CNFs is based on the reduced-rank VSTAR model and is similar to the definition of cointegration in cointegrated-VAR models, in Engle and Kozicki (1993). The general case of more than one combination vectors is definitely worth considering and this paper is a first step towards this goal by developing a test for CNFs whose properties are straightforward to study.

Because the asymptotic distribution of a statistical test can provide a poor approximation in finite samples. A second aim of this paper is to provide a bootstrap version of our test for CNFs. The bootstrap critical values are obtained while relying on the estimation of the VSTAR model with CNFs, which is a nonlinear reduced-rank regression. Therefore, an algorithm extended from the solution for linear reduced-rank regression in, for example, Anderson (1951) and Reinsel and Velu (1998), is provided. Furthermore, in the VSTAR models with CNFs, the reduced-rank matrix can be specified to allow for a parameter in the matrix to measure how far the data departs from the null of CNFs. This departure is illustrated in the simulation experiments as well.

The remainder of this paper is organized as follows. Section 2 presents VSTAR models containing CNFs. Section 3 derives the limiting distribution of the proposed Wald-type test for CNFs. Section 4 proposes a bootstrap version of the test and an algorithm for NRRR. Section 5 illustrates the finite-sample properties and the bootstrap critical values of the proposed test by simulation studies, and Section 6 concludes. The proofs can be found in the Appendix.

2. VSTAR MODELS AND CNFS

Consider an $(n \times 1)$ vector of observations $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})' = (y_{1t}, \mathbf{y}'_{2t})'$ generated from a vector nonlinear autoregression model

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^p \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \left(\tilde{\boldsymbol{\mu}} + \sum_{k=1}^p \boldsymbol{\Gamma}_k \mathbf{y}_{t-k} \right) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c}) + \mathbf{v}_t \quad (2.1)$$

in which $\mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \mathbf{\Omega})$. Consider $\mathbf{y}_{-p}, \mathbf{y}_{-p+1}, \dots, \mathbf{y}_{-1}, \mathbf{y}_0$ to be fixed. Assume that the process \mathbf{y}_t is stationary and ergodic and that fourth-order moments exist and are finite. The ST function $G(\mathbf{s}_t; \gamma, \mathbf{c})$ is a bounded function on $[0, 1]$ and is at least fourth-order differentiable with respect to γ in a neighborhood of $\gamma = 0$. Given any value of $G(\mathbf{s}_t; \gamma, \mathbf{c})$ between zero and one, all of the roots of $|\mathbf{I}_n - \sum_{k=1}^p (\mathbf{\Phi}_k + \mathbf{\Gamma}_k G) z^k| = 0$ lie outside the unit circle.

CNFs in VSTAR models have been defined in Section 3 in Li and He (2013). When there are CNFs in \mathbf{y}_t , we can find a nonzero vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\mathbf{y}_t$ is linear. When CNFs exist, the relation of CNFs among \mathbf{y}_t can be expressed by the following equation:

$$y_{1t} = \mu_1 + \boldsymbol{\alpha}'\mathbf{y}_{2t} + \sum_{k=1}^p \phi'_{1k}\mathbf{y}_{t-k} + u_t \quad (2.2)$$

in which μ_1 , $\boldsymbol{\alpha}^*$ and ϕ_{1k} are constant parameters, and $u_t \sim i.i.d.N(0, \sigma_v^2)$. In this statement, there only exists one CNF vector $\boldsymbol{\alpha}$. (There are not any other linearly independent vectors that remove the nonlinearity). Generally, the rank of the $n \times (np + 1)$ matrix $\mathbf{\Gamma}^* = (\tilde{\boldsymbol{\mu}} \ \mathbf{\Gamma}_1 \ \dots \ \mathbf{\Gamma}_p)$ is required to be less than n for the presence of CNFs. If the rank of $\mathbf{\Gamma}^*$ equals $n - 1$, then the simplest case is (2.2). We construct a transformation of model (2.1) that specifies the matrix $\mathbf{\Gamma}^*$, to study how the model departs from the null of CNFs. The specification might be used for further study considering a matrix replacing the vector $\boldsymbol{\alpha}$. Let \mathbf{A} and \mathbf{B} be defined as

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{I}_{n-r} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{B}_{22} \end{pmatrix},$$

then

$$\mathbf{\Gamma}^* = \mathbf{A}\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{A}_2\mathbf{B}_{11} & \mathbf{A}_2\mathbf{B}_{12} + \mathbf{B}_{22} \end{pmatrix}, \quad (2.3)$$

where \mathbf{A}_2 is an $(n - r) \times r$ matrix, \mathbf{B}_{11} is an $r \times (np + 1 - n + r)$ matrix with full rank, \mathbf{B}_{12} is an $r \times (n - r)$ matrix and \mathbf{B}_{22} is an $(n - r) \times (n - r)$ matrix. When

$\mathbf{B}_{22} = \mathbf{0}$, $\mathbf{\Gamma}^*$ has a reduced rank and can be specified as \mathbf{AB} with $\mathbf{A} = (\mathbf{I}_r, \mathbf{A}'_2)'$ and $\mathbf{B} = (\mathbf{B}_{11}, \mathbf{B}_{12})$. When $\mathbf{B}_{22} \neq \mathbf{0}$, $\mathbf{\Gamma}^*$ has a full rank n . This specification allows for the parameters in \mathbf{B}_{22} that measure the departure from CNFs. Similar to Kleibergen and van Dijk (1994), such a specification is by no means unique; however, every parameterization contains a parameter that reflects a departure from CNFs.

Imposing a simple null hypothesis, $r = n - 1$, model (2.2) is reasonable for testing if there exist CNFs in \mathbf{y}_t . In this triangular representation, the regressor \mathbf{y}_{2t} is not orthogonal to the error term \tilde{v}_t , and then we rewrite (2.2) as follows in order that the nonlinear least square (NLS) estimation be applicable:

$$y_{1t} = \mu_1 + \boldsymbol{\alpha}^{*'} \mathbf{y}_{2t} + \sum_{k=1}^p \phi'_{1k} \mathbf{y}_{t-k} + \sum_{s=-l}^l \boldsymbol{\delta}'_s \mathbf{v}_{2t-s} + u_t \quad (2.4)$$

in which $\mathbf{v}_{2t} = \mathbf{y}_{2t} - \boldsymbol{\mu}_2 - \boldsymbol{\Phi}_2^* \mathbf{y}_{t-1} - (\tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\Gamma}_2^* \mathbf{y}_{t-1}) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c})$ when $p = 1$, and u_t is uncorrelated with \mathbf{v}_{2t-s} for $s = -l, -l + 1, \dots, l$. In the correctly specified model, $l = 0$ is only necessary to impose to correct the correlations between \mathbf{y}_{2t} and e_t because \mathbf{v}_t is not serially correlated. However, a model with leads and lags is more robust because the model is not easy to specify exactly. Although the order p can be moderately large, this solution does not require a large number of observations, because l is not necessary to be different from zero. In the case that the model is not specified correctly and p is not small, a small l would not cause much trouble because serial correlation usually shows up at short lags.

To be more general, we rewrite the VSTAR model (2.1) in matrix notation as

$$\mathbf{y}_t = \boldsymbol{\Phi}^* \mathbf{x}_t + \boldsymbol{\Gamma}^* \mathbf{z}_t + \mathbf{v}_t = \boldsymbol{\Phi}^* \mathbf{x}_t + \mathbf{ABz}_t + \mathbf{v}_t \quad (2.5)$$

in which $\boldsymbol{\Phi}^* = (\boldsymbol{\mu} \ \boldsymbol{\Phi}_1 \ \dots \ \boldsymbol{\Phi}_p)$, $\boldsymbol{\Gamma}^* = (\tilde{\boldsymbol{\mu}} \ \boldsymbol{\Gamma}_1 \ \dots \ \boldsymbol{\Gamma}_p) = \mathbf{AB}$, $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$ and $\mathbf{z}_t = \mathbf{x}_t G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c})$, where \mathbf{A} is an $n \times r$ matrix and \mathbf{B} is a $r \times (np + 1)$ matrix ($r \leq n$). When $r = n$, that $\boldsymbol{\Gamma}^*$ is of full row rank indicates that \mathbf{y}_t is modeled in a nonlinear VSTAR model, each of which presents nonlinearity, and \mathbf{y}_t does not

have common nonlinearities. When $r = 0$, \mathbf{y}_t is modeled in a linear VAR model in the sense that $\mathbf{\Gamma}^*$ is a zero matrix (every linear combination of \mathbf{y}_t is linear). When $0 < r < n$, $\mathbf{\Gamma}^*$ has reduced-rank r , and there are $n - r$ linearly independent combinations of \mathbf{y}_t that are linear; therefore, there are r CNFs. We multiply an $(n - r) \times n$ matrix $\mathbf{H} = (-\mathbf{A}_2, \mathbf{I}_{n-r})$ and an $r \times n$ matrix $(\mathbf{I}_r, \mathbf{0})$ on both sides of (2.5), respectively, and partition \mathbf{y}_t as $(\mathbf{y}'_{2t}, \mathbf{y}'_{1t})'$, where \mathbf{y}_{1t} is an $(n - r) \times 1$ vector and \mathbf{y}_{2t} is an $r \times 1$ vector, a triangular representation of (2.5) with r CNFs, which is given as follows:

$$\begin{aligned}\mathbf{y}_{1t} &= \mathbf{A}_2 \mathbf{y}_{2t} + \tilde{\mathbf{\Phi}}^* \mathbf{x}_t + \mathbf{HABz}_t + \mathbf{u}_t \\ \mathbf{y}_{2t} &= \mathbf{\Phi}_2^* \mathbf{x}_t + \mathbf{\Gamma}_2^* \mathbf{z}_t + \mathbf{v}_{2t}\end{aligned}\tag{2.6}$$

in which $\tilde{\mathbf{\Phi}}^* = \mathbf{H}\mathbf{\Phi}^*$, $\mathbf{HA} = \mathbf{0}$, $\mathbf{u}_t = \mathbf{Hv}_t$, $\mathbf{\Phi}_2^* = (\mathbf{0}, \mathbf{I}_r)\mathbf{\Phi}_2$, $\mathbf{\Gamma}_2^* = (\mathbf{0}, \mathbf{I}_r)\mathbf{\Gamma}_2$, and $\mathbf{v}_{2t} = (\mathbf{0}, \mathbf{I}_r)\mathbf{v}_t$. Here, (2.6) is equivalent to (2.5) when there exist CNFs ($0 < r < n$). \mathbf{AB} has a general definition in (2.3) when $r = n$. The proposed residual-based test in Section 3 can be used to test CNFs in (2.6) when the null hypothesis is $n - 1$ CNFs or $r = n - 1$ (there is one linearly independent linear combination) against the alternative of no CNFs or $r = n$.

3. ASYMPTOTIC TEST

Davidson and Mackinnon (1993) illustrates many estimation issues of nonlinear regression models. The results in Chapter 8 are applicable for (2.4), which states that NLS estimation is equivalent to maximum likelihood (ML) estimation when the error term is assumed to be normally and independently distributed with constant variance (Chapter 8.10) and the ML estimator has three properties of consistency (Theorem 8.2), asymptotic normality (Theorem 8.3) and asymptotic efficiency (Chapter 8.8).

Let $\boldsymbol{\theta}$ be the parameter vector that contains all of the parameters that appear in (2.4). Suppose that the parametric model is asymptotically identified, and the contributions $\{l_t(\boldsymbol{\theta})\}_{t=1}^{\infty}$ satisfy the regularity conditions of the weak uniform law

of large numbers. The following Proposition 1 is given by applying the results in Davidson and Mackinnon (1993).

Proposition 1. *Let $\hat{\boldsymbol{\theta}}_T$ denote an NLS or ML estimator¹ of the parameter $\boldsymbol{\theta}$ in (2.4) under the assumption that $u_t \sim i.i.d.N(0, \sigma_u^2)$. Under the conditions in Theorem 8.2-8.3 in Davidson and Mackinnon (1993), $\hat{\boldsymbol{\theta}}_T$ is consistent and asymptotically efficient, and $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$ converges to a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$ as $T \rightarrow \infty$.*

The assumption of an *i.i.d* distribution of the error term is suggested here because it is assumed in a linear VAR model and univariate STAR model in the literature. However, time series data could further require a model that has serially correlated errors. Note that Davidson and Mackinnon (1993) also provides more discussion about relaxing the assumption of having an independent and identical distribution for the error term, which is useful to make the proposed method widely applicable.

Suppose that \hat{u}_t denotes the residuals from (2.4). Consider a nonlinear regression for \hat{u}_t :

$$\hat{u}_t = \beta_0 + \beta_1 \hat{u}_{t-1} + \beta_2 \hat{u}_{t-1}^2 + e_t = \boldsymbol{\beta}' \hat{\mathbf{x}}_t + e_t, \quad (3.1)$$

in which $\boldsymbol{\beta}$ is a vector of parameters, $\hat{\mathbf{x}}_t = (1, \hat{u}_{t-1}, \hat{u}_{t-1}^2)'$ is a vector of functions of the parameters $\boldsymbol{\theta}$ in (2.4), and the error term e_t is *i.i.d.*. In this case, the inference based on the usual NLS estimation is generally invalid. However, it is natural to have the following Proposition 2 when $\boldsymbol{\theta}$ in (2.4) is known.

Proposition 2. *When $\boldsymbol{\theta}$ in (2.4) is known, the model for u_t would be*

$$u_t = \beta_0^* + \beta_1^* u_{t-1} + \beta_2^* u_{t-1}^2 + e_t = \boldsymbol{\beta}^{*'} \mathbf{x}_t + e_t^*, \quad (3.2)$$

which is from (3.1), replacing the estimated parameters with known values. Let $\hat{\boldsymbol{\beta}}_T^$ denote an NLS estimator of $\boldsymbol{\beta}^*$ in (3.2). Under the conditions stated before, the Central Limit Theorem applies, and $\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}_0^*)$ converges to a multivariate*

¹Refer to Cramér definition, MLEs of Type 2 in (8.12) in Davidson and Mackinnon (1993).

normal distribution $N(\mathbf{0}, \Sigma_\beta)$ as $T \rightarrow \infty$, where β_0^* is the asymptotic mean of $\hat{\beta}_T^*$, which is zero.

Proposition 2 is quite straightforward under regularity conditions. Imposing the relation between $\sqrt{T}(\hat{\beta}_T - \beta_0)$ as a function of $\hat{\theta}_T$ and $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ by Taylor expansion, the asymptotic inference of $\sqrt{T}(\hat{\beta}_T - \beta_0)$ can be obtained from the joint distribution of $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$. Aside from a reminder, $\sqrt{T}(\hat{\beta}_T - \beta_0)$ equals a linear combination of $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$. Pierce (1982) proposed a general approach for the asymptotic results of the statistic, which involves substituting estimates for unknown parameters θ without the ordinarily tedious calculations. To achieve this goal, asymptotic joint normality of $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is required, and the fact that the asymptotic marginal distributions of $\sqrt{T}(\hat{\beta}_T - \beta_0)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$ are independent makes the calculation of the asymptotic variance of $\sqrt{T}(\hat{\beta}_T - \beta_0)$ simpler using Pierce's idea because $\hat{\theta}_T$ is asymptotically efficient from Proposition 1. Moreover, the covariance between $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$, which is often difficult to obtain, is not necessary to know in the final result shown in Pierce (1982). However, we can analytically give the covariance matrix of $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$ in their joint distribution. Details are in the following Proposition 3.

Proposition 3. $\sqrt{T}(\hat{\beta}_T^* - \beta_0^*)$ and $\sqrt{T}(\hat{\theta}_T - \theta_0)$ are asymptotically jointly normally distributed, $N(\mathbf{0}, \Sigma)$, in which

$$\Sigma = \begin{pmatrix} \sigma_e^2 \Omega_3^{-1} & -\Omega_3^{-1} \Omega_4 \mathcal{H}^{-1}(\theta_0) \\ -\mathcal{H}^{-1}(\theta_0) \Omega_4' \Omega_3^{-1} & -\mathcal{H}^{-1}(\theta_0) \end{pmatrix},$$

where $\Omega_3 = \text{plim}(\sum \mathbf{x}_t \mathbf{x}_t' / T)$ and $\Omega_4 = \text{plim}(\sum \mathbf{x}_t \frac{\partial u_t}{\partial \theta'} / T)$.

Proof of Proposition 3. See Appendix A. □

Theorem 1 below illustrates how to obtain the asymptotic distribution of $\hat{\beta}_T$ by applying the results in Pierce (1982) to residual-based inference. Note that the techniques in Theorem 1 are applicable to different forms of the regression of \hat{u}_t

besides the polynomial model (3.1). However, different forms affect the variance-covariance matrix in (3.3).

Theorem 1. *Consider (2.4) and (3.1). If the conditions in Proposition 1 hold, then the NLS estimator of $\boldsymbol{\beta}$ in (3.1) when $\boldsymbol{\theta}$ in (2.4) is estimated (denoted $\hat{\boldsymbol{\beta}}_T$) and satisfies*

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \sigma_e^2 \boldsymbol{\Omega}_3^{-1} + \mathbf{B} \mathcal{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{B}'), \quad (3.3)$$

in which $\mathcal{H}(\boldsymbol{\theta}_0)$ is the limiting Hessian matrix, $\mathbf{B} = \boldsymbol{\Omega}_3^{-1} \boldsymbol{\Omega}_4$, $\boldsymbol{\Omega}_3 = \text{plim}(\sum \mathbf{x}_t \mathbf{x}_t' / T)$ and $\boldsymbol{\Omega}_4 = \text{plim}(\sum \mathbf{x}_t \frac{\partial \hat{u}_t}{\partial \boldsymbol{\theta}'} / T)$.

Proof of Theorem 1. See Appendix A. □

The powerful aspect of the approach of Pierce's paper is that it is quite general. Propositions 1-3 are essentially regularity conditions, in which the only critical aspect is that the estimator of the nuisance parameter $\boldsymbol{\theta}$ is efficient, without which the theory fails. The approach is general in not needing many specific conditions, but it is limited to where everything is asymptotically normal.

To test whether the residual \hat{u}_t is linear or nonlinear, the null hypothesis is set up as $H_0 : \mathbf{R} \boldsymbol{\beta} = \mathbf{0}$, where \mathbf{R} is the corresponding constraint in the null hypothesis.

From Theorem 1, it is straightforward to obtain the asymptotic distribution of the estimator of the test statistic

$$T \mathbf{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \{ \mathbf{R}(\sigma_e^2 \boldsymbol{\Omega}_3^{-1} + \mathbf{B} \mathcal{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{B}') \mathbf{R}' \}^{-1} (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)' \mathbf{R}', \quad (3.4)$$

which is an asymptotical Chi-square distribution. The number of degrees of freedom is the number of restrictions under the null. In practice, σ_e^2 and \mathbf{B} can be consistently estimated by $\hat{\sigma}_e^2 = \sum \hat{u}_t^2 / (T - 1 - m)$ (m is the number of parameters in (3.1)) and $\hat{\mathbf{B}} = \hat{\boldsymbol{\Omega}}_3^{-1} \hat{\boldsymbol{\Omega}}_4$ with $\hat{\boldsymbol{\Omega}}_3 = \sum \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t' / (T - 1)$ and $\hat{\boldsymbol{\Omega}}_4 = \sum \hat{\mathbf{x}}_t \frac{\partial \hat{u}_t}{\partial \boldsymbol{\theta}'} / (T - 1)$, which is stated at the end of the proof of Theorem 1.

4. BOOTSTRAP TEST

The bootstrap test of CNFs is a version of (3.4) that uses bootstrap critical values instead of asymptotic critical values to improve finite-sample properties of the test. The idea of the bootstrap test is to treat the estimated parameters as the population values in the Monte Carlo experiments.

The bootstrap version of our test requires the estimated parameters in the restricted VSTAR model. However, the VSTAR model containing CNFs is a type of reduced-rank regression (RRR). The estimation of RRR is given by minimizing the sum of squared residual subjects to a reduced-rank restriction on the coefficient matrices. The problem is not a simple least squares estimation. The solution of linear RRR is related to canonical analysis and leads to the solution of a generalized eigenvalue problem, see, for example, Anderson (1951) and Reinsel and Velu (1998). In our nonlinear case, we apply and extend the RRR algorithm by imposing a grid search optimization of the slope and location parameters (γ, \mathbf{c}) . Given (γ, \mathbf{c}) , the VSTAR model is reduced as a linear model, where the RRR algorithm could be directly applied. Consider the following VSTAR(1) model with a reduced-rank matrix coefficient as an example without loss of generality,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + (\tilde{\boldsymbol{\mu}} + \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1}) G(\mathbf{s}_t; \gamma, \mathbf{c}) + \mathbf{v}_t \\ &= \boldsymbol{\Phi}^* \mathbf{x}_t + \boldsymbol{\Gamma}^* \mathbf{z}_t + \mathbf{v}_t, \end{aligned} \quad (4.1)$$

where $\mathbf{x}_t = (1, \mathbf{y}'_{t-1})'$, $\mathbf{z}_t = \mathbf{x}_t G(\mathbf{s}_t; \gamma, \mathbf{c})$, $\boldsymbol{\Phi}^* = (\boldsymbol{\mu}, \boldsymbol{\Phi}_1)$, $\boldsymbol{\Gamma}^* = (\tilde{\boldsymbol{\mu}}, \boldsymbol{\Gamma}_1)$, and $\boldsymbol{\Gamma}^* = \mathbf{A}\mathbf{B}$ has a reduced rank. Imposing the grid search of (γ, \mathbf{c}) , the RRR algorithm is updated as follows:

- (i) Given a set of (γ, \mathbf{c}) , calculate sample variance-covariance matrices

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{yz.x} &= \hat{\boldsymbol{\Sigma}}_{yz} - \hat{\boldsymbol{\Sigma}}_{yx} \hat{\boldsymbol{\Sigma}}_{xx}^{-1} \hat{\boldsymbol{\Sigma}}_{xz}, \\ \hat{\boldsymbol{\Sigma}}_{zz.x} &= \hat{\boldsymbol{\Sigma}}_{zz} - \hat{\boldsymbol{\Sigma}}'_{xz} \hat{\boldsymbol{\Sigma}}_{xx}^{-1} \hat{\boldsymbol{\Sigma}}_{xz}, \end{aligned}$$

where $\hat{\boldsymbol{\Sigma}}_{yz} = T^{-1} \sum_{t=1}^T \mathbf{y}_t \mathbf{z}'_t$, $\hat{\boldsymbol{\Sigma}}_{xx} = T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t$, $\hat{\boldsymbol{\Sigma}}_{xz} = T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{z}'_t$, $\hat{\boldsymbol{\Sigma}}_{zz} =$

$T^{-1} \sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t'$. Then, regress \mathbf{y}_t on \mathbf{x}_t and \mathbf{z}_t to obtain the full-rank OLS residual $\hat{\mathbf{v}}_t$, and calculate $\hat{\Sigma}_{\mathbf{v}} = T^{-1} \sum_{t=1}^T \hat{\mathbf{v}}_t \hat{\mathbf{v}}_t'$.

(ii) Find the eigenvalues of the matrix $\hat{\Sigma}_{\mathbf{v}}^{-1/2} \hat{\Sigma}_{yz.x} \hat{\Sigma}_{zz.x}^{-1} \hat{\Sigma}'_{yz.x} \hat{\Sigma}_{\mathbf{v}}^{-1/2}$ with the ordered eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_n$. The associated eigenvectors of the r largest eigenvalues are denoted by $\hat{\zeta}_1, \hat{\zeta}_2, \dots, \hat{\zeta}_r$. Normalize those eigenvectors in such a way that $\hat{\zeta}_i' \hat{\zeta}_i = 1$ ($i = 1, 2, \dots, r$).

(iii) Calculate the estimators of RRR, $\hat{\mathbf{A}} = \hat{\Sigma}_{\mathbf{v}}^{1/2} (\hat{\zeta}_1, \dots, \hat{\zeta}_r)$,

$\hat{\mathbf{B}} = (\hat{\zeta}_1, \dots, \hat{\zeta}_r)' \hat{\Sigma}_{\mathbf{v}}^{-1/2} \hat{\Sigma}_{yz.x} \hat{\Sigma}_{zz.x}^{-1}$, $\hat{\Gamma}^* = \hat{\mathbf{A}} \hat{\mathbf{B}}$, and $\hat{\Phi}^* = (\hat{\Sigma}_{yx} - \hat{\mathbf{A}} \hat{\mathbf{B}} \hat{\Sigma}_{zx}) \hat{\Sigma}_{xx}^{-1}$.

(iv) Calculate the criterion

$$tr \left\{ \hat{\Sigma}_{\mathbf{v}}^{-1/2} \frac{1}{T} \sum_{t=1}^T \left(\mathbf{y}_t - \hat{\Phi}^* \mathbf{x}_t - \hat{\mathbf{A}} \hat{\mathbf{B}} \mathbf{z}_t \right) \left(\mathbf{y}_t - \hat{\Phi}^* \mathbf{x}_t - \hat{\mathbf{A}} \hat{\mathbf{B}} \mathbf{z}_t \right)' \hat{\Sigma}_{\mathbf{v}}^{-1/2} \right\}. \quad (4.2)$$

(v) For different (γ, \mathbf{c}) in the grid search set, repeat (i)-(iv) above and compare (4.2). Choose the value of (γ, \mathbf{c}) that leads to the minimum value of (4.2) as the estimators $\hat{\gamma}$ and $\hat{\mathbf{c}}$.

For each bootstrap replication, a simulated data set $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ is recursively generated from (4.1) via the simulated error term $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_T$ and is used to estimate the parameters in the restricted VSTAR model and the covariance matrix of \mathbf{v}_t by the algorithm above. The disturbance term \mathbf{v}_t^* in the bootstrap step is randomly drawn from the normal distribution with a zero mean, and the covariance matrix is estimated from the sample covariance of the simulated \mathbf{v}_t ($t = 1, \dots, T$). Then, a bootstrap data set is generated recursively from the estimated VSTAR model, where the starting values $\mathbf{y}_{t-p}^b, \mathbf{y}_{t-p+1}^b, \dots, \mathbf{y}_{t-1}^b$ are generated by randomly drawn blocks of the simulated data $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$. More specifically, in each bootstrap replication, we first generate a data set from (4.1), in which \mathbf{v}_t is randomly drawn from $N(\mathbf{0}, \Sigma_{\mathbf{v}})$ with $\Sigma_{\mathbf{v}}$ given. We estimate model (4.1) by the proposed algorithm and estimate $\Sigma_{\mathbf{v}}$ from the residuals. Then, the bootstrap data set is generated from

$$\mathbf{y}_t^b = \hat{\boldsymbol{\mu}} + \hat{\Phi}_1 \mathbf{y}_{t-1}^b + \left(\hat{\boldsymbol{\mu}} + \hat{\Gamma}_1 \mathbf{y}_{t-1}^b \right) G(\mathbf{s}_t^b; \hat{\gamma}, \hat{\mathbf{c}}) + \mathbf{v}_t^* \quad (4.3)$$

in which \mathbf{v}_t^* is randomly drawn from a bivariate normal distribution with a mean of zero and the estimated covariance matrix is $\mathbf{v}_t, \hat{\Sigma}_v$. Our bootstrap test then can be obtained by Monte Carlo experiments using bootstrap samples, and we construct its quantiles as bootstrap critical values from the approximate distribution used in the following simulation studies.

5. FINITE-SAMPLE PROPERTIES

The Monte Carlo experiments in this subsection study the finite-sample properties of the Wald test in (3.4) and the bootstrap critical values. The desired sample sizes are $T = 100, 200, 500, 1000, 2000$. We discarded the first 500 out of $T + 500$ observations to minimize the initial effects from the simulated time series data in each experiment, and each experiment was performed 10,000 times.

The data generation process (DGP) for this size study follows

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \left(0.05 + \begin{pmatrix} 0.12 & -0.18 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \right) G(s_t; \gamma, c) + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \end{aligned} \quad (5.1)$$

$t = 1, 2, \dots, T$, in which $G(\mathbf{s}_t; \gamma, \mathbf{c}) = (1 + \exp\{-5(y_{2t-1} - 0.5)\})^{-1}$, and $(v_{1t}, v_{2t})'$ are generated from a bivariate normal distribution with a mean of zero and a covariance matrix

$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

In our experiments, different covariances and the F form of the test have been examined as well, but both of them have not substantially changed the finite-sample properties of the test. Therefore, the corresponding results are not present here.

The ML estimation of parameter $\boldsymbol{\theta}$ in (2.4) is achieved from

$$y_{1t} = \mu_1 + \alpha y_{2t} + \phi_{11} y_{1t-1} + \phi_{12} y_{2t-1} + \delta v_{2t}^* + u_t$$

where $v_{2t}^* = y_{2t} - \mu_2 - \phi_{21}y_{1t-1} - \phi_{22}y_{2t-1} - (\tilde{\mu}_2 + \Gamma_{21}y_{1t-1} + \Gamma_{22}y_{2t-1})/(1 + \exp\{-\gamma(y_{2t-1} - c)\})$, which is, after reparameterization, equivalent to

$$y_{1t} = \mu_1^* + \alpha^* y_{2t} + \phi_1^* y_{1t-1} + \phi_2^* y_{2t-1} \\ + (\tilde{\mu}_2^* + \Gamma_{21}^* y_{1t-1} + \Gamma_{22}^* y_{2t-1})/(1 + \exp\{-\gamma(y_{2t-1} - c)\}) + u_t.$$

The residual \hat{u}_t is obtained from the estimated model above. Consider model (3.1) of \hat{u}_t under the null of linearity. Because u_t is assumed to not have serial correlations, it is likely that \hat{u}_{t-1} is exogenous to \hat{u}_t . Moreover, zero-mean is assumed for u_t as well. Thereafter, to study the empirical size, it is reasonable to additionally exclude the constant term and the lagged regressor \hat{u}_{t-1} from (3.1), as follows:

$$\hat{u}_t = \beta_2 \hat{u}_{t-1}^2 + e_t. \quad (5.2)$$

In both cases, the null hypothesis is $H_0 : \beta_2 = 0$. (3.4) is then calculated with

$$\frac{\partial \hat{u}_t}{\partial \theta} = -(1, y_{2t}, y_{1t-1}, y_{2t-1}, \hat{G}, y_{1t-1} \hat{G}, y_{2t-1} \hat{G}, \\ (\hat{\mu}_2^* + \hat{\Gamma}_{21}^* y_{1t-1} + \hat{\Gamma}_{22}^* y_{2t-1}) \hat{G} (1 - \hat{G}) (y_{2t-1} - \hat{c}), \\ (\hat{\mu}_2^* + \hat{\Gamma}_{21}^* y_{1t-1} + \hat{\Gamma}_{22}^* y_{2t-1}) \hat{G} (\hat{G} - 1) \hat{\gamma})$$

and $\hat{G} = (1 + \exp\{-\hat{\gamma}(y_{2t-1} - \hat{c})\})^{-1}$ in this case.

In Table 1, the empirical size of the test using model (3.1) and asymptotic critical values is close to the nominal level, and the test using (5.2) has size distortions, which results in under-size estimates. Afterward, bootstrap critical values are imposed to improve the effective finite-sample size, especially when the test in (5.2) has size distortions. The results clearly indicate that the bootstrap version of our test is considerably more accurate than the asymptotic version when sample size is small.

To study the finite-sample power of our test, we consider two DGPs as follows,

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\ &+ \left(\begin{pmatrix} 0.1 \\ -0.16 \end{pmatrix} + \begin{pmatrix} 0.24 & -0.18 \\ 0.12 & -0.36 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \right) G(s_t; \gamma, c) + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \end{aligned} \quad (5.3)$$

and

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0.1G(s_t^*; \gamma, c) \\ -0.16G(s_t; \gamma, c) \end{pmatrix} + \begin{pmatrix} 0.24 & -0.18 \\ 0.12 & -0.36 \end{pmatrix} \begin{pmatrix} y_{1t-1}G(s_t^*; \gamma, c) \\ y_{2t-1}G(s_t; \gamma, c) \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \end{aligned} \quad (5.4)$$

in which (5.4) is actually an extension of the VSTAR model with different non-linearity for each series, which indicates that CNFs are not likely to occur in the system regardless of whether the coefficient matrix has a reduced rank or not. In the experiments, we set $s_t^* = y_{1t-1}$ and $s_t = y_{2t-1}$.

Size-adjusted power using both asymptotic critical values and bootstrap critical values are computed to eliminate the effects of possible size distortions. From Table 2, the power of the bootstrap version of our test has considerably improved compared to the asymptotic version, which has size distortions when sample size is small. It is not surprising that the bootstrap test does not improve much when the size estimate of our test is around the nominal level, but it still improves when the sample size is small, such as $T = 100$. Overall, the power using DGP (5.4) is better than the power using DGP (5.3). This indicates that our test has increasing power while the CNFs have a decreasing possibility of existing. As the model moves further from containing CNFs, the power of the test increases substantially.

Last, we examine the rejection rate of our test by varying the values of the parameters in \mathbf{B}_{22} . The DGP for this study, how the rejection rate changes to how much departure the DGP is from the null, is designed along with the specification

Table 1: Empirical size

Models	T	Asymptotic critical values				Bootstrap critical values			
		given significant levels				given significant levels			
		0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
(3.1)	200	0.1957	0.0954	0.0471	0.0091	0.2128	0.1039	0.0547	0.0107
	500	0.1956	0.0971	0.0481	0.0096	0.1981	0.0973	0.0500	0.0096
	1000	0.1977	0.0989	0.0506	0.0083	0.1993	0.1042	0.0501	0.0104
	2000	0.1929	0.0951	0.0507	0.0100	0.1982	0.0964	0.0508	0.0098
(5.2)	200	0.1095	0.0402	0.0145	0.0012	0.2106	0.1040	0.0543	0.0102
	500	0.1126	0.0421	0.0155	0.0011	0.1974	0.0988	0.0503	0.0099
	1000	0.1127	0.0450	0.0155	0.0009	0.1984	0.1045	0.0506	0.0098
	2000	0.1099	0.0442	0.0163	0.0013	0.1980	0.0951	0.0500	0.0099

Table 2: Empirical power

DGPs	T	Asymptotic critical values				Bootstrap critical values			
		given significant levels				given significant levels			
		0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
(3.1)									
(5.3)	200	0.1955	0.1006	0.0489	0.0108	0.2126	0.1088	0.0555	0.0118
	500	0.1941	0.0992	0.0508	0.0101	0.1976	0.0955	0.0519	0.0103
	1000	0.2005	0.0990	0.0497	0.0081	0.1964	0.1051	0.0514	0.0105
	2000	0.1926	0.0947	0.0505	0.0097	0.1986	0.0959	0.0505	0.0093
(5.4)	200	0.2853	0.1696	0.0987	0.0292	0.3047	0.1785	0.1105	0.0309
	500	0.4143	0.2785	0.1825	0.0666	0.4172	0.2797	0.1857	0.0671
	1000	0.6061	0.4655	0.3414	0.1531	0.5909	0.4539	0.3315	0.1522
	2000	0.8357	0.7285	0.6144	0.3690	0.8386	0.7308	0.6145	0.3663
(5.2)									
(5.3)	200	0.1955	0.1006	0.0489	0.0108	0.3082	0.1886	0.1208	0.0385
	500	0.1941	0.0992	0.0508	0.0101	0.2981	0.1754	0.1116	0.0374
	1000	0.2005	0.0990	0.0497	0.0081	0.2954	0.1841	0.1140	0.0362
	2000	0.1926	0.0947	0.0505	0.0097	0.2940	0.1763	0.1058	0.0346
(5.4)	200	0.2853	0.1696	0.0987	0.0292	0.4096	0.2783	0.1964	0.0823
	500	0.4143	0.2785	0.1825	0.0666	0.5245	0.3927	0.2988	0.1539
	1000	0.6061	0.4655	0.3414	0.1531	0.6845	0.5744	0.4669	0.2816
	2000	0.8357	0.7285	0.6144	0.3690	0.8859	0.8204	0.7438	0.5537

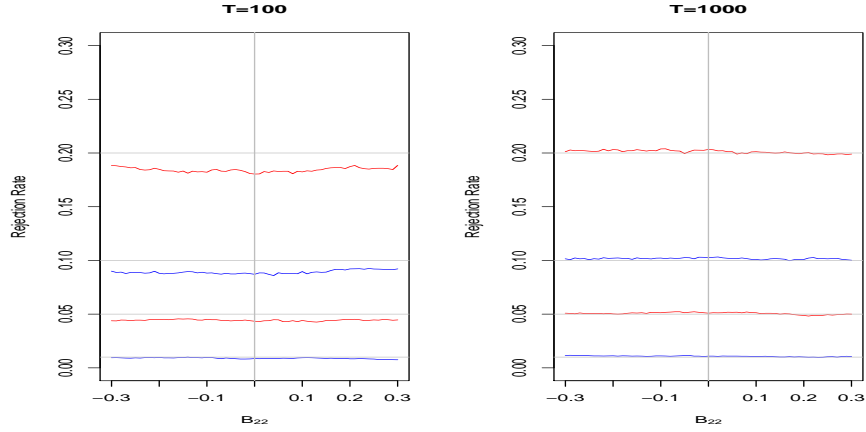


Figure 1: Rejection rates to different B_{22} , $T = 100$ and $T = 1000$.

of (2.3) as follows:

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\ &+ \left(\begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix} + \begin{pmatrix} 0.24 & -0.36 \\ 0.12 & -0.18 + B_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \right) G(s_t; \gamma, c) + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \\ \mathbf{A} &= \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}, \text{ and } \mathbf{B} = \begin{pmatrix} 0.1 & 0.24 & -0.36 \\ 0 & 0 & B_{22} \end{pmatrix}. \end{aligned}$$

In this experiment, we should be careful with the parameter settings from which the generated series are stationary. The values of B_{22} above is given from -0.3 to 0.3 for the stationary condition. When $B_{22} = 0$, y_{1t} and y_{2t} have CNFs. We can check, by this experiment, if the lack of power using the DGP in (5.3) depends on the particular parameter values that have been chosen. Figure 1 shows that the varying of B_{22} does not affect the lack of power.

6. CONCLUSIONS

This paper investigates CNFs in VSTAR models via testing the hypothesis in the residual regressions. The proposed Wald-type test has an asymptotic Chi-square distribution. The method for constructing the inferential statistics to test CNFs

can be naturally applied to other nonlinear models of residuals. As the model departs further from containing CNFs, the power of the test increases substantially. Imposing bootstrap critical values, for which an algorithm to estimate VSTAR with a reduced-rank matrix coefficient (due to CNFs) is provided, also moderately improves the finite-sample properties of the test. Our test will be applicable for determining the number of CNFs after being further developed. In addition to the empirical size and power, a specification of the reduced-rank matrix coefficient is studied to measure the departure of the data from the null of CNFs. However, when the test is lack of power in the experiments using the DGP in (5.3), the varying of the particular parameter values does not affect the lack of power.

Acknowledgments

Li is grateful to Sune Karlsson, Kenneth Carling and Johan Lyhagen for invaluable advice and thanks School of Technology and Business Studies at Högskolan Dalarna for financial support. Special thanks to Donald A. Pierce for his patient question-answering and advice.

References

- Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**(1): 1–36.
- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical statistics* **22**: 327–351.
- Davidson, R. and Mackinnon, J. G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica* **55**(2): 251–276.

- Engle, R. and Kozicki, S. (1993). Testing for common features, *Journal of Business and Economic Statistics* **11**(4): 369–380.
- Fu, B., Li, W.-K. and Fung, W.-K. (2002). Testing model adequacy for dynamic panel data with intercorrelation, *Biometrika* **89**(3): 591–601.
- Kleibergen, F. and van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* **10**(3/4): 514–551.
- Li, D. and He, C. (2013). Testing common nonlinear features in vector nonlinear autoregressive models, *Manuscript* .
- Pagan, A. (1986). Two stage and related estimators and their applications, *The Review of Economic Studies* **53**(4): 517–538.
- Pena, E. A. and Slate, E. H. (2006). Global validation of linear model assumptions, *Journal of the American Statistical Association* **101**(473): 341–354.
- Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration, *Econometrica* **58**: 165–93.
- Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics, *The Annals of Statistics* **10**(2): 475–478.
- Pierce, D. A. and Schafer, D. W. (1986). Residuals in generalized linear models, *Journal of the American Statistical Association* **81**(396): 977–986.
- Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters, *The Annals of Statistics* **10**(2): 462–474.
- Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression: Theory and Applications*, Springer, New York.
- Vella, F. and Verbeek, M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias, *Journal of Econometrics* **90**: 239–263.

White, H. (2001). *Asymptotic Theory for Econometricians. Revised edition.*,
Emerald Group Publishing Limited, UK.

APPENDIX

A. Proofs

Proof of Proposition 3. The estimates $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\beta}}_T^*$ are calculated by

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = (-T^{-1}\mathbf{H}(\bar{\boldsymbol{\theta}}))^{-1} (T^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0))$$

in which $\bar{\boldsymbol{\theta}}$ is a convex combination of $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$, $\mathbf{H}(\bar{\boldsymbol{\theta}})$ is the Hessian matrix and the gradient vector

$$\mathbf{g}(\boldsymbol{\theta}_0) = \begin{pmatrix} \sum_{t=1}^T \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_1} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \vdots \\ \sum_{t=1}^T \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix}$$

(suppose that $\boldsymbol{\theta}$ is k -dimensional) and

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}_0^*) = \left(T^{-1} \sum \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(T^{-1/2} \sum \mathbf{x}_t e_t \right).$$

It follows that

$$\begin{pmatrix} \sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}_0^*) \\ \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} T^{-1} \sum \mathbf{x}_t \mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & -T^{-1}\mathbf{H}(\bar{\boldsymbol{\theta}}) \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2} \sum \mathbf{x}_t e_t \\ T^{-1/2} \mathbf{g}(\boldsymbol{\theta}_0) \end{pmatrix} \\ \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Sigma})$$

in which

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega}_1^{-1} \boldsymbol{\Omega}_2 \boldsymbol{\Omega}_1^{-1} = \begin{pmatrix} \sigma_e^2 \boldsymbol{\Omega}_3^{-1} & -\boldsymbol{\Omega}_3^{-1} \boldsymbol{\Omega}_4 \mathcal{H}^{-1}(\boldsymbol{\theta}_0) \\ -\mathcal{H}^{-1}(\boldsymbol{\theta}_0) \boldsymbol{\Omega}_4' \boldsymbol{\Omega}_3^{-1} & -\mathcal{H}^{-1}(\boldsymbol{\theta}_0) \end{pmatrix},$$

where $\mathcal{H}(\boldsymbol{\theta}_0)$ is the limiting Hessian. Because

$$\begin{pmatrix} T^{-1} \sum \mathbf{x}_t \mathbf{x}'_t & \mathbf{0} \\ \mathbf{0} & -T^{-1} \mathbf{H}(\bar{\boldsymbol{\theta}}) \end{pmatrix} \xrightarrow{p} \boldsymbol{\Omega}_1 = \begin{pmatrix} \boldsymbol{\Omega}_3 & \mathbf{0} \\ \mathbf{0} & -\mathcal{H}(\boldsymbol{\theta}_0) \end{pmatrix}$$

and

$$E \begin{pmatrix} \mathbf{x}_t e_t \\ \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_1} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \vdots \\ \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_t e_t \\ \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_1} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \vdots \\ \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix}' = \begin{pmatrix} \sigma_e^2 \boldsymbol{\Omega}_3 & \boldsymbol{\Omega}_{4t} \\ \boldsymbol{\Omega}'_{4t} & -\mathcal{H}_t(\boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow{p} \boldsymbol{\Omega}_2,$$

where

$$\boldsymbol{\Omega}_3 = plim \left(T^{-1} \sum \mathbf{x}_t \mathbf{x}'_t \right)$$

and

$$\boldsymbol{\Omega}_{4t} = E \left(e_t \mathbf{x}_t \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)$$

with $(1/T) \sum_{t=1}^T \boldsymbol{\Omega}_{4t} \xrightarrow{p} \boldsymbol{\Omega}_4$ and $(1/T) \sum_{t=1}^T \mathcal{H}_t(\boldsymbol{\theta}_0) \xrightarrow{p} \mathcal{H}(\boldsymbol{\theta}_0)$ as $T \rightarrow \infty$. \square

Proof of Theorem 1. It follows from Proposition 1-3 that three conditions in Pierce (1982) are satisfied. Applying the result (1.3) in Pierce (1982), we obtain

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \sigma_e^2 \boldsymbol{\Omega}_3^{-1} + \mathbf{B} \mathcal{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{B}'),$$

in which $\mathbf{B} = \lim_{T \rightarrow \infty} E(\partial(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}_0^*)/\partial \boldsymbol{\theta}') = -\boldsymbol{\Omega}_3^{-1} \boldsymbol{\Omega}_4$.

To illustrate how to consistently estimate the covariance matrix in (3.3), we notice that

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) = \left(\frac{1}{T} \sum \hat{\mathbf{x}}_t \hat{\mathbf{x}}'_t \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum \hat{\mathbf{x}}_t e_t \right)$$

and correspondingly

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}_0^*) = \left(\frac{1}{T} \sum \mathbf{x}_t \mathbf{x}'_t \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum \mathbf{x}_t e_t^* \right),$$

where we assume further that, as $T \rightarrow \infty$,

$$\begin{aligned} \text{plim} \left(\frac{1}{T} \sum \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t' \right) &= \text{plim} \left(\frac{1}{T} \sum \mathbf{x}_t \mathbf{x}_t' \right) \\ &= \text{plim} \left(\frac{1}{T} \sum E(\hat{\mathbf{x}}_t \hat{\mathbf{x}}_t') \right) = \text{plim} \left(\frac{1}{T} \sum E(\mathbf{x}_t \mathbf{x}_t') \right) = \boldsymbol{\Omega}_3. \end{aligned}$$

Under the null $e_t^* = u_t$. Denote \mathbf{b}_T as $\sum \mathbf{x}_t u_t / T$.

$$\begin{aligned} E \left(\frac{\partial \mathbf{b}_T}{\partial \boldsymbol{\theta}'} \right) &= E \left(\frac{1}{T} \sum \left(\frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\theta}'} u_t + \mathbf{x}_t \frac{\partial u_t}{\partial \boldsymbol{\theta}'} \right) \right) \\ &= E \left(\frac{1}{T} \sum \frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\theta}'} u_t \right) + E \left(\frac{1}{T} \sum \mathbf{x}_t \frac{\partial u_t}{\partial \boldsymbol{\theta}'} \right) \\ &= E \left(E \left(\frac{1}{T} \sum \frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\theta}'} u_t \mid \mathbf{x}_t \right) \right) + E \left(\frac{1}{T} \sum \mathbf{x}_t \frac{\partial u_t}{\partial \boldsymbol{\theta}'} \right) \\ &= E \left(\frac{1}{T} \sum \mathbf{x}_t \frac{\partial u_t}{\partial \boldsymbol{\theta}'} \right), \end{aligned}$$

since $E \left(\sum \frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\theta}'} u_t / T \mid \mathbf{x}_t \right) = \mathbf{0}$. Thus,

$$\mathbf{B} = \boldsymbol{\Omega}_3^{-1} E \left(\frac{1}{T} \sum \mathbf{x}_t \frac{\partial u_t}{\partial \boldsymbol{\theta}'} \right)$$

in which

$$\begin{aligned} u_t &= y_{1t} - \mu_1 - \boldsymbol{\alpha}^{*'} \mathbf{y}_{2t} - \sum_{k=1}^p \phi_{1k}' \mathbf{y}_{t-k} \\ &\quad - \sum_{s=-l}^l \boldsymbol{\delta}'_s (\mathbf{y}_{2t-s} - \boldsymbol{\mu}_2 - \boldsymbol{\Phi}_{21} \mathbf{y}_{2t-1-s} - (\tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\Gamma}_{21} \mathbf{y}_{2t-1-s})) G(\mathbf{s}_{t-s}; \boldsymbol{\gamma}, \mathbf{c}) \\ &= y_{1t} + \mu_1^* + \sum_{k=1}^p \phi_{1k}^{*'} y_{1t-k} + \sum_{s=-l}^{\max\{p, l+1\}} \boldsymbol{\delta}'_s \mathbf{y}_{2t-s} + \sum_{s=-l}^l (\tilde{\boldsymbol{\mu}}_2^* + \boldsymbol{\Gamma}_{21}^* \mathbf{y}_{2t-1-s}) G(\mathbf{s}_{t-s}; \boldsymbol{\gamma}, \mathbf{c}) \end{aligned}$$

after reparameterized, and then

$$\begin{aligned} \frac{\partial u_t}{\partial \boldsymbol{\theta}'} = & (1, y_{1t-1}, \dots, y_{1t-p}, \mathbf{y}_{2t-\max\{p, (l+1)\}}, \dots, \mathbf{y}_{2t+l}, \\ & G(\mathbf{s}_{t-l}; \gamma, \mathbf{c}), \mathbf{y}_{2t-1-l} G(\mathbf{s}_{t-l}; \gamma, \mathbf{c}), g_{t-l}(\gamma), g_{t-l}(c), \\ & \vdots \\ & G(\mathbf{s}_{t+l}; \gamma, \mathbf{c}), \mathbf{y}_{2t-1+l} G(\mathbf{s}_{t+l}; \gamma, \mathbf{c}), g_{t+l}(\gamma), g_{t+l}(c))' \end{aligned}$$

where $g_{t-s}(\gamma) = (\tilde{\mu}_2^* + \mathbf{\Gamma}_{21}^* \mathbf{y}_{2t-1-s}) (1 + e^{-\gamma(s_t-s-c)})^{-2} (s_{t-s} - c) e^{-\gamma(s_t-s-c)}$ and $g_{t-s}(c) = (\tilde{\mu}_2^* + \mathbf{\Gamma}_{21}^* \mathbf{y}_{2t-1-s}) (-\gamma) (1 + e^{-\gamma(s_t-s-c)})^{-2} e^{-\gamma(s_t-s-c)}$.

\mathbf{B} can be consistently estimated as

$$\left(\frac{1}{T} \sum \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t' \right)^{-1} \left(\frac{1}{T} \sum \hat{\mathbf{x}}_t \hat{\mathbf{z}}_t \right)$$

by applying the law of large numbers for sequences of mixing random variables (see White (2001, Theorem 3.47 and 3.49)), where $\hat{\mathbf{z}}_t$ equals to the vector $\frac{\partial u_t}{\partial \boldsymbol{\theta}'}$ with all parameters replaced by their estimates.

□