ÖREBRO UNIVERSITY
SCHOOL OF BUSINESS

WORKING PAPER

8/2013

FORECASTING WITH VECTOR NONLINEAR TIME SERIES MODELS

DAO LI, CHANGLI HE
STATISTICS

ISSN 1403-0586

http://www.oru.se/Institutioner/Handelshogskolan-vid-Orebro-universitet/Forskning/Publikationer/Working-papers/
Örebro University School of Business
701 82 Örebro
SWEDEN

Forecasting with Vector Nonlinear Time Series Models

Dao Li*, Changli He†

First version: February 2012; Second version: October 2012,

This version: August 12, 2013

Abstract

In this study, vector nonlinear time series models are used for forecasting. Point forecasts are numerically obtained via bootstrapping. Our procedure is illustrated by two examples, each of which involves an application of macroeconomic data. Point forecast evaluation concentrates on forecast equality and encompassing. From these two applications, the forecasts from nonlinear models contribute useful information absent in the forecasts from linear models.

KEYWORDS: point forecast, forecast evaluation, nonlinearity.

1. Introduction

One of the primary concerns in macroeconomic time series data evaluation is the presence of nonlinearities. Because most existing linear time series models cannot approximate the nonlinear behavior of macroeconomic data, richer models are required. Modeling and forecasting using nonlinear models have been proposed more frequently in recent literature. To date, neither nonlinear models nor linear models can generally outperform after a large number of macroeconomic variables have been studied. However, combining forecasts obtained from different models has provided positive signs in several recent publications.

Point forecasts from a nonlinear parametric model usually do not have analytic solutions when we consider multi-step ahead forecasts. A Monte Carlo simulation

*Correspondence to: Dao Li, Örebro University School of Business, Örebro, Sweden. The first version has been presented at the 32nd International Symposium on Forecasting, Boston, June 2012.

†Changli He: School of Technology and Business Studies, Dalarna University, Borlänge, Sweden.

or bootstrapping method is usually applied in practice to approximate the integrations for the conditional expectations. Teräsvirta (2006, Chapter 8) illustrates a method to calculate the point forecasts from univariate nonlinear models, assuming that errors are independent. However, this paper addresses forecasting from vector nonlinear parametric models, in which several issues can be extended for empirical interest. For example, the correlations between variables should be included when multivariate models are involved, or a modification in numerical computation is necessary when serial correlations appear or endogenous variables are present. Block bootstrap is used in simple multivariate models, while model-based bootstrap is applied when serial correlations exist in the errors.

This study evaluates point forecasts, using two nonlinear models as examples. Forecast equality and encompassing are evaluated by applying several tests, following methods described in West (1996). The two examples are the nonlinear VAR model for stationary data and the nonlinear cointegration model for nonstationary data. An important empirical interest of nonstationary economic data is the purchasing power parity (PPP) system. The stationarity of the real exchange rate often fails in linear models. Imposing a nonlinear cointegration relation, this study investigates whether a nonlinear cointegration relation contributes to forecasting in the PPP system. For stationary systems, this study considers the growth of consumption and income instead of the levels themselves, to concentrate on the stationary nonlinear relation, thus cointegration is not possible.

The remainder of the paper is organized as follows. Section 2 describes the forecasting procedure using vector nonlinear time series models, including the two examples. Section 3 describes the accuracy measures used to evaluate the forecasts. After presenting two data sets and the modeling pre-treatment in Section 4, the forecasting results are presented in Section 5. The conclusions are presented in Section 6.

2. Point Forecasts with Nonlinear Models

2.1 Univariate forecasting

Teräsvirta (2006, Chapter 8) establishes a general framework for forecasting using a univariate nonlinear model. Assume a univariate time series y_t modeled in a nonlinear form as follows:

$$y_t = g(\mathbf{x}_{t-1}) + e_t, \tag{2.1}$$

where $g(\mathbf{x}_{t-1})$ is a nonlinear function of the regressors, $\mathbf{x}_{t-1}$, and the error term, $e_t \sim i.i.d.(0, \sigma^2)$. Let $\mathcal{F}_T$ denote the information set available up to time T and $f^y_{T+h|T} = E\left(y_{T+h}\,|\mathcal{F}_T\right)$ denote an h-step ahead point forecast of y_t, given $\mathcal{F}_T$. We summarize the procedure of forecasting from a univariate nonlinear model in the following.

When $h = 1$, the forecast of y_{T+1}, given $\mathcal{F}_T$, is analytically derived as follows:

$$f^y_{T+1|T} = E\left(y_{T+1}\,|\mathcal{F}_T\right) = E\left(g(\mathbf{x}_T) + e_{T+1}\,|\mathcal{F}_T\right) = g(\mathbf{x}_T),$$

because $E\left(e_{T+1}\,|\mathcal{F}_T\right) = 0$. However, when $h \geq 2$, the forecast of y_{T+h}, given $\mathcal{F}_T$, must usually be obtained numerically. Suppose that $\mathbf{x}_t$ is generated from the previous information, $\mathbf{x}_{t-1}$, and the independent random errors, $\boldsymbol{\eta}_t$, as $\mathbf{x}_t = g^*(\mathbf{x}_{t-1}) + \boldsymbol{\eta}_t$. For example, when $h = 2$, the two-step forecast, $E\left(y_{T+2}\,|\mathcal{F}_T\right)$, is defined as follows:

$$\begin{aligned} f^y_{T+2|T} =& E\left(y_{T+2}\,|\mathcal{F}_T\right) = E\left(g(\mathbf{x}_{T+1}) + e_{T+2}\,|\mathcal{F}_T\right) = E\left(g(\mathbf{x}_{T+1})\,|\mathcal{F}_T\right) && (2.2)\\ =& E\left(g(g^*(\mathbf{x}_T) + \boldsymbol{\eta}_{T+1})\,|\mathcal{F}_T\right) \\ =& \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(g^*(\mathbf{x}_T) + \boldsymbol{\eta}_{T+1}) dF(\boldsymbol{\eta}_{T+1}), && (2.3) \end{aligned}$$

where $F(\boldsymbol{\eta}_{T+1})$ is the joint cumulative function of the random variables in the vector $\boldsymbol{\eta}_{T+1}$. Because $\mathbf{x}_{T+1}$ in (2.2) is not known, given $\mathcal{F}_T$, the forecast of $g(\mathbf{x}_{T+1})$ depends on the distribution of $\boldsymbol{\eta}_{T+1}$. Consequently, it is necessary to approximate the integration in (2.3) numerically. The usual approach is to perform simulations for $\boldsymbol{\eta}_{T+1}$ by assuming a distribution of $\boldsymbol{\eta}_t$ or bootstrapping the residuals, $\hat{\boldsymbol{\eta}}_T$, to approximate the integrals in (2.3). As an example of the latter option, when the two-step ahead forecast is undertaken, the integral is approximated as follows:

$$\hat{f}^{y(b)}_{T+2|T} = \hat{E}\left(y_{T+2}\left|\mathcal{F}_T\right.\right) = \frac{1}{N}\sum_{j=1}^{N}\left(g(g^*(\mathbf{x}_T) + \hat{\boldsymbol{\eta}}^{(j)}_{T+1})\right)$$

where $\hat{\boldsymbol{\eta}}^{(j)}_{T+1}(j = 1, 2, ..., N)$ are independently drawn from the residuals $\hat{\boldsymbol{\eta}}_t$ $(t = 1, 2, ..., T)$ with replacement.

2.2 Multivariate forecasting

The forecasting procedure in Section 2.1 was developed for univariate nonlinear models with independent errors. However, the correlations between elements of the errors should also be considered when the forecasting procedure is extended to multivariate models. Serially correlated errors or endogenous variables, which can be important, should also be included. Thus, the forecasting procedure must be extended further. We will use two examples to illustrate both these issues. To be as general as possible, one of the examples is the forecast of stationary time series and the other is the forecast of nonstationary time series.

Starting from stationary variables, the autoregressive model is a popular class of time series models applied to dynamic structures. The nonlinear autoregressive models interest many modelers or forecasters for stationary data when a pattern of nonlinearity is likely. The following unrestricted and restricted vector smooth-transition autoregressive (VSTAR) models, (2.4) and (2.6), belong to the class from which we have the benefit of capturing nonlinear dynamics.

Suppose a class of vector nonlinear autoregressions is desired to forecast a group of stationary time series. Let $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t}, ..., y_{nt})'$ denote an $(n \times 1)$ vector time series. In this example, the following vector smooth-transition autoregression

(VSTAR) is considered:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \left(\tilde{\boldsymbol{\mu}} + \sum_{k=1}^{p} \boldsymbol{\Gamma}_k \mathbf{y}_{t-k} \right) G\left(\mathbf{s}_t; \gamma, \mathbf{c}\right) + \mathbf{v}_t. \tag{2.4}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Phi}_k$, $\tilde{\boldsymbol{\mu}}$, $\boldsymbol{\Gamma}_k$, γ and $\mathbf{c}$ are all parameters, and $G(\mathbf{s}_t; \gamma, \mathbf{c})$ is a logistic function of the transition variables $\mathbf{s}_t$ and $G(\mathbf{s}_t; \gamma, \mathbf{c}) = (1+\exp\{-\gamma \prod_{j=1}^{k} (s_{jt} - c_j)\})^{-1}$. The error term $\mathbf{v}_t$ is independent and identically distributed. However, the n elements in $\mathbf{v}_t$ are correlated. Assume that the process $\mathbf{y}_t$ is stationary and ergodic and that fourth-order moments exist and are finite. When $\mathbf{y}_t$ contains common nonlinear features (CNFs), the combination of $\mathbf{y}_t$ is reduced to a linear form, as follows:

$$\boldsymbol{\alpha}' \mathbf{y}_t = \boldsymbol{\alpha}' \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\alpha}' \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \boldsymbol{\alpha}' \mathbf{v}_t, \tag{2.5}$$

where $\boldsymbol{\alpha}' \tilde{\boldsymbol{\mu}} = \mathbf{0}$ and $\boldsymbol{\alpha}' \boldsymbol{\Gamma}_k = \mathbf{0}$ for all k. Therefore, the restricted VSTAR model can be rewritten as follows:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \boldsymbol{\alpha}_\perp \left(\tilde{\boldsymbol{\mu}}^* + \sum_{k=1}^{p} \boldsymbol{\beta}_k' \mathbf{y}_{t-k} \right) G\left(\mathbf{s}_t; \gamma, \mathbf{c}\right) + \mathbf{v}_t, \tag{2.6}$$

where $\boldsymbol{\alpha}_\perp \tilde{\boldsymbol{\mu}}^* = \tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\alpha}_\perp \boldsymbol{\beta}_k' = \boldsymbol{\Gamma}_k$, where $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_k$ and $\tilde{\boldsymbol{\mu}}^*$ are parameters. Multiplying both sides of (2.6) by a vector, $\boldsymbol{\alpha}$, results in (2.5) because there exists a nonzero vector $\boldsymbol{\alpha}$ satisfying $\boldsymbol{\alpha}' \boldsymbol{\alpha}_\perp = \mathbf{0}$ when $\mathbf{y}_t$ contains CNFs. Without loss of generality, we use the VSTAR model with order $p = 1$ for simplicity. Usually, the transition variable $\mathbf{s}_t$ is set as the lagged values of $\mathbf{y}_t$, denoted by $\mathbf{y}_{t-d}$, where the delay parameter, d, affects the forecasts at a different-step ahead. Because $\mathbf{y}_{T+h-d} \,|\mathcal{F}_T$ may or may not be knwon, depending on whether d is not less than or less than h.

The one-step ahead forecast of $\mathbf{y}_t$, given $\mathcal{F}_T$, is detailed as follows:

$$\begin{aligned} f^{\mathbf{y}}_{T+1|T} &= E\left(\mathbf{y}_{T+1} \,|\mathcal{F}_T\right) = E\left(\boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_T + \boldsymbol{\alpha}_\perp \left(\tilde{\boldsymbol{\mu}}^* + \boldsymbol{\beta}_1' \mathbf{y}_T\right) G\left(\mathbf{s}_{T+1}; \gamma, \mathbf{c}\right) + \mathbf{v}_{T+1} \,|\mathcal{F}_T\right) \\ &= \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_T + \boldsymbol{\alpha}_\perp \left(\tilde{\boldsymbol{\mu}}^* + \boldsymbol{\beta}_1' \mathbf{y}_T\right) E(G\left(\mathbf{s}_{T+1}; \gamma, \mathbf{c}\right) |\mathcal{F}_T), \end{aligned}$$

where $E(G(\mathbf{s}_{T+1};\gamma,\mathbf{c})\,|\mathcal{F}_T)$ is unobserved when $d=0$, while the expectation is trivial when $d>0$ and $G(\mathbf{s}_{T+1};\gamma,\mathbf{c})$ is observed. Similarly, the two-step ahead forecast of $\mathbf{y}_t$ is recursively derived as follows:

$$\begin{aligned}
f^{\mathbf{y}}_{T+2|T} =& E\left(\mathbf{y}_{T+2}\,|\mathcal{F}_T\right) \\
=& E\left(\boldsymbol{\mu}+\boldsymbol{\Phi}_1\mathbf{y}_{T+1}+\boldsymbol{\alpha}_\perp\left(\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\beta}_1'\mathbf{y}_{T+1}\right)G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)+\mathbf{v}_{T+2}\,|\mathcal{F}_T\right) \\
=& \boldsymbol{\mu}+\boldsymbol{\Phi}_1\boldsymbol{\mu}+\boldsymbol{\Phi}_1\boldsymbol{\Phi}_1\mathbf{y}_T+\boldsymbol{\Phi}_1\boldsymbol{\alpha}_\perp\left(\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\beta}_1'\mathbf{y}_T\right)G\left(\mathbf{s}_{T+1};\gamma,\mathbf{c}\right) \\
&+\boldsymbol{\alpha}_\perp\tilde{\boldsymbol{\mu}}^*E\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\mu}E\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right) \\
&+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\Phi}_1\mathbf{y}_TE\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'E\left(\mathbf{v}_{T+1}G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right) \\
&+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\alpha}_\perp\left(\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\beta}_1'\mathbf{y}_T\right)G\left(\mathbf{s}_{T+1};\gamma,\mathbf{c}\right)E\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right),
\end{aligned} \tag{2.7}$$

where $E\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)$ and $E\left(\mathbf{v}_{T+1}G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)$ need numerical approximations when $d=1$. However, when $d\geq 2$, $f^{\mathbf{y}}_{T+2|T}$ is reduced to the following:

$$\begin{aligned}
f^{\mathbf{y}}_{T+2|T} =& \boldsymbol{\mu}+\boldsymbol{\Phi}_1\boldsymbol{\mu}+\boldsymbol{\Phi}_1\boldsymbol{\Phi}_1\mathbf{y}_T+\boldsymbol{\Phi}_1\boldsymbol{\alpha}_\perp\left(\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\beta}_1'\mathbf{y}_T\right)G\left(\mathbf{s}_{T+1};\gamma,\mathbf{c}\right) \\
&+\left(\boldsymbol{\alpha}_\perp\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\mu}+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\Phi}_1\mathbf{y}_T\right)G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right) \\
&+\boldsymbol{\alpha}_\perp\boldsymbol{\beta}_1'\boldsymbol{\alpha}_\perp\left(\tilde{\boldsymbol{\mu}}^*+\boldsymbol{\beta}_1'\mathbf{y}_T\right)G\left(\mathbf{s}_{T+1};\gamma,\mathbf{c}\right)G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right),
\end{aligned}$$

because $G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)$ is observed at time T when $d\geq 2$.

Generally speaking, the h-step ahead forecast, $f^{\mathbf{y}}_{T+h|T}$, is obtained recursively in the same way as for $f^{\mathbf{y}}_{T+2|T}$. When $d\geq h$, $f^{\mathbf{y}}_{T+h|T}$ has an analytic form; however, numerical approaches are required when $d<h$. When the approximations of the conditional expectations, such as $E\left(\mathbf{v}_{T+1}G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)$, are numerically calculated via bootstrapping, individually drawing different elements from the corresponding residuals is not sufficient to capture the dependent structures between the elements in $\mathbf{v}_t$. A simple modification of re-sampling for the dependence in $\mathbf{v}_t$ is then used to calculate $f^{\mathbf{y}}_{T+h|T}$.

A class of time series models, complementary to the class above, is used for the vector unit root processes. The cointegrating regression models are a popular choice for series that have long-run equilibrium. We therefore apply the cointegrated processes from nonlinear cointegrating regression models to frecasting. Beyond introducing the nonlinear cointegrating regressions, we also define the nonlinear cointegration. Let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ be an $(n \times 1)$ vector time series in which each of the series is nonstationary. The vector time series $\mathbf{y}_t$ is nonlinearly cointegrated if there is an $(n \times 1)$ time-varying or random vector $\boldsymbol{\alpha}_t^* = (\alpha_{1t}, \alpha_{2t}, ..., \alpha_{nt})'$ such that the nonlinear combination of the series $\mathbf{y}_t$ is stationary. A formal definition can be found in Li and He (2013b). We normalize the cointegrating vector with a nonzero element to be unity and correspondingly partition $\boldsymbol{\alpha}_t^*$ as $(1, \boldsymbol{\alpha}_t')'$. The proposed nonlinear cointegrating regression model is then defined in a partitioned form for $\mathbf{y}_t = (y_{1t}, \mathbf{y}_{2t}')'$ as follows:

$$y_{1t} = a_1 + \delta t + \boldsymbol{\alpha}_t' \mathbf{y}_{2t} + u_t \tag{2.8a}$$

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{y}_{2,t-1} + \mathbf{v}_t, \tag{2.8b}$$

where each α_{it} $(i = 2, ..., n)$ in the random or time-varying parameter vector $\boldsymbol{\alpha}_t$ is defined as $\alpha_{it} = \alpha_i G_i(\mathbf{s}_t; \gamma, \mathbf{c})$ for each i, where $G(\mathbf{s}_t; \gamma, \mathbf{c})$ is the logistic function of $\mathbf{s}_t$, and a_1, δ, $\mathbf{a}_2 = (a_2, ..., a_n)'$, γ and $\mathbf{c}$ are constant parameters. The error terms $(u_t, \mathbf{v}_t')' = (u_t, v_{2t}, ..., v_{nt})'$ are serially correlated. In this study, the logistic functions $G(\mathbf{s}_t; \gamma, \mathbf{c})$ in (2.6) and (2.8a) are in the same form, but the transition variables $\mathbf{s}_t$ are often chosen as $\mathbf{y}_{t-d}$ in (2.6) and $\mathbf{y}_{t-d}$ or $\Delta \mathbf{y}_{t-d}$ in (2.8a). In this model, the delay parameter, d, in $G(\mathbf{s}_t; \gamma, \mathbf{c})$ also affects the forecasts at a different-step ahead, similar to the previous VSTAR model. The proposed model (2.8a)-(2.8b) is a triangular representation of cointegration. Although the triangular representation is not necessarily the only option to model a cointegrated system, some evidence leads to its appeal for forecasting (see Clements and Hendry; 1995). This is a good and important example where the endogenous variables are present, and the assumption of serially correlated errors is reasonable. The included endogenous variables introduce difficulties in the forecasting procedure because they are unobserved at a certain time when the regression, y_{1t}, is not observed. Therefore, the vector version of (2.1) should compensate for models such as triangular cointegrating regressions by considering $\mathbf{x}_{t-j}$ $(j \geq 0)$ rather than $j > 0$. Moreover, serially correlated errors also cause some changes in the implemented bootstrapping for the conditional expectations because of the correlation structure in the residuals.

To forecast y_{1t} in (2.8a), there is another correction that will be implemented in the estimation strategy. Leads and lags of $\mathbf{y}_{2t}$ are added in (2.8a) in order that the least squares estimation is applicable. For that reason, we rewrite the error as $u_t = \sum_{s=-p}^{p} \boldsymbol{\zeta}'_s \Delta \mathbf{y}_{2,t-s} + \tilde{u}_t$ and impose it on the model to estimate and forecast. Then, the h-step ahead forecasts of y_{1t} are generally derived as follows:

$$\begin{aligned} f^{y_1}_{T+h|T} =& E\left(a_1 + \delta(T+h) + \boldsymbol{\alpha}'_{T+h}\mathbf{y}_{2,T+h} + \sum_{s=-p}^{p} \boldsymbol{\zeta}'_s \Delta \mathbf{y}_{2,T+h-s} + \tilde{u}_{T+h} \,|\mathcal{F}_T \right) \\ =& a_1 + \delta(T+h) + E(\boldsymbol{\alpha}'_{T+h} \,|\mathcal{F}_T)(\mathbf{a}_2 h + \mathbf{y}_{2,T}) + \sum_{s=1}^{h} E\left(\boldsymbol{\alpha}'_{T+h}\mathbf{v}_{T+s} \,|\mathcal{F}_T\right) \\ &+ \sum_{s=-p}^{p} \boldsymbol{\zeta}'_s E(\Delta \mathbf{y}_{2,T+h-s} \,|\mathcal{F}_T) + E(\tilde{u}_{T+h} \,|\mathcal{F}_T), \end{aligned} \tag{2.9}$$

from which the following four categorizations (a)-(d) are summarized according to the values of h, p and d. For example, $\sum_{s=-p}^{p} \boldsymbol{\zeta}'_s E(\Delta \mathbf{y}_{2,T+h-s} \,|\mathcal{F}_T) = (\sum_{s=-p}^{p} \boldsymbol{\zeta}'_s)\mathbf{a}_2 + \sum_{s=-p}^{p} \boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s} \,|\mathcal{F}_T)$ if $h > p$, and $\sum_{s=-p}^{p} \boldsymbol{\zeta}'_s E(\Delta \mathbf{y}_{2,T+h-s} \,|\mathcal{F}_T) = (\sum_{s=-p}^{h-1} \boldsymbol{\zeta}'_s)\mathbf{a}_2 + \sum_{s=-p}^{h-1} \boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s} \,|\mathcal{F}_T) + \sum_{s=h}^{p} \boldsymbol{\zeta}'_s \Delta \mathbf{y}_{2,T+h-s}$ if $h \leq p$; $\boldsymbol{\alpha}_{T+h}$ is observed if h is not larger than d, and vice versa.

(a) If $h \leq d \leq p$ or $h \leq p \leq d$ then:

$$\begin{aligned} f^{y_1}_{T+h|T} =& a_1 + \delta(T+h) + \boldsymbol{\alpha}'_{T+h}\left(\mathbf{a}_2 h + \mathbf{y}_{2,T} + \sum_{s=1}^{h} E(\mathbf{v}_{T+s} \,|\mathcal{F}_T) \right) \\ &+ \left(\sum_{s=-p}^{h-1} \boldsymbol{\zeta}'_s \right) \mathbf{a}_2 + \sum_{s=-p}^{h-1} \boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s} \,|\mathcal{F}_T) + \sum_{s=h}^{p} \boldsymbol{\zeta}'_s \Delta \mathbf{y}_{2,T+h-s} + E(\tilde{u}_{T+h} \,|\mathcal{F}_T) \end{aligned}$$

(b) If $d < h \leq p$ then:

$$
\begin{aligned}
f^{y_1}_{T+h|T} =& a_1 + \delta(T+h) + E\left(\boldsymbol{\alpha}'_{T+h}\left|\mathcal{F}_T\right.\right)\left(\mathbf{a}_2 h + \mathbf{y}_{2,T}\right) + E\left(\boldsymbol{\alpha}'_{T+h}\sum_{s=1}^{h}\mathbf{v}_{T+s}\left|\mathcal{F}_T\right.\right) \\
&+ \left(\sum_{s=-p}^{h-1}\boldsymbol{\zeta}'_s\right)\mathbf{a}_2 + \sum_{s=-p}^{h-1}\boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s}\left|\mathcal{F}_T\right.) + \sum_{s=h}^{p}\boldsymbol{\zeta}'_s \Delta\mathbf{y}_{2,T+h-s} + E(\tilde{u}_{T+h}\left|\mathcal{F}_T\right.)
\end{aligned}
$$

(c) If $p < h \leq d$ then:

$$
\begin{aligned}
f^{y_1}_{T+h|T} =& a_1 + \delta(T+h) + \boldsymbol{\alpha}'_{T+h}\left(\mathbf{a}_2 h + \mathbf{y}_{2,T} + \sum_{s=1}^{h} E(\mathbf{v}_{T+s}\left|\mathcal{F}_T\right.)\right) \\
&+ \left(\sum_{s=-p}^{p}\boldsymbol{\zeta}'_s\right)\mathbf{a}_2 + \sum_{s=-p}^{p}\boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s}\left|\mathcal{F}_T\right.) + E(\tilde{u}_{T+h}\left|\mathcal{F}_T\right.)
\end{aligned}
$$

(d) If $d \leq p < h$ or $p \leq d < h$ then:

$$
\begin{aligned}
f^{y_1}_{T+h|T} =& a_1 + \delta(T+h) + E\left(\boldsymbol{\alpha}'_{T+h}\left|\mathcal{F}_T\right.\right)\left(\mathbf{a}_2 h + \mathbf{y}_{2,T}\right) + E\left(\boldsymbol{\alpha}'_{T+h}\sum_{s=1}^{h}\mathbf{v}_{T+s}\left|\mathcal{F}_T\right.\right) \\
&+ \left(\sum_{s=-p}^{p}\boldsymbol{\zeta}'_s\right)\mathbf{a}_2 + \sum_{s=-p}^{p}\boldsymbol{\zeta}'_s E(\mathbf{v}_{T+h-s}\left|\mathcal{F}_T\right.) + E(\tilde{u}_{T+h}\left|\mathcal{F}_T\right.)
\end{aligned}
$$

Forecasting $\mathbf{y}_{2t}$ from (2.8b) is much simpler than forecasting y_{1t}. The h-step ahead forecast of $\mathbf{y}_{2t}$, conditional on $\mathcal{F}_T$, is given by the following:

$$
f^{\mathbf{y}_2}_{T+h|T} = E\left(\mathbf{y}_{2,T+h}\left|\mathcal{F}_T\right.\right) = \mathbf{a}_2 h + \mathbf{y}_{2T} + \sum_{s=1}^{h} E\left(\mathbf{v}_{T+s}\left|\mathcal{F}_T\right.\right), \tag{2.10}
$$

where $E\left(\mathbf{v}_{T+s}\left|\mathcal{F}_T\right.\right) \neq \mathbf{0}$ in general because $\mathbf{v}_t$ contains serial correlations.

The extension to multivariate nonlinear forecasting is of great interest because of an existing requirement from, e.g., the two models illustrated above. Generally, an analytical solution does not exist for any horizons, even for a one-step head forecast. Assume that the following model represents a multivariate nonlinear model:

$$
\mathbf{y}_t = g(\mathbf{x}_{t-j}) + \mathbf{e}_t,
$$

where $\mathbf{e}_t$ contains the contemporaneous and serial correlations and $j \geq 0$. The variables $\mathbf{x}_t$ do not imply the exclusion of any variables in $\mathbf{y}_t$. The one-step ahead forecast of $\mathbf{y}_t$, given $\mathcal{F}_T$, is generally defined as follows:

$$f^{\mathbf{y}}_{T+1|T} = E\left(\mathbf{y}_{T+1} \left| \mathcal{F}_T \right.\right) = E\left(g(\mathbf{x}_{T+1-j}) + \mathbf{e}_{T+1} \left| \mathcal{F}_T \right.\right) \neq g(\mathbf{x}_{T+1-j}),$$

where $\mathbf{e}_t$ has nonzero serial correlations or $j = 0$. Therefore, an analytical solution is not available, even when $h = 1$. Further, looking at the two-step ahead forecast, the following is derived:

$$\begin{aligned} f^{\mathbf{y}}_{T+2|T} =& E\left(\mathbf{y}_{T+2} \left| \mathcal{F}_T \right.\right) = E\left(g(\mathbf{x}_{T+2-j}) + \mathbf{e}_{T+2} \left| \mathcal{F}_T \right.\right) \\ =& \begin{cases} g(\mathbf{x}_{T+2-j}) + E\left(\mathbf{e}_{T+2} \left| \mathcal{F}_T \right.\right), j > 1 \\ E\left(g(g^*(\mathbf{x}_T) + \boldsymbol{\eta}_{T+1}) + \mathbf{e}_{T+2} \left| \mathcal{F}_T \right.\right), j = 1 \\ E\left(g(g^*(g^*(\mathbf{x}_T + \boldsymbol{\eta}_{T+1})) + \boldsymbol{\eta}_{T+2}) + \mathbf{e}_{T+2} \left| \mathcal{F}_T \right.\right), j = 0. \end{cases} \end{aligned}$$

The h-step ahead forecast of $\mathbf{y}_t$ can then be generally obtained in the same way, as follows:

$$\begin{aligned} f^{\mathbf{y}}_{T+h|T} =& E\left(\mathbf{y}_{T+h} \left| \mathcal{F}_T \right.\right) = E\left(g(\mathbf{x}_{T+h-j}) + \mathbf{e}_{T+h} \left| \mathcal{F}_T \right.\right) \\ =& E\left(g(g^*(\mathbf{x}_{T+h-j-1}) + \boldsymbol{\eta}_{T+h-j}) + \mathbf{e}_{T+h} \left| \mathcal{F}_T \right.\right) = \cdots \\ =& E\left(g(\mathbf{x}_T; \boldsymbol{\eta}_{T+h-j}, ..., \boldsymbol{\eta}_{T+1}) + \mathbf{e}_{T+h} \left| \mathcal{F}_T \right.\right) \\ =& \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[g(\mathbf{x}_T; \boldsymbol{\eta}_{T+h-j}, ..., \boldsymbol{\eta}_{T+1}) + \mathbf{e}_{T+h}\right] dF(\boldsymbol{\eta}_{T+h-j}, ..., \boldsymbol{\eta}_{T+1}, \mathbf{e}_{T+h}) \end{aligned}$$

in which $F(\boldsymbol{\eta}_{T+h-j}, ..., \boldsymbol{\eta}_{T+1}, \mathbf{e}_{T+h})$ is the joint distribution of dependent errors (we do not specify different cases here by assuming $h - j > 1$ for longer horizons), and for notation simplification, let $g(\mathbf{x}_T; \boldsymbol{\eta}_{T+h-j}, ..., \boldsymbol{\eta}_{T+1})$ denote that the calculation is based on the observation $\mathbf{x}_t$ up to time T and the joint distribution of dependent errors, $\boldsymbol{\eta}_{T+1}, ..., \boldsymbol{\eta}_{T+h-j}$. In this case, neither independently drawing nor block bootstrapping is applicable because for example $\mathbf{e}_{T+h}$ is conditional on $\mathbf{e}_T$, which is correlated with $\mathbf{e}_{T+h}$. Fortunately, model-based bootstrapping is available and necessary for this case.

2.3 Bootstrap Approximations

Re-sampling a single observation at a time is commonly performed in bootstrapping from independent data and is sufficient to calculate the forecasts in Section 2.1. In contrast to the ordinary bootstrap, block bootstrap is applicable to dependent data by re-sampling blocks of consecutive observations to preserve the dependence structure of the original observations. Related literature include Politis and Romano (1994) and Nordman (2009). In general, centering residuals before bootstrapping can reduce the random bias in bootstrap approximations; in addition, the studentized residuals can be useful.

However, the expectation is conditional on the information up to time T, i.e., the bootstrapping for $\mathbf{e}_{T+h}$ from $\hat{\mathbf{e}}_t$ $(t = 1, 2, ..., T)$ depens on $\mathbf{e}_T$ and $\hat{\mathbf{e}}_T$. In other words, re-sampling blocks to preserve the dependent structure does not apply when the re-sampling for $\mathbf{e}_{T+h}$ depends on $\mathbf{e}_T$ becaue $\mathbf{e}_{T+h}$ and $\mathbf{e}_T$ are correlated. Another option for re-sampling dependent data is the model-based bootstrap. Suppose the correlations in $\mathbf{e}_t$ can be expressed as a $VMA(\infty)$ representation, $\mathbf{e}_t = \sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\varepsilon}_{t-j}$, where $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_\tau$ $(t \neq \tau)$ are uncorrelated. It is straightforward to obtain the forecast $E(\mathbf{e}_{T+s} | \mathcal{F}_T) = E(\sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\varepsilon}_{T+s-j} | \mathcal{F}_T)$ by applying the model-based bootstrap of $\hat{\mathbf{e}}_t$ $(t = 1, ..., T)$. In practice, a truncated $VMA(q)$ or $VAR(p)$ model for $\hat{\mathbf{e}}_t$ $(t = 1, ..., T)$ can be used.

In the following, we will present an example of re-sampling for the conditional expectations in (2.7) using block bootstrap, and for (2.9)-(2.10) using model-based bootstrap. In (2.7), the approximations of the conditional expectations can be computed as follows:

$$\hat{E}\left(\mathbf{v}_{T+1} G\left(\mathbf{s}_{T+2}; \gamma, \mathbf{c}\right) | \mathcal{F}_T\right) = \begin{pmatrix} \hat{E}\left(v_{1,T+1} G\left(\mathbf{s}_{T+2}; \gamma, \mathbf{c}\right) | \mathcal{F}_T\right) \\ \vdots \\ \hat{E}\left(v_{n,T+1} G\left(\mathbf{s}_{T+2}; \gamma, \mathbf{c}\right) | \mathcal{F}_T\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{N} \sum_{j=1}^{N} \hat{v}_{1,T+1}^{(j)} G\left(\hat{\mathbf{y}}_{T+2-d}^{(j)}; \gamma, \mathbf{c}\right) \\ \vdots \\ \frac{1}{N} \sum_{j=1}^{N} \hat{v}_{n,T+1}^{(j)} G\left(\hat{\mathbf{y}}_{T+2-d}^{(j)}; \gamma, \mathbf{c}\right) \end{pmatrix}$$

and

$$\hat{E}\left(G\left(\mathbf{s}_{T+2};\gamma,\mathbf{c}\right)|\mathcal{F}_T\right)=\frac{1}{N}\sum_{j=1}^{N}G\left(\hat{\mathbf{y}}_{T+2-d}^{(j)};\gamma,\mathbf{c}\right),$$

where a simple block bootstrap is imposed with block size one. We first center the raw residual $\hat{\mathbf{v}}_t$ $(t=1,2,...,T)$ to limit the bias such that $\tilde{\mathbf{v}}_t=\hat{\mathbf{v}}_t-\bar{\hat{\mathbf{v}}}_t$. Then, for each j, the vectors $\tilde{\mathbf{v}}_{T+1}^{(j)}$ are independently drawn with replacement from $\tilde{\mathbf{v}}_t$ including the correlations between the elements in vector $\mathbf{v}_t$. After that, we calculate $\tilde{\mathbf{v}}_{T+1}^{(j)}+\bar{\hat{\mathbf{v}}}_t$ as $\hat{\mathbf{v}}_{T+1}^{(j)}$. Thereafter, $\hat{y}_{i,T+1}^{(j)}$ are calculated from $\hat{\mathbf{y}}_{T+1}^{(j)}=\boldsymbol{\mu}+\boldsymbol{\Phi}_1\mathbf{y}_T+\boldsymbol{\alpha}_{\perp}\left(\tilde{\mu}+\boldsymbol{\beta}_1'\mathbf{y}_T\right)G(\hat{\mathbf{y}}_{T+1-d}^{(j)};\gamma,\mathbf{c})+\hat{\mathbf{v}}_{T+1}^{(j)}$.

In (2.9)-(2.10), the conditional expectations are computed as follows:

$$\hat{E}(\tilde{u}_{T+h}|\mathcal{F}_T)=\frac{1}{N}\sum_{j=1}^{N}\hat{\tilde{u}}_{T+h}^{(j)},$$

$$\hat{E}(\mathbf{v}_{T+s}|\mathcal{F}_T)=\begin{pmatrix}\frac{1}{N}\sum_{j=1}^{N}\hat{v}_{2,T+s}^{(j)}\\ \vdots\\ \frac{1}{N}\sum_{j=1}^{N}\hat{v}_{n,T+s}^{(j)}\end{pmatrix},s=1,2,...,h,$$

$$\hat{E}\left(\boldsymbol{\alpha}_{T+h}|\mathcal{F}_T\right)=\begin{pmatrix}\hat{E}\left(\alpha_{2,T+h}|\mathcal{F}_T\right)\\ \vdots\\ \hat{E}\left(\alpha_{n,T+h}|\mathcal{F}_T\right)\end{pmatrix}=\begin{pmatrix}\frac{1}{N}\sum_{j=1}^{N}\hat{\alpha}_{2,T+h}\\ \vdots\\ \frac{1}{N}\sum_{j=1}^{N}\hat{\alpha}_{n,T+h}\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{N}\sum_{j=1}^{N}\alpha_2G_2(\hat{\mathbf{s}}_{2,T+h}^{(j)};\gamma_2,\mathbf{c}_2)\\ \vdots\\ \frac{1}{N}\sum_{j=1}^{N}\alpha_nG_n(\hat{\mathbf{s}}_{n,T+h}^{(j)};\gamma_n,\mathbf{c}_n)\end{pmatrix},$$

and

$$\hat{E}\left(\boldsymbol{\alpha}_{T+h}'\mathbf{v}_{T+s}|\mathcal{F}_T\right)=\frac{1}{N}\sum_{j=1}^{N}\left(\sum_{i=2}^{n}\alpha_i\hat{v}_{i,T+s}^{(j)}G_i(\hat{\mathbf{s}}_{i,T+h}^{(j)};\gamma_i,\mathbf{c}_i)\right),s=1,...,h.$$

where $\hat{\mathbf{s}}_{i,T+h}^{(j)}=a_i+\hat{v}_{i,T+h-d}^{(j)}$, when s_{it} is set as $\Delta y_{it-d}=a_i+v_{it}$ in the model.

A VAR model is fitted for the residuals $(\hat{\hat{u}}_t, \hat{\mathbf{v}}_t')'$ $(t = 1, 2, ..., T)$, and then, this estimated model produces its own residuals, $\hat{\boldsymbol{\varepsilon}}_t$, which are assumed *i.i.d.* in the VAR model. Using the recursive relation in the VAR model, the bootstrapping of $(\hat{\hat{u}}_t, \hat{\mathbf{v}}_t')'$ is achieved via the estimated VAR model and the simple block bootstrap for $\hat{\boldsymbol{\varepsilon}}_t$.

3. Forecast Evaluation

The evaluation of a point forecast primarily includes the adequacy of the model and a comparison of two or more models. Measures to compare model quality usually include ratios or differences of functions of forecast errors (known as forecast equality) and a correlation between one model's forecast error and the other model's forecast or a correlation between one model's forecast error and the difference between two models' forecast errors (both of the two forms are known as forecast encompassing). Testing forecast equality was proposed by Granger and Newbold (1977) and involves testing the equality between variances of two normally distributed random variables in Morgan (1939). Then, it was further developed and extended for many difference cases, covering multi-step ahead, non-*i.i.d.*, non-normal and nested models. Related literature include Meese and Rogoff (1988), Mizrach (1995), West and Cho (1995) and a general context by Diebold and Mariano (1995). Chong and Hendry (1986) and Harvey et al. (1997) proposed testing the conditional efficiency of competitive forecasts (forecast encompassing) in two different forms.

Let $\hat{\mathbf{y}}_{it}$ denote the forecast from model i $(i = 1, 2)$ and let $\mathbf{e}_{it} = \mathbf{y}_{it} - \hat{\mathbf{y}}_{it}$ denote the forecast error from model i. We compare two models, our nonlinear model and a linear benchmark model for each example in Section 2.2. The dimension of $\mathbf{y}_{it}$ is small, no greater then three for our applications. The function of forecast error used in forecast evaluation is usually called the loss function, which is commonly expressed as $l(\mathbf{e}_{it}) = \mathbf{e}_{it}'\mathbf{K}\mathbf{e}_{it}$, where the weighting matrix, $\mathbf{K}$, is a positive definite symmetric square matrix. As usual, we set $\mathbf{K} = \mathbf{I}$. The measure we apply is a multivariate version of the root-mean-squared forecast errors (RMSFEs),

$\sqrt{E(\mathbf{e}'_{it}\mathbf{e}_{it})} = \sqrt{\text{trace}(var(\mathbf{e}_{it}))}$ (see Christoffersen and Diebold; 1998).

The equality for forecast loss measures the relative perforemance of the two forecasts. Testing forecast equality determines if the population mean of the loss-differential series, $E(l(\mathbf{e}_{it}) - l(\mathbf{e}_{jt}))$, is zero. If the null of equal forecast accuracy is rejected, then it would imply a better forecast. However, the preference of a particular model in terms of forecast accuracy does not mean that the non-preferred forecast does not contain any useful information absent in the preferred forecast. Therefore, tests for forecast encompassing are developed to discover if the non-preferred forecast can contribute more information. Testing for forecast encompassing tests if the correlation between $\mathbf{e}_{it}$ and $\mathbf{e}_{it} - \mathbf{e}_{jt}$ is zero, or if the correlation between $\mathbf{e}_{it}$ and $\hat{\mathbf{y}}_{jt}$ is zero. If the null is not rejected, then the test suggests that the forecast $\hat{\mathbf{y}}_{it}$ is conditionally efficient with respect to $\hat{\mathbf{y}}_{jt}$, i.e., $\hat{\mathbf{y}}_{it}$ is encompassing $\hat{\mathbf{y}}_{jt}$. Otherwise, $\hat{\mathbf{y}}_{jt}$ can provide useful information absent in $\hat{\mathbf{y}}_{it}$.

4. Data and Modeling

The *first* data set in Figure 1 shows the monthly percentage growth of personal consumption expenditures, $z_t^{(C)}$, and percentage growth of personal disposable income, $z_t^{(I)}$, in the United States, from January 1985 to December 2011 (324 observations)*. Both of them are seasonally adjusted at annual rates. We use the observations from 1985:l to 2006:l2 (264 observations) for modeling and use the remaining data for out-of-sample forecasting.

Modeling strategy begins with testing for nonlinearity, where testing linearity against nonlinearity is one option. However, when we suspect a unit root, testing the null of a linear unit root process against a stationary nonlinear process becomes another option. We refer to Luukkonen et al. (1988) for the possibility of the former option and Li and He (2013a) for the later one. If the null of linearity is rejected, then including CNFs becomes part of the study also. When nonlinear structures appear in both series, $z_t^{(C)}$ and $z_t^{(I)}$, we apply the derived LM test in Li and He (2013a) to examine the CNFs in the system. If the system is likely to

*The data from 2009:1 to 2011:12 was revised in year 2012.

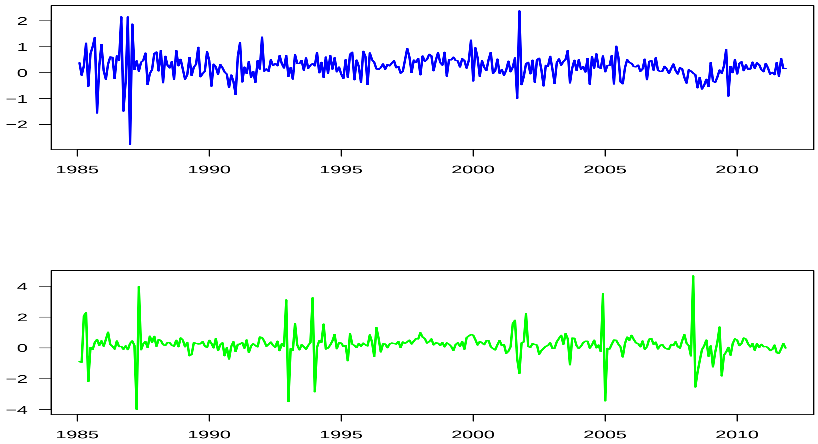

Figure 1: Percentage change of personal consumption expenditures (upper) and disposable income (lower), United States, billions of 2005 dollars, monthly, 1985:1-2011:12.

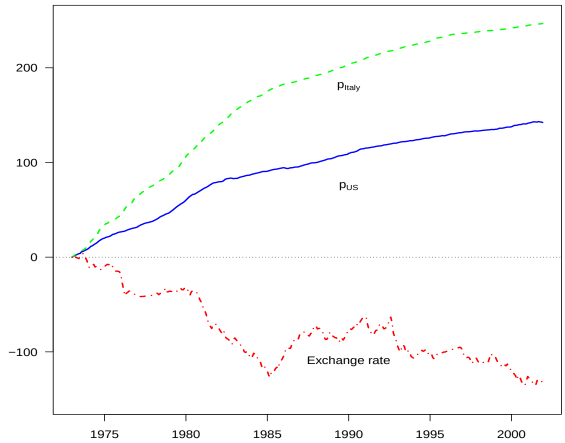

Figure 2: The price level of the United States (solid line, blue) and Italy (dashed line, green) and the dollar-lira exchange rate (dot-dashed line, red), monthly, 1973:1-2001:12.

have CNFs, then the system can be modeled by a VSTAR model with CNFs, using (2.6). Model specification, i.e., choosing the nonlinear components of $G(\mathbf{s}_t, \gamma, \mathbf{c})$, is achieved by testing for nonlinearity simultaneously, and we choose the one that makes the series most significantly nonlinear. The estimation is numerically obtained by minimizing the sum of squared residuals, subject to a reduced rank restriction on the coefficient matrices (the algorithm is provided in Li; 2013). Li

Table 1: The unit root tests.

time series	s_t	DF statistics t	DF statistics F	critical values t	critical values F
y_{1t}	Δy_{1t-1}	-0.8124	2.1339	-3.43	4.76
	y_{1t-1}	-0.5234	5.7210	-3.75	5.64
y_{2t}	Δy_{2t-1}	-0.7025	3.9011	-3.43	4.76
	y_{2t-1}	-0.1287	4.5767	-3.75	5.64
y_{3t}	Δy_{3t-1}	-1.8076	1.6904	-3.43	4.76
	y_{3t-1}	0.5478	1.8508	-3.75	5.64

and He (2013a) analysed this data set and using the modelling procedure outlined above they adopted a VSTAR with CNFs. The same model specification will be used here. Lag selection is additionally involved by AIC criterion ending up with $p = 3$. Estimating the model on the data up 2006:12 yields:

$$\begin{aligned}\begin{pmatrix}\hat{z}_t^{(C)}\\ \hat{z}_t^{(I)}\end{pmatrix} &= \begin{pmatrix}0.4502\\ 0.3296\end{pmatrix} + \begin{pmatrix}-0.5596 & -0.0028\\ -0.0379 & -0.4703\end{pmatrix}\begin{pmatrix}z_{t-1}^{(C)}\\ z_{t-1}^{(I)}\end{pmatrix}\\ &+ \begin{pmatrix}-0.1788 & 0.1022\\ 0.1518 & -0.2219\end{pmatrix}\begin{pmatrix}z_{t-2}^{(C)}\\ z_{t-2}^{(I)}\end{pmatrix} + \begin{pmatrix}-0.1328 & 0.0736\\ 0.3031 & -0.1673\end{pmatrix}\begin{pmatrix}z_{t-3}^{(C)}\\ z_{t-3}^{(I)}\end{pmatrix}\\ &+ \begin{pmatrix}-0.3411\\ -0.4651\end{pmatrix}(0.3595 - 0.5479 z_{t-1}^{(C)} - 0.3791 z_{t-1}^{(I)} - 0.1350 z_{t-2}^{(C)}\\ &+ 0.0413 z_{t-2}^{(I)} - 0.7514 z_{t-3}^{(C)} + 0.2033 z_{t-3}^{(I)})\hat{G}(\mathbf{s}_t;\gamma,\mathbf{c}),\end{aligned}$$

where:

$$\hat{G}(\mathbf{s}_t;\gamma,\mathbf{c}) = \left(1 + \exp\left\{-7.58(z_{t-7}^{(C)} - 0.30)/\hat{\sigma}_{z_{t-7}^{(C)}}\right\}\right)^{-1}$$

and $\hat{\sigma}_{z_{t-7}^{(C)}} = 0.5146$. The Ljung-Box test is used for the diagnosis of residuals, and the residuals are not likely to have correlations.

The *second* data set includes three monthly time series from 1973:1 to 2001:12 (348 observations), the price level of the United States, y_{1t}, the price level of Italy, y_{2t} and the dollar-lira exchange rate, y_{3t}. Natural logs of the raw data were taken and multiplied by 100, and then normalized by subtracting the initial value for 1973:l in Figure 2. We use the observations from 1973:l to 1996:l2 for modeling and use the remaining data for out-of-sample forecasting.

Building ST cointegrating regression models is initiated by testing for ST coin-

Table 2: Testing linear cointegration against nonlinear cointegration.

s_{2t}	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	14.10	14.03	14.83	21.81	14.70	14.52
s_{2t}	$\Delta y_{2,t-2}$	$\Delta y_{2,t-2}$	$\Delta y_{2,t-2}$	$\Delta y_{2,t-2}$	$\Delta y_{2,t-2}$	$\Delta y_{2,t-2}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	16.82	14.16	**21.86**	14.96	14.55	14.67
s_{2t}	$\Delta y_{3,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-1}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	0.79	0.59	14.17	14.74	1.85	0.63
s_{2t}	$\Delta y_{3,t-2}$	$\Delta y_{3,t-2}$	$\Delta y_{3,t-2}$	$\Delta y_{3,t-2}$	$\Delta y_{3,t-2}$	$\Delta y_{3,t-2}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	0.85	0.61	14.20	14.34	0.70	1.81
s_{2t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	0.49	0.07	16.02	14.45	0.06	0.06
s_{2t}	$\Delta y_{1,t-2}$	$\Delta y_{1,t-2}$	$\Delta y_{1,t-2}$	$\Delta y_{1,t-2}$	$\Delta y_{1,t-2}$	$\Delta y_{1,t-2}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-2}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-2}$	$\Delta y_{3,t-1}$	$\Delta y_{3,t-2}$
statistics	0.26	0.41	15.76	16.50	0.23	0.22

Critical value at 5% significance level is 2.40. The bold face test is the most significantly rejected.

tegration. First, we test for the unit root in an individual time series, y_{it}, using a nonlinear framework, see for example He and Sandberg (2006). If the null of the unit root cannot be rejected, then the system is recognized as nonstationary. Next we investigate the evidence of nonlinearity in the nonstationary system, $\mathbf{y}_t$, by testing linear cointegration against ST cointegration, as proposed in Li and He (2013b). If the null of linearity is rejected, then we specify the ST cointegrating regression models by choosing the nonlinear component in the ST function $G(\mathbf{s}_t, \gamma, \mathbf{c})$ and also the variable for the regression variable, y_{1t}. The principle is that the specified model should give the most significant rejection in testing for ST cointegration. The estimation of the specified model is based on nonlinear least squares (NLS). Numerically, the NLS estimators are set by an iterative minimization algorithm with initial parameter values chosen by grid search. Finally, we diagnose the cointegration relation by additional tests of stationarity for residuals such as the KPSS test, reported in Kwiatkowski et al. (1992), or a nonparametric method such as the rank test, used in Breitung (2001).

The null hypothesis that y_{it} is nonstationary is not rejected by the unit root tests, as shown in Table 1. It is therefore possible to state that each time series is a unit root process. The results of testing for nonlinearity are presented in Table 2. The null of linearity is significantly rejected in the nonstationary system. The test of nonlinearity is also used to specify the model, i.e., the one that presents the most nonlinear relation in the system is chosen. Among the choices in Table 2, the transition variables $s_{2t} = \Delta y_{2,t-2}$ and $s_{3t} = \Delta y_{2,t-1}$ are most optimal in the sense that they contribute the most to the rejection of the null. The linear and nonlinear combination of the PPP system is depicted in Figure 3 and the corresponding estimated cointegrating vector is suggested as follows:

$$\begin{aligned}\hat{y}_{1t} =& -\underset{(0.379)}{2.92} + \underset{(0.003)}{0.10}\, t + \underset{(0.005)}{0.46}\, y_{2t} + \underset{(0.004)}{0.05}\, y_{3t} \\ &+ \underset{(0.006)}{0.04}\, y_{2t} \left(1 + \exp\{-\underset{(0.238)}{1.78}(\Delta y_{2,t-2} - \underset{(0.060)}{1.15})/\hat{\sigma}_{\Delta y_{2,t-2}}\}\right)^{-1} \\ &- \underset{(0.004)}{0.02}\, y_{3t} \left(1 + \exp\{-\underset{(0.719)}{2.74}(\Delta y_{2,t-1} - \underset{(0.142)}{1.09})/\hat{\sigma}_{\Delta y_{2,t-1}}\}\right)^{-1} \\ &+ \underset{(0.119)}{0.75}\, \Delta y_{2,t+1} + \underset{(0.122)}{0.65}\, \Delta y_{2,t+2} + \underset{(0.123)}{0.59}\, \Delta y_{2,t+3} + \underset{(0.115)}{1.01}\, \Delta y_{2,t+4} + \underset{(0.117)}{0.29}\, \Delta y_{2t} \\ &- \underset{(0.207)}{1.61}\, \Delta y_{2,t-2} + \underset{(0.121)}{0.33}\, \Delta y_{2,t-3} + \underset{(0.110)}{0.85}\, \Delta y_{2,t-4} - \underset{(0.017)}{0.12}\, \Delta y_{3,t+4} - \underset{(0.018)}{0.09}\, \Delta y_{3t} \\ &- \underset{(0.020)}{0.05}\, \Delta y_{3,t-1} - \underset{(0.020)}{0.10}\, \Delta y_{3,t-2} - \underset{(0.020)}{0.05}\, \Delta y_{3,t-3} - \underset{(0.019)}{0.07}\, \Delta y_{3,t-4},\end{aligned}$$

where $\hat{\sigma}_{\Delta y_{2,t-1}} = 0.585$ and $\hat{\sigma}_{\Delta y_{2,t-2}} = 0.582$. When we apply the KPSS test to diagnose residuals, the test cannot be rejected with a p-value equal to 0.1, which indicates that the residuals are stationary.

5. Results

For each data set, we forecast using the recursive scheme for 60 months at different horizons $h = 1, 3, 6, 12$. We re-estimate the model each time an observation is added, but we do not re-specify the model. Block bootstrap is applied to the system of consumption and income including the correlations between the errors of consumption and income. The model-based bootstrap is applied to the US-IT

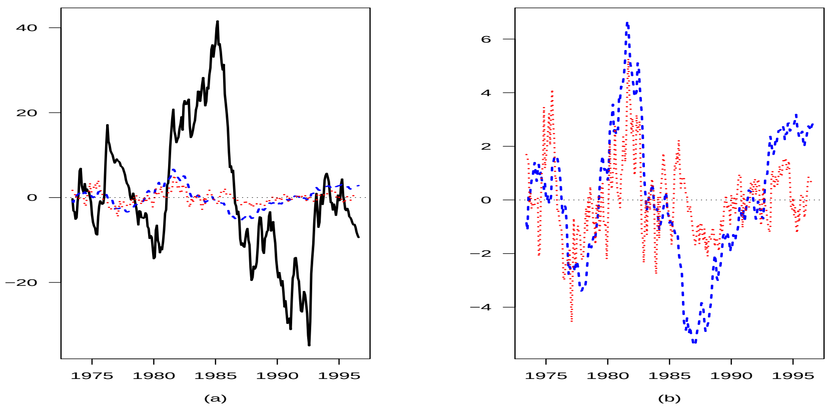

Figure 3: Panel (a): Linear combination with given cointegrating vector $(1, -1, -1)'$ (solid, black), estimated cointegrating vector (dashed, blue), and nonlinear combination (dotted, red). Panel (b): Magnification of estimated linear and nonlinear combinations in panel (a).

PPP system, allowing serial correlations in the errors.

The root-mean-squared forecast errors (RMSFEs) are reported in Table 3. The numbers in parentheses are the ratios of the RMSFEs results to a linear benchmark model. We first compare our VSTAR model to VAR for consumption and income as a system. Then, for consumption and income individually, we compare our VSTAR model to the VAR model and to the AR model, respectively. US-IT as a system is analyzed by comparing our ST-cointegration model and the nonstationary autoregressive model with correlated errors to the random walk model. Then, we further compare three series to their random walk models individually.

Tables 4, 5 and 6 contain results from the formal evaluation tests described in Section 3 for equal forecast accuracy and forecast encompassing. The critical values used are the 95% quantiles of the Chi-square distribution with two and three degrees of freedom (5.99 and 7.81) and standard normal distribution (±1.96). Table 4 gives the results of testing if two forecasts are euqally well. If the null of euqal accruracy is rejected, then the forecast with lower RMSFE will be preferred. The results in Table 5 is intended to show if linear models can contribute useful information absent from nonlinear models, and those in Table 6 if nonlinear models can contribute additional information. We apply those two forms of testing

Table 3: Forecasting performance in terms of RMSFE.

Forecasts	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Consumption and Income from VSTAR	1.0100	0.9905	1.0087	1.0640
(ratio to VAR benchmark)	(1.0097)	(1.0092)	(1.0044)	(1.0018)
Consumption from VSTAR	0.4424	0.3775	0.4011	0.4186
(ratio to VAR benchmark)	(0.9916)	(1.0243)	(0.9943)	(0.9951)
(ratio to AR benchmark)	(1.1171)	(1.0821)	(1.0477)	(0.9775)
Income from VSTAR	0.9079	0.9158	0.9255	0.9782
(ratio to VAR benchmark)	(1.0145)	(1.0067)	(1.0064)	(1.0031)
(ratio to AR benchmark)	(0.9787)	(1.0279)	(1.0023)	(1.0069)
US-IT from ST-Cointegration	3.2421	5.5041	7.2004	10.9217
(ratio to RW benchmark)	(1.3150)	(1.1480)	(1.2489)	(1.4440)
US-IT from AR with correlated errors	2.9526	5.4061	7.1319	10.8216
(ratio to RW benchmark)	(1.1976)	(1.1276)	(1.2370)	(1.4308)
US price from ST-Cointegration	1.3658	1.1426	1.1996	1.8015
(ratio to RW benchmark)	(3.9611)	(1.3316)	(0.7634)	(0.6131)
US price from AR with correlated errors	0.2678	0.4861	0.6759	1.0338
(ratio to RW benchmark)	(0.7766)	(0.5665)	(0.4302)	(0.3518)
IT price from ST-Cointegration	0.1888	0.3882	0.6793	1.5364
(ratio to RW benchmark)	(0.3151)	(0.2181)	(0.1909)	(0.2162)
EX rate from ST-Cointegration	2.9343	5.3702	7.0672	10.6620
(ratio to RW benchmark)	(1.2027)	(1.1297)	(1.2337)	(1.4032)

forecast encompassing to our data; however, we skip the second form for PPP data because it is not meaningful when the variables are not stationary.

For the system of consumption and income, the forecasts from the nonlinear model and the linear model are similar. Further studies of the differences can be made by formal tests. They are likely to conclude that equal forecast accuracy is not rejected in most of cases, which indicates that the forecast accuracy between nonlinear models and linear models is not particularly different. Both nonlinear and linear models can contribute more information to combined forecasts.

For the US-IT system, the random walk model is not proven to be better. Considering the price levels in the U.S. and in Italy from the ST cointegrating regression model, the relative RMSFE reduces as the forecast horizon increases, and most of them are less than one. That means that the cointegration relation has a positive effect on long-run forecasts. However, equal forecast accuracy is not likely to be rejected in the cases that the random walk model is better than coin-

Table 4: Testing equal forecast accuracy.

Benchmark	$h=1$	$h=3$	$h=6$	$h=12$
Consumption and income from VAR	0.0160	0.0564	0.0205	0.0096
Consumption and income from 2 AR	0.0786	0.0464	0.0904	0.0825
Consumption from VAR	-1.1533	1.8907	**-2.2437**	-0.7463
Consumption from AR	1.5419	1.4570	1.4357	-1.0398
Income from VAR	1.0047	0.6877	**2.5569**	1.2350
Income from AR	-0.7763	1.8261	0.5070	1.4477
US-IT (Coin) from RW	5.7672	0.0353	1.1915	7.1874
US-IT (AR) from RW	4.9268	7.8116	3.7136	**16.139**
US price (Coin) from RW	**5.8654**	**2.2385**	**-5.5458**	**-2.3754**
US price (AR) from RW	**-2.5747**	**-2.9279**	**-2.7657**	**-2.2935**
IT price from RW	**-7.0350**	**-7.5316**	**-7.1108**	**-6.1657**
Ex rate from RW	**2.4021**	1.4596	1.3747	1.2037

Table 5: Testing if nonlinear forecasts encompass linear forecasts.

Benchmark	$h=1$	$h=3$	$h=6$	$h=12$
Form 1: correlation between $\mathbf{e}_{it}$ and $\mathbf{e}_{it}-\mathbf{e}_{jt}$				
Consumption and Income from VAR	0.0289	0.0668	0.0395	0.0034
Consumption and Income from 2 AR	0.1501	0.0759	0.0850	0.0827
Consumption from VAR	-0.8940	1.9903	**-2.2328**	-0.7052
Consumption from AR	1.8262	1.7289	1.8017	0.3179
Income from VAR	1.4550	1.2397	**3.5058**	1.3206
Income from AR	1.4599	**3.1715**	1.2335	1.5315
Form 2: correlation between $\mathbf{e}_{it}$ and $\hat{\mathbf{y}}_{jt}$				
Consumption and Income from VAR	0.3997	0.4404	0.1549	0.0816
Consumption and Income from 2 AR	0.3072	-0.4173	0.1561	0.0783
Consumption from VAR	1.6471	**2.2365**	1.8768	1.3772
Consumption from AR	**2.6120**	**3.2417**	**2.2674**	**2.6228**
Income from VAR	1.6575	**2.4286**	1.7936	1.3518
Income from AR	**2.4548**	**5.0009**	**2.3056**	**2.5675**
US-IT (Coin) from RW	0.8439	0.6147	0.0982	0.1808
US-IT (AR) from RW	0.3845	0.0875	1.2364	0.2765
US price (Coin) from RW	**4.3949**	**4.2362**	0.4988	0.4287
US price (AR) from RW	1.5396	0.1270	-0.1005	-0.2403
IT price from RW	0.9146	0.3874	0.1542	0.2026
Ex rate from RW	**2.9899**	1.5558	1.7102	1.4291

Table 6: Testing if linear forecasts encompass nonlinear forecasts.

Benchmark	$h=1$	$h=3$	$h=6$	$h=12$
Form 1: correlation between $\mathbf{e}_{it}$ and $\mathbf{e}_{it}-\mathbf{e}_{jt}$				
Consumption and Income from VAR	0.0616	0.0147	0.0780	0.0479
Consumption and Income from 2 AR	0.8156	-0.0031	-0.0233	0.0479
Consumption from VAR	1.3624	-1.7802	**2.2594**	0.7871
Consumption from AR	-0.9958	-0.7375	-0.7280	**2.1104**
Income from VAR	-0.0117	0.5815	-1.7441	-1.1528
Income from AR	**2.3877**	0.1817	0.5260	-1.3634
Form 2: correlation between $\mathbf{e}_{it}$ and $\hat{\mathbf{y}}_{jt}$				
Consumption and Income from VAR	0.3875	0.2407	0.1395	0.0747
Consumption and Income from 2 AR	0.2680	0.0999	0.1340	0.0680
Consumption from VAR	1.8860	**2.3242**	1.8767	1.4022
Consumption from AR	1.6596	1.7732	1.5324	1.2005
Income from VAR	**4.1575**	**3.6878**	**2.2239**	**2.4284**
Income from AR	**3.9855**	**3.7555**	**2.1299**	**2.3317**
US-IT (Coin) from RW	**12.292**	**26.562**	**27.277**	5.9527
US-IT (AR) from RW	**12.901**	**24.198**	**7.9872**	**10.881**
US price (Coin) from RW	-1.3298	-0.5041	**6.3580**	**8.9636**
US price (AR) from RW	**4.8190**	**4.0639**	**3.9412**	**3.2593**
IT price from RW	**13.350**	**14.592**	**14.395**	**14.328**
Ex rate from RW	**-12.706**	-0.5793	-1.0278	-1.1647

tegrating regression and is likely to be rejected when the cointegrating regression is better. Nonlinear forecasts mostly encompass linear ones, but the nonlinear model is likely to contribute additional information to the combined forecasts.

6. Conclusions

This study presents studies of point forecasting from multivariate nonlinear time series models. Block bootstrap and model-based bootstrap are imposed for numerical integrations. Testing for forecast equality and for encompassing is used for evaluation. Two examples are illustrated; one is a stationary system of the growth of consumption and income using VSTAR models with CNFs and the other one is a nonstationary PPP system using the ST cointegrating regression model. Nonlinear models contribute additional information absent from linear models. It is likely that combined forecasts will generally outperform both nonlinear and linear ones. However, how to construct combined forecasts for our cases remains as future work.

Acknowledgments

Li is grateful to Sune Karlsson, Kenneth Carling and Johan Lyhagen for invaluable advice. Li also thanks School of Technology and Business Studies at Högskolan Dalarna for financial support.

References

Breitung, J. (2001). Rank tests for nonlinear cointegration, *Journal of Business & Economic Statistics* **19**(3): 331–340.

Chong, Y. and Hendry, D. (1986). Econometric evaluation of linear macro-economic models, *Review of Economic Studies* **53**: 671–690.

Christoffersen, P. F. and Diebold, F. X. (1998). Cointegration and long-horizon forecasting, *Journal of Business & Economic Statistics* **16**(4): 450–458.

Clements, M. P. and Hendry, D. F. (1995). Forecasting in cointegrated systems, *Journal of Applied Econometrics* **10**: 127–146.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**(3): 253–263.

Granger, C. and Newbold, P. (1977). *Forecasting Economic Time Series*, Academic Press, New York.

Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the quality of prediction mean squared errors, *International Journal of Forecasting* **13**: 281–291.

He, C. and Sandberg, R. (2006). Dickey-fuller type of tests against nonlinear dynamic models, *Oxford Bulletin of Economics and Statistics* **68**(s1): 835–861.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics* **54**: 159–178.

Li, D. (2013). Residual-based inference for common nonlinear features, *Manuscript* .

Li, D. and He, C. (2013a). Testing common nonlinear features in vector nonlinear autoregressive models, *Manuscript* .

Li, D. and He, C. (2013b). Testing linear cointegration against smooth-transition cointegration, *Manuscript* .

Meese, R. and Rogoff, K. (1988). Was it real? the exchange rate - interest differential over the modern floating rate period, *Journal of finance* **43**: 933–948.

Mizrach, B. (1995). Forecast comparison in l_2. Manuscript, Rutgers University.

Morgan, W. (1939). A test for significance of the difference between two variances in a sample from a normal bivariate population, *Biometrika* **31**: 13–19.

Nordman (2009). A note on the stationary bootstrap's variance, *The Annals of Statistics* **37**(1): 359–370.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *Journal of the American Statistical Association* **89**(428): 1303–1313.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, North-Holland, pp. 413–457.

West, K. and Cho, D. (1995). The predictive ability of several models of exchange rate volatility, *Journal of econometrics* **69**: 367–391.

West, K. D. (1996). Asymptotic inference about predictive ability, *Econometrica* **64**(5): 1067–1084.