Bruder, Stefan

Working Paper

Comparing several methods to compute joint prediction regions for path forecasts generated by vector autoregressions

Working Paper, No. 181

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Bruder, Stefan (2014) : Comparing several methods to compute joint prediction regions for path forecasts generated by vector autoregressions, Working Paper, No. 181, University of Zurich, Department of Economics, Zurich, https://doi.org/10.5167/uzh-101244

This Version is available at:
https://hdl.handle.net/10419/111239

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 181

Comparing Several Methods to Compute Joint Prediction Regions for Path Forecasts Generated by Vector Autoregressions

Stefan Bruder

November 2014

Comparing Several Methods to Compute Joint Prediction Regions for Path Forecasts Generated by Vector Autoregressions

Stefan Bruder
Department of Economics
University of Zurich
stefan.bruder@econ.uzh.ch

October, 2014

Abstract

Path forecasts, defined as sequences of individual forecasts, generated by vector autoregressions are widely used in applied work. It has been recognized that a profound econometric analysis requires, besides the path forecast, a joint prediction region that contains the whole future path with a prespecified coverage probability. The forecasting literature offers several different methods of computing joint prediction regions, where the existing methods are either bootstrap based or rely on asymptotic results. The aim of this paper is to investigate the finite-sample performance of three methods for constructing joint prediction regions in various scenarios via Monte Carlo simulations.

KEY WORDS: Path Forecast; Joint Prediction Region; Monte Carlo Simulation.

JEL CLASSIFICATION NOS: C15, C32, C53

1 Introduction

Prediction is one of the key objectives in wide areas of applied time series analysis. This corresponds to the following representative scenario: Given an observed k-dimensional time series $\{y_1, \ldots, y_T\}$, one is interested in predicting the future path of one of the variables, that is, $\{\hat{y}_{T,i}(1), \ldots, \hat{y}_{T,i}(H)\}$ for some $H \in \mathbb{N}_{>0}$ and $i \in \{1, \ldots, k\}$. For example, national banks publish predictions of the monthly core inflation for the next twelve months or commercial banks publish predictions of government bond yields over several time periods.

One of the workhorse models used for the computation of a path forecast $\hat{Y}_{T,i}(H) := (\hat{y}_{T,i}(1), \ldots, \hat{y}_{T,i}(H))'$ in applied work is the vector autoregression originally proposed by Sims (1980). According to Stock and Watson (2001), the reasons for the widespread use of the VAR model are its simplicity and yet the capability of capturing rich dynamics in multivariate time series. Current research involving path forecasts generated by VAR models include Meyer and Zaman (2013) or Schorfheide and Song (2012).

Nevertheless, the amount of information about the future path of the variable of interest that is actually obtained by computing $\hat{Y}_{T,i}(H)$ is almost negligible because the future path will be different from the path forecast generated by a VAR($\hat{p}$) model with probability one at least for continuous distributions. Thus, a profound econometric analysis requires, besides $\hat{Y}_{T,i}(H)$, information about the uncertainty about the *entire* future path $Y_{T,H,i} := (y_{T+1,i}, \ldots, y_{T+H,i})'$. In other words, there is a strong need for a *joint prediction region* (JPR) that contains the entire future path with a prespecified coverage probability.

In the context of VAR models, the literature offers different methods to construct a joint prediction region for $Y_{T,H,i}$ with prespecified coverage probability $(1-\alpha)$. Jordà and Marcellino (2010) propose an asymptotic method to construct a rectangular and symmetric joint prediction region. The method is on the one hand based on the assumption that the conditional distribution of the prediction errors is asymptotically normal and on the other hand on the application of results by Scheffé (1953, 1959) and Bowden (1970). Staszewska-Bystrova (2011) proposes a heuristic bootstrap-based method. The joint prediction region with coverage probability $(1-\alpha)$ is constructed as the envelope of the remaining $(1-\alpha)\%$ of generated conditional bootstrap paths that survived a heuristic iterative elimination procedure. Staszewska-Bystrova and Winker (2013) propose a method that constructs a joint prediction region by using a threshold accepting optimization heuristic on the generated conditional bootstrap paths[1]. Finally, the method of Wolf and Wunderli (2014) is also a bootstrap-based method, but has a sound theoretical foundation. Their symmetric and rectangular joint prediction region is based on the bootstrap predictive distribution of the standardized prediction errors.

In this paper, the finite-sample performance of the methods of Jordà and Marcellino (2010),

[1]The method of Staszewska-Bystrova and Winker (2013) is basically a refinement of Staszewska-Bystrova (2011). However, the computational burden is enormous and yet the performance is generally inferior to the one of Staszewska-Bystrova (2011), for details see Staszewska-Bystrova and Winker (2013, Section 4). Thus, the method of Staszewska-Bystrova and Winker (2013) is omitted from the following performance analysis.

Staszewska-Bystrova (2011) and Wolf and Wunderli (2014) for constructing a joint prediction region based on a VAR model is investigated in various scenarios through Monte Carlo simulations. First, the data generating process is correctly specified and the true lag order is assumed to be known. Second, the data generating process is again correctly specified but the lag order is estimated using the Bayesian information criterion (BIC). Third, the data generating process is misspecified, that is, the data is generated according to a vector moving average process or a threshold vector autoregressive process, but the joint prediction regions are computed based on a VAR($\hat{p}_{\text{BIC}}$) model. Fourth, the performance of the joint prediction regions based on a VAR($\hat{p}$) model are compared to the performance of the joint prediction regions based on a AR($\hat{p}$) model. A joint prediction region based on an AR(p) model is obtained by fitting an univariate autoregressive model to $\{y_{1,i}, \ldots, y_{T,i}\}$. The empirical coverages are reported for all the various scenarios. For some of the scenarios the volumes, defined as the average of the geometric-average widths, indicating the size of the joint prediction regions are also reported.

The remainder of the paper is organized as follows: Section 2 introduces vector autoregressive processes, the different methods of constructing joint prediction regions and the employed bootstrap method. Section 3 describes in detail the Monte Carlo experiment and presents the results of simulation and section 4 finally concludes.

2 Model

2.1 Vector Autoregression

Consider a k-dimensional VAR(p) process:

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \epsilon_t, \tag{1}$$

where y_t is a k-dimensional random vector, the A_i are fixed $k \times k$ coefficient matrices, ν is a k-dimensional vector of fixed intercept terms, and $\{\epsilon_t\}$ is a k-dimensional i.i.d. process with $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon$. The covariance matrix Σ_ϵ is assumed to be positive definite with finite elements. Any VAR(p) process has a kp-dimensional VAR(1) representation

$$Y_t = V + \mathbf{A} Y_{t-1} + U_t,$$

where

$$Y_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, V := \begin{bmatrix} v \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{A} := \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}, U_t := \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

A VAR(p) process is stable and stationary if

$$\det\left(I_k - A_1 z^1 - \ldots - A_p z^p\right) \neq 0 \quad \text{for } z \in \mathbb{C}, |z| \leq 1.$$

A stationary VAR(p) process admits a Moving Average (MA) representation of the following form

$$y_t = \mu + \sum_{i=0}^{+\infty} \phi_i \epsilon_{t-i},$$

where $\mu = \mathbb{E}[y_t] = (I_k - A_1 - \ldots - A_p)^{-1} \nu$ and the ϕ_i are fixed $k \times k$ MA-coefficient matrices.

2.2 Estimation and Finite Sample-Bias

The parameters of a VAR(p) process, $\beta := \text{vec}(\nu, A_1, \ldots, A_p)$, are consistently estimated by the standard procedure of Least Squares (LS). The LS estimator can be written in the following closed form expression $\hat{\beta}_{LS} = ((ZZ')Z \otimes I_k)y$, where $Z = [Z_0, \ldots, Z_T]_{[(kp+1)\times T)]}$ with $Z_t' := [1\ y_t' \cdots y_{t-p+1}']$ and $y := \text{vec}((y_1, \ldots, y_T))$.

The number of lags, if unknown, is estimated by minimizing the Bayesian information criterion[2] (BIC) over a compact set of lag orders $S \subset \mathbb{N}_{>0}$

$$\hat{p}_{BIC} \in \underset{m \in S \subset \mathbb{N}_{>0}}{\arg\min}\ \text{BIC}(m) = \log\left(\left|\hat{\Sigma}_\epsilon(m)\right|\right) + \frac{\log(T)}{T} m k^2,$$

where $|\hat{\Sigma}_\epsilon(m)|$ denotes the determinant of the estimated covariance matrix of ϵ_t based on a VAR(m) process. The BIC is a consistent order selection criteria, that is, $\hat{p}_{BIC} \xrightarrow{p} p$, where $\xrightarrow{p}$ denotes convergence in probability as $T \to \infty$. A more detailed discussion about parameter estimation and the order selection in vector autoregressions can be found in Lütkepohl (2005, Section 3, Section 4).

It is a well-known fact that the presence of lagged endogenous variables in vector autoregressions entails that the LS estimator of β is biased in finite-samples, that is, $\mathbb{E}[\hat{\beta}_{LS}] \neq \beta$. As a consequence, correcting the LS estimates for its finite-sample bias is desirable. The literature offers two basic approaches of estimating the finite-sample bias of the least squares estimator; bias estimates based on closed-form formulas or bias estimates based on bootstrap techniques.

Closed-form formulas have been derived by Yamamoto and Kunitomo (1981), Nicholls and Pope (1988) and Pope (1990)[3]. The closed-form formulas are all based on asymptotic approximations of the finite-sample distribution of the least square estimator. Using asymptotic approximations removes the bias up to first order; for details see Yamamoto and Kunitomo (1981) or Pope (1990). Engsted and Pedersen (2014) show that the formula of Yamamoto and Kunitomo (1981) and Pope (1990) are, although independently developed, in fact numerically identical. From a computational point of view these closed-form solutions are easy to implement and fast in terms of execution time.

[2] Alternatively, the lag order can be estimated using the Akaike information criterion (AIC) or the corrected Akaike information criterion (AICc) of Hurvich and Tsai (1993). However, using the BIC results in a more parsimonious model which is desirable for Monte Carlo simulations.

[3] Pope (1990) presents the same bias formula as Nicholls and Pope (1988) but shows that it is still valid under milder assumptions than in the work of Nicholls and Pope (1988).

A nonparametric bootstrap procedure to estimate the bias can be found in Kilian (1998). This bootstrap procedure removes the first-order bias of the LS estimator; for details see Kilian (1998). The bootstrap procedure is also straightforward to implement, but the computational burden is substantial, which makes it less practical for a Monte Carlo simulation[4].

Taking into account the trade-off between fast execution and accuracy of parameter estimates, the choice falls on the bias correction of Pope (1990). Choosing the closed-form formulas of Pope (1990) can be justified by the simulation study of Engsted and Pedersen (2014), which shows for one thing that both approaches indeed yield a significant reduction in bias in finite samples and for another thing that the performance in terms of bias reduction of both approaches is very similar for stationary processes.

Pope (1990) derives the following approximation for the bias of $\hat{\mathbf{A}}$

$$\text{Bias}(\hat{\mathbf{A}}) = -\frac{b}{T} + \mathcal{O}(T^{-\frac{3}{2}}), \tag{2}$$

where

$$b = \Sigma_U \left[(I_{kp} - \mathbf{A})^{-1} + \mathbf{A}' \left(I_{kp} - \left(\mathbf{A}'\right)^2 \right)^{-1} + \sum_{i=1}^{k} \lambda_i \left(I_{kp} - \lambda_i \mathbf{A}' \right)^{-1} \right] \Sigma_Y^{-1}.$$

Here, I_{kp} denotes the $kp \times kp$ identity matrix, λ_i denotes the i-th eigenvalue of $\mathbf{A}$, Σ_Y denotes the covariance matrix of Y_t and Σ_U denotes the covariance matrix of U_t . Neglecting higher order terms and replacing true parameters by its LS estimates yields the following estimate for the finite-sample bias of $\hat{\mathbf{A}}$ and $\hat{V}$

$$\widehat{\text{Bias}}(\hat{\mathbf{A}}) = -\frac{1}{T}\hat{\Sigma}_U \left[\left(I_{kp} - \hat{\mathbf{A}} \right)^{-1} + \hat{\mathbf{A}}' \left(I_{kp} - \left(\hat{\mathbf{A}}' \right)^2 \right)^{-1} + \sum_{i=1}^{k} \hat{\lambda}_i \left(I_{kp} - \hat{\lambda}_i \hat{\mathbf{A}}' \right)^{-1} \right] \hat{\Sigma}_Y^{-1} \tag{3}$$

$$\widehat{\text{Bias}}(\hat{V}) = -\widehat{\text{Bias}}(\hat{\mathbf{A}}) \left(I_{kp} - \hat{\mathbf{A}} \right)^{-1} \hat{V}. \tag{4}$$

The bias-corrected parameter estimates are then given by

$$\hat{\mathbf{A}}^{BC} = \hat{\mathbf{A}}_{LS} - \widehat{\text{Bias}}(\hat{\mathbf{A}}) \quad \text{and} \quad \hat{V}^{BC} = \hat{V}_{LS} - \widehat{\text{Bias}}(\hat{V}).$$

Thus, given an observed sample, there are two different estimates of the parameters of a VAR(p) process available

$$\hat{\beta}_{LS} := \text{vec}(\hat{\nu}, \hat{A}_1, \ldots, \hat{A}_p) \quad \text{and} \quad \hat{\beta}_{LS}^{BC} := \text{vec}(\hat{\nu}^{BC}, \hat{A}_1^{BC}, \ldots, \hat{A}_p^{BC}).$$

[4] Bauer et al. (2012) propose a refined nonparametric bootstrap approach, the so-called "inverse bootstrap bias correction". In their simulation study it is shown that the inverse bootstrap method yields a slightly more accurate bias estimate than the bootstrap method of Kilian (1998) and the closed-form solution of Pope (1990). However, this improvement comes at the cost of an even greater computational burden than the bootstrap procedure of Kilian (1998). The inverse bootstrap method is therefore not suitable for a Monte Carlo simulation.

2.3 Path Forecast and Prediction Error

A path forecast of length H for the i-th variable of a VAR(p) process, based on an observed time series of length T, consists of H individual point forecasts and is denoted by

$$\hat{Y}_{T,i}(H) \coloneqq (\hat{y}_{T,i}(1), \ldots, \hat{y}_{T,i}(H))', \quad \text{for } i = 1, \ldots, k.$$

Given an estimate $\widetilde{\beta}$ for the parameters of the VAR(p) process, the individual point forecasts are computed via the following standard forecasting recursion

$$\hat{y}_T(h) = \widetilde{\nu} + \widetilde{A}_1 \hat{y}_T(h-1) + \cdots + \widetilde{A}_p \hat{y}_T(h-p), \quad \text{for } h = 1, \ldots, H, \tag{5}$$

where $\hat{y}_T(j) = y_{T-j}$ if $j \leq 0$. Using the bias-corrected LS estimates $\hat{\beta}_{LS}^{BC}$ yields the path forecast $\hat{Y}_{T,i}^{BC}(H)$, which will be used in the Monte Carlo simulation throughout.

The estimated prediction error for $h \in \{1, \ldots, H\}$ is given by $\hat{u}_{T+h} \coloneqq y_{T+h} - \hat{y}_T(h)$. Following the standard literature, see for example Lütkepohl (2005, Section 3.5), there are two estimators of the covariance matrix of $\hat{e}_{T+h}$

$$\hat{\Sigma}_y(h) = \sum_{i=0}^{h-1} \hat{\phi}_i \hat{\Sigma}_\epsilon \hat{\phi}_i' \quad \text{and} \quad \hat{\Sigma}_{\hat{y}}(h) = \sum_{i=0}^{h-1} \hat{\phi}_i \hat{\Sigma}_\epsilon \hat{\phi}_i' + \frac{\hat{\Omega}(h)}{T},$$

where an explicit formula for $\hat{\Omega(h)}$ can be found in Lütkepohl (2005, Section 3.5.2). $\hat{\Sigma}_{\hat{y}}(h)$ incorporates the uncertainty originated from the estimation of the parameters. Thus, given an observed sample, there are four different estimates of the forecast error covariance matrix

$$\hat{\Sigma}_y(h) \quad , \quad \hat{\Sigma}_y^{BC}(h) \quad , \quad \hat{\Sigma}_{\hat{y}}(h) \quad \text{and} \quad \hat{\Sigma}_{\hat{y}}^{BC}(h).$$

2.4 JPR of Jordà and Marcellino (2010)

The joint prediction region for $Y_{T,i,H}$ of Jordà and Marcellino (2010) is based on the assumption that the conditional distribution of the prediction errors is asymptotically normal, that is,

$$\sqrt{T}\left(\hat{Y}_{T,i}(H) - Y_{T,H,i} | y_T, y_{T-1}, \ldots\right) \xrightarrow{d} \mathcal{N}\left(0, \Xi_{i,H}\right), \tag{6}$$

where $\xrightarrow{d}$ denotes convergence in distribution as $T \to \infty$. It is further assumed that a consistent estimator of $\Xi_{i,H}$ exists. Based on this distributional assumption, they derive an elliptical joint prediction region using the method of Scheffé (1953, 1959):

$$\left\{\widetilde{Y} \in \mathbb{R}^H : T(\hat{Y}_{T,i}(H) - \widetilde{Y})' \hat{\Xi}_{i,H}^{-1} (\hat{Y}_{T,i}(H) - \widetilde{Y}) \leq \chi^2_{(1-\alpha),H}\right\}.$$

Finally, Jordà and Marcellino (2010) apply the lemma of Bowden (1970) to the elliptical joint prediction region to come up with the following joint prediction region with a nominal coverage probability of $(1-\alpha)$

$$\hat{Y}_{T,i}^{BC}(H) \pm P \left[\sqrt{\frac{\chi^2_{(1-\alpha),H}}{h}} \mathbf{1}_{[H \times 1]}\right], \tag{7}$$

where $\mathbf{1}_{[H\times 1]}$ is a H-dimensional vector of ones, P is the lower-triangular Cholesky decomposition of $T^{-1}\hat{\Xi}_{i,H}$ and $\chi^2_{(1-\alpha),h}$ denotes the $(1-\alpha)$-quantile of a χ^2 distribution with h degrees of freedom. The joint prediction region in (7) is then refined by a step-down recursive procedure to obtain

$$\hat{Y}^{BC}_{T,i}(H) \pm P \left[\sqrt{\frac{\chi^2_{(1-\alpha),h}}{h}} \right]_{h=1}^{H}. \tag{8}$$

However, Staszewska-Bystrova (2013) shows that the coverage probability of the joint prediction region based on (8) is severely distorted if P exhibits negative entries[5]. Staszewska-Bystrova (2013) proposes instead the following *modified* Scheffé joint prediction region

$$\text{JPR}_{\text{Scheffé}} := \hat{Y}^{BC}_{T,i}(H) \pm |P| \left[\sqrt{\frac{\chi^2_{(1-\alpha),h}}{h}} \right]_{h=1}^{H}. \tag{9}$$

Jordà and Marcellino (2010) completely ignore the fact that $\hat{\beta}_{LS}$ is biased in finite samples, P is therefore computed based on the raw LS-estimates. A critical discussion about the Scheffé joint prediction region is found in Wolf and Wunderli (2014, Section 3.3).

2.5 JPR of Staszewska-Bystrova (2011)

The neighbouring path (NP) method of Staszewska-Bystrova (2011) is a bootstrap based approach. More specifically, a total of B bootstrap paths $\hat{Y}^*_{T,H,i,b} := \left(y^*_{T+1}, \ldots, y^*_{T+H}\right)'$ are generated conditional on the original data $\{y_{T-p+1}, \ldots, y_T\}$ resulting in a bootstrap predictive distribution which approximates the unknown distribution of the path forecast. The transformation of the predictive distribution into a joint prediction region is then made by the following heuristic iterative procedure: Remove the particular bootstrap path that is the furthest away from $\hat{Y}^{BC}_{T,i}(H)$, where the distance is measured by the (squared) Euclidean norm[6]. Repeat this procedure until $\alpha \times B$ bootstrap paths are removed. The joint prediction region for $Y_{T,i,H} = (y_{T+1,i}, \ldots, y_{T+H,i})'$, denoted by NP Heuristic, with a nominal coverage of $(1-\alpha)$ is then given by the envelope of the remaining $(1-\alpha)\times B$ paths

$$\text{JPR}_{\text{NP Heuristic}} := \begin{bmatrix} l^*_{1,(1-\alpha)}, u^*_{1,(1-\alpha)} \\ l^*_{2,(1-\alpha)}, u^*_{2,(1-\alpha)} \\ \vdots \\ l^*_{H,(1-\alpha)}, u^*_{H,(1-\alpha)} \end{bmatrix}, \tag{10}$$

where $l^*_{h,(1-\alpha)}$ denotes the lower bound of the envelope of the remaining $(1-\alpha)\times B$ bootstrap paths at forecast horizon h and $u^*_{h,(1-\alpha)}$ denotes the corresponding upper bound. Note that

[5] The simulation study of Wolf and Wunderli (2014) shows that without the absolute value correction of Staszewska-Bystrova (2013) the empirical coverage can be even close to zero for AR(p) models. Simulations run by the author show that the same is indeed true for VAR(p) models.

[6] Alternatively, Staszewska-Bystrova (2011) suggests the L_1-norm, that is $\sum_{h=1}^{H} |\hat{y}_{T,i}(h) - y^*_{T+h}|$. However, the Euclidian norm seems to works better according to the simulation study in Staszewska-Bystrova (2011).

the joint prediction region of Staszewska-Bystrova (2011) is not symmetric about $\hat{Y}_{T,i}^{BC}$ and has a jagged shape due to the way of construction. A critical discussion about the NP Heuristic method, especially the lack of asymptotic validity, is found in Wolf and Wunderli (2014, Section 3.3).

Remark. The method of Staszewska-Bystrova and Winker (2013) basically replaces the heuristic iterative elimination procedure of Staszewska-Bystrova (2011) by a sophisticated threshold accepting optimization heuristic. However, the computational burden of the proposed threshold accepting method is enormous and the extensive simulation study of Staszewska-Bystrova and Winker (2013) demonstrates that the NP Heuristic method of Staszewska-Bystrova (2011) generally outperforms the method of Staszewska-Bystrova and Winker (2013). Thus, in order to be able to conduct the simulation study within a reasonable amount of time, the threshold accepting method is omitted.

2.6 JPR of Wolf and Wunderli (2014)

The method of Wolf and Wunderli (2014) is based on the bootstrap predictive distribution of the standardized prediction errors conditional on the original data $\{y_{T-p+1}, \ldots, y_T\}$. The aforementioned predictive distribution is obtained by computing B standardized bootstrap prediction errors $\hat{S}_{T,b,i}^{*}(H) := (\hat{u}_{T,i,b}^{*}(1)/\hat{\sigma}_{T,b}^{*}(1), \ldots, \hat{u}_{T,i,b}^{*}(H)/\hat{\sigma}_{T,b}^{*}(H))'$, where $\hat{u}_{T,i,b}^{*}(h) := \hat{y}_{T,i,b}^{*}(h) - y_{T+h,i,b}^{*}$ and $\hat{\sigma}_{T,b}^{*}(h)$ denotes the bootstrap prediction standard error. Subsequently, the empirical distribution of $\max_{H,b}^{*} := \|\hat{S}_{T,b,i}^{*}(H)\|_{\infty}$ is obtained, where $\|\cdot\|_{\infty}$ denotes the maximum norm[7]. The symmetric and rectangular joint prediction region of Wolf and Wunderli (2014) with a nominal coverage of $(1-\alpha)$ is then given by

$$\text{JPR}_{\text{WW}} := \hat{Y}_{T,i}^{BC}(H) \pm \begin{bmatrix} \hat{d}_{|\cdot|,(1-\alpha)}^{\max,*} \cdot \hat{\sigma}_{T}^{BC}(1) \\ \hat{d}_{|\cdot|,(1-\alpha)}^{\max,*} \cdot \hat{\sigma}_{T}^{BC}(2) \\ \vdots \\ \hat{d}_{|\cdot|,(1-\alpha)}^{\max,*} \cdot \hat{\sigma}_{T}^{BC}(H) \end{bmatrix}, \tag{11}$$

where $\hat{d}_{|\cdot|,(1-\alpha)}^{\max,*}$ denotes the $(1-\alpha)$ quantile of the empirical distribution of $\max_{H,b}^{*}$ and $\hat{\sigma}_T(h)$ denotes the prediction standard error for forecast period h. The bootstrap procedure already incorporates the estimation uncertainty (through parameter re-estimation), therefore $\hat{\sigma}_{T}^{BC}(h)$ is computed by $\sqrt{(\sum_{i=0}^{h-1} \hat{\phi}_i^{BC} \hat{\Sigma}_{\epsilon}^{BC} (\hat{\phi}_i^{BC})')_{hh}}$.

2.7 Bootstrap

Bootstrap data $\{y_1^{*}, \ldots, y_T^{*}, y_{T+1}^{*}, \ldots, y_{T+H}^{*}\}$, used for the NP Heuristic and the WW joint prediction region, are generated by the following four step bootstrap procedure of Pascual et al. (2011)

[7] For $x \in \mathbb{R}^d$, the maximum norm is defined as $\|x\|_{\infty} := \max\{|x_1|, \ldots, |x_d|\}$.

Step 1: Given $\hat{\beta}_{LS}^{BC}$, $\{y_t\}_{t=1}^{T}$ and the corresponding series of centered and rescaled[8] residuals $\{\hat{\epsilon}_t\}_{t=p+1}^{T}$, generate a bootstrap sample $\{y_1^*, \ldots, y_T^*\}$ via the following recursion

$$y_t^* = \begin{cases} y_t & \text{if } t = 1, \ldots, p \\ \hat{\nu} + \hat{A}_1 y_{t-1}^* + \ldots + \hat{A}_p y_{t-p}^* + e_t^* & \text{if } t = p+1, \ldots, T \end{cases},$$

where e_t^* is a random draw with replacement from the empirical distribution of $\{\hat{\epsilon}_t\}_{t=p+1}^{T}$.

Step 2: Obtain $\hat{\beta}_{LS}^{*,BC} := (\hat{\nu}^{*,BC}, \hat{A}_1^{*,BC}, \ldots, \hat{A}_p^{*,BC})$ by fitting a VAR($\hat{p}_{BIC}^*$) model to the bootstrap sample $\{y_1^*, \ldots, y_T^*\}$.

Step 3: Generate $\{y_{T+1}^*, \ldots, y_{T+H}^*\}$ via

$$y_{T+h}^* = \nu^* + \hat{A}_1^* y_{T+h-1}^* + \ldots + \hat{A}_p^* y_{T+h-p}^* + e_h^*, \quad \text{for } h = 1, \ldots, H,$$

where $y_{T+j}^* = y_{T+j}$ if $j \leq 0$ and e_h^* is a random draw with replacement from the empirical distribution of $\{\hat{\epsilon}_t\}_{t=p+1}^{T}$.

Step 4: Repeat steps 1 to 3 B times.

The bootstrap procedure of Pascual et al. (2011) is asymptotically valid under some regularity conditions, that is $\hat{Y}_{T,H,i}^*$ converges in distribution to $Y_{T,H,i}$ as $T \to \infty$, for the proof see Pascual et al. (2011, p.14–15).

Remark. In Staszewska-Bystrova (2011), the bootstrap data $\{y_1^*, \ldots, y_T^*, y_{T+1}^*, \ldots, y_{T+H}^*\}$ is actually generated using the bootstrap-after-bootstrap procedure of Kim (2001). In contrast to the previously outlined procedure of Pascual et al. (2011), the procedure of Kim (2001) is based on the backward representation of a VAR(p) model. However, using the backward representation for generating bootstrap predictive distributions has some serious disadvantages, for a discussion see Pascual et al. (2011, Section 1). The similar finite-sample performance of both approaches justifies the use of the Pascual et al. (2011) boostrap instead of the bootstrap-after-bootstrap procedure.

3 Monte Carlo Simulation

3.1 Data Generating Processes

The basis data generating process (DGP) is a bivariate VAR(1) process previously used in Amihud and Hurvich (2004), Amihud et al. (2009), and Engsted and Pedersen (2014)

$$\textbf{DGP 1} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.80 & 0.10 \\ 0.10 & 0.85 \end{pmatrix} y_{t-1} + \epsilon_t. \tag{12}$$

[8] The centering and rescaling is suggested as in Stine (1987).

The process in (12) is stationary with corresponding characteristic roots of $\rho_1 = (1.385, 1.077)'$. In order to cover a broader range of stationarity characteristics, the coefficient matrix of the process in (12) is modified, resulting in the following data generating processes and corresponding characteristic roots

$$\textbf{DGP 2} \quad y_t = \begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}-0.80 & 0.10\\0.10 & -0.85\end{pmatrix} y_{t-1} + \epsilon_t, \quad \rho_2 = (-1.385, -1.077)' \tag{13}$$

$$\textbf{DGP 3} \quad y_t = \begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}0.30 & 0.10\\0.10 & 0.35\end{pmatrix} y_{t-1} + \epsilon_t, \quad \rho_3 = (2.336, 4.506)' \tag{14}$$

$$\textbf{DGP 4} \quad y_t = \begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}-0.30 & 0.10\\0.10 & -0.35\end{pmatrix} y_{t-1} + \epsilon_t, \quad \rho_4 = (-2.336, -4.506)' \tag{15}$$

$$\textbf{DGP 5} \quad y_t = \begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}0.30 & 0.10\\0.10 & -0.35\end{pmatrix} y_{t-1} + \epsilon_t, \quad \rho_5 = (-2.740, 3.174)'. \tag{16}$$

The sixth data generating process is a bivariate VAR(4) process originally considered in Kilian (2001), but also used in the simulation study of Staszewska-Bystrova (2013)

$$\textbf{DGP 6} \quad y_t = \begin{pmatrix}1\\1\end{pmatrix} + A^{(6,1)} y_{t-1} + A^{(6,2)} y_{t-2} + A^{(6,3)} y_{t-3} + A^{(6,4)} y_{t-4} + \epsilon_t, \tag{17}$$

where

$$A^{(6,1)} := \begin{pmatrix}0.6362 & -0.0012\\0.0190 & 0.5782\end{pmatrix}, A^{(6,2)} := \begin{pmatrix}-0.0168 & -0.0285\\0.5211 & -0.3041\end{pmatrix}$$

$$A^{(6,3)} := \begin{pmatrix}0.0273 & -0.0028\\0.1568 & 0.2229\end{pmatrix}, A^{(6,4)} := \begin{pmatrix}0.1517 & -0.0198\\-0.7600 & -0.3168\end{pmatrix}.$$

The error process $\{\epsilon_t\}$ is assumed to be an i.i.d. process according to one of the following three distributions

- $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$. A multivariate normal distribution with covariance matrix Σ_ϵ.
- $\epsilon_t := \frac{1}{\sqrt{3}} \times \tilde{\epsilon}$, where $\tilde{\epsilon}_t \sim t_3(0, \Sigma_\epsilon)$. $\tilde{\epsilon}_t$ follows a multivariate t-distribution with 3 degrees of freedom. The covariance matrix of $\tilde{\epsilon}_t$ is given by $3 \times \Sigma_\epsilon$. As a result, each random variate of $\tilde{\epsilon}_t$ is scaled by $\frac{1}{\sqrt{3}}$ to ensure that $\text{Var}(\epsilon_t) = \Sigma_\epsilon$. In the following, this distribution of ϵ_t is just called t-distribution.
- $\epsilon_t := C_{[k\times k]} \left(\frac{1}{\sqrt{6}} \times \left(\tilde{\epsilon}_t - 3_{[k\times 1]} \right) \right)$, where $\tilde{\epsilon}_i \sim \chi^2_3 \quad \forall i \in \{1, \ldots, k\}$ and C is the lower-triangular Cholesky decomposition of Σ_ϵ. $\tilde{\epsilon}_t$ is first centered and standardized to have a covariance matrix equal to the identity matrix. The premultiplication with the Cholesky-factor of Σ_ϵ ensures that $\text{Var}(\epsilon_t) = \Sigma_\epsilon$. In the following, this distribution of ϵ_t is just called χ^2-distribution.

The covariance matrices of the six data generating processes are given by

$$\Sigma_\epsilon^{1,2,3,4,5} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_\epsilon^{6} = \begin{pmatrix} 0.025 & 0.009 \\ 0.009 & 0.387 \end{pmatrix} \times 10^{-3},$$

where Σ_ϵ^6 is again taken from Kilian (2001) as the corresponding data generating process.

3.2 Misspecified Models

The assumption about the underlying true data generating process is crucial in applied work. However, it is an assumption that is not verifiable in practice. It is therefore instructive to investigate the finite-sample performance of each of the three methods of constructing joint prediction regions if the underlying model is misspecified. More specifically, the data is generated by either a vector moving average (VMA) process or a threshold vector autoregressive (TVAR) process. The joint prediction regions are then computed with the *same* methodology as in the previous section, that is, based on a VAR(p) model, where the lag order p is estimated using the BIC.

The basis data generating process is a VMA(1) process used in Galbraith et al. (2002)

$$\textbf{DGP 7:} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \underbrace{\begin{pmatrix} 0.20 & 0.10 \\ 0.10 & 0.60 \end{pmatrix}}_{:=M_1} \epsilon_{t-1} + \epsilon_t, \quad \rho_7 = (-5.669, -1.604)', \tag{18}$$

where ρ_7 denotes the roots of $\det(I_2 + M_1 z)$. The coefficient matrix of the process in (18) is then modified, resulting in the following additional data generating processes and corresponding characteristic roots

$$\textbf{DGP 8:} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0.20 & 0.10 \\ 0.10 & 0.60 \end{pmatrix} \epsilon_{t-1} + \epsilon_t, \quad \rho_8 = (-1.633, 4.710)' \tag{19}$$

$$\textbf{DGP 9:} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0.20 & 0.10 \\ 0.10 & -0.60 \end{pmatrix} \epsilon_{t-1} + \epsilon_t, \quad \rho_9 = (-5.669, -1.604)' \tag{20}$$

$$\textbf{DGP 10:} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0.80 & 0.10 \\ 0.10 & -0.60 \end{pmatrix} \epsilon_{t-1} + \epsilon_t, \quad \rho_{10} = (1.189, 1.790)' \tag{21}$$

$$\textbf{DGP 11:} \quad y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1.20 & 0.10 \\ 0.10 & 0.90 \end{pmatrix} \epsilon_{t-1} + \epsilon_t, \quad \rho_{11} = (-1.150, -0.813)'. \tag{22}$$

Remark. A VMA(q) process has a pure VAR(∞) representation if $\det(I_k + M_1 z^1 + \ldots + M_q z^q) \neq 0$ for $z \in \mathbb{C}, |z| \leq 1$. Thus, DGP 7 – DGP 10 exhibit a VAR(∞) representation, whereas DGP 11 does not. This implies that the processes in (18), (19), (20) and (21) can be approximated by a finite VAR process.

It is assumed that $\{\epsilon_t\}$ is an i.i.d. process with $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$, where Σ_ϵ is taken from Galbraith et al. (2002) and given by

$$\Sigma_\epsilon = \begin{pmatrix} 1.00 & 0.50 \\ 0.50 & 1.00 \end{pmatrix}.$$

The last considered data generating process is a non-linear TVAR(1) process already used by Tsay (1998). The process is given by

$$\textbf{DGP 12:} \quad y_t = \begin{cases} A_1^{(1)} y_{t-1} + \epsilon_t^{(1)}, & \text{if } y_{1,t-1} < 0 \\ A_1^{(2)} y_{t-1} + \epsilon_t^{(2)}, & \text{if } y_{1,t-1} \geq 0 \end{cases}, \tag{23}$$

where

$$A_1^{(1)} = \begin{pmatrix} 0.70 & 0.00 \\ 0.30 & 0.70 \end{pmatrix} \quad \text{and} \quad A_1^{(2)} = \begin{pmatrix} -0.70 & 0.00 \\ -0.30 & -0.70 \end{pmatrix}$$

The error process is assumed to be an i.i.d. process with $\epsilon_t^{(i)} \sim \mathcal{N}(0, \Sigma_i) \quad i \in \{1, 2\}$. The corresponding covariance matrices are also taken from Tsay (1998) and given by

$$\Sigma_1 = \begin{pmatrix} 1.00 & 0.20 \\ 0.20 & 1.00 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} 1.00 & -0.30 \\ -0.30 & 1.00 \end{pmatrix}.$$

3.3 Multivariate vs. Univariate

Consider the following scenario: One observes a k-dimensional time series and is interested in a path forecast and corresponding joint prediction region for the i-th variable. Besides the previously discussed multivariate methodology, there is basically the alternative of constructing a path forecast and corresponding joint prediction region based on an univariate AR($\hat{p}$) model fitted to $\{y_{1,i}, \ldots, y_{T,i}\}$. More specifically, the data is generated according to the multivariate DGP 1 – DGP 6, the path forecast and corresponding joint prediction regions are then computed based on VAR($\hat{p}$) and AR($\hat{p}$). For simplicity, $\{\epsilon_t\}$ is assumed to be an i.i.d. process with $\epsilon_t \sim \mathcal{N}(0, I_2)$.

3.4 Design

The nominal coverage probability of each of the joint prediction regions is 90%. The empirical coverage of a particular joint prediction region is computed in the usual way, that is, the number of continuations that are completely covered by a joint prediction region divided by the total number of continuations. In particular, $1,000$ time series samples $\{y_1, \ldots, y_T\}$ are generated according to the specified data generating processes, each with 100 independent continuations $\{y_{T+1}, \ldots, y_{T+H}\}$. As a result, the empirical coverages are computed based on $100,000$ continuations and are therefore very accurate.

The forecast horizon is $H \in \{6, 12, 24\}$. The sample size is $T \in \{100, 400\}$. The number of bootstrap samples for the WW and the NP Heuristic joint prediction region is $B = 1,000$

throughout. The lag order is initially assumed to be known and afterwards estimated using the BIC. For the cases where the lag order is determined using the BIC, the maximum lag order is 10, that is, the BIC is minimized over $S = \{1, \dots, 10\}$.

In order to be able to compare the volume of the joint prediction regions the geometric average width is computed in each of the 1,000 Monte Carlo repetitions. More specifically, each of the joint prediction regions can be characterized by the Cartesian product of H individual prediction intervals, $\text{PI}_h = [l_h, u_h] \subset \mathbb{R}$, where u_h and l_h denote the upper and the lower bound of the joint prediction region at a given forecast horizon $h \in \{1, \dots, H\}$. The geometric-average widths is then computed by $w_{\text{geometric}} := (\prod_{i=1}^{H} w_i)^{\frac{1}{H}}$, where $w_i := u_h - l_h$. The empirical geometric-average width, which will be reported, is then computed by taking the average

$$\bar{w}_{\text{geometric}} = \frac{1}{1000} \sum_{m=1}^{1000} \left(\prod_{i=1}^{H} w_{m,i} \right)^{\frac{1}{H}},$$

where $w_{m,i} := u_{m,i} - l_{m,i}$.

3.5 Results

The results for the six data generating processes (correctly specified), assuming the true lag order is known, are summarized in Tables 1 and 2. The corresponding results when the true lag order is unknown and instead estimated using the BIC are presented in Tables 3 and 4. The empirical geometric-average widths for the cases where the lag is estimated are presented in Table 5 and 6. The empirical coverages for the vector moving average models are presented in Table 7 and the corresponding empirical geometric-average widths are found in Table 9 . The results for the threshold vector autoregressive model are summarized in Tables 8 and 10. Finally, the results of the univariate vs. multivariate experiment can be found in Table 11.

The main conclusions from the Monte Carlo simulations can be summarized as follows.

- The (modified) Scheffé joint prediction region of Jordà and Marcellino (2010) is not robust with respect to the stationarity characteristics of the data generating processes (correctly specified). More precisely, the method performs satisfactory for models whose roots are near the non-stationary region. However, this is definitely not the case for models with roots far away from the non-stationary region. In this case, the method suffers from massive undercoverage, especially at the long forecast horizons $H \in \{12, 24\}$. Concerning the volume of the (modified) Scheffé joint prediction regions, there are two noticeable insights. First, the volume is also not robust, meaning that for some data generating processes the volume is strictly increasing in H, but for other data generating processes the volume is only weakly increasing or in particular scenarios even decreasing in the forecast horizon. Second, there are scenarios where $\bar{w}_{\text{Scheffé}} > \max\{\bar{w}_{\text{WW-JPR}}, \bar{w}_{\text{NP Heuristic}}\}$, but the coverage is strictly below the coverages of the other two methods.

- The NP Heuristic joint prediction region of Staszewska-Bystrova (2011) features generally good performances for the short forecasting horizon $H = 6$. For $H \in \{12, 24\}$, the performance depends on the stationarity characteristics and the distribution of the errors, with good performances for processes whose roots are close to the non-stationarity region, but with mild to substantial undercoverage for the other scenarios. However, the dependence on the process characteristics declines when $T = 400$. Estimating the lag order has almost no noticeable influence on the performance. The volume of the NP Heuristic joint prediction region is increasing in H and weakly decreasing in the sample size T.

- The joint prediction region of Wolf and Wunderli (2014) is generally robust with respect to the various data generating processes (correctly specified). The performance is very good for $T = 400$ regardless of whether the lag order is estimated or not. The WW joint prediction region is still very reliable for $T = 100$ and $H \in \{6, 12\}$ with empirical coverages close to 90%, again without any noticeable disadvantage when p is estimated. However, there is significant undercoverage for some scenarios with fat-tailed or skewed errors when $H = 24$. The volume of the joint prediction region by Wolf and Wunderli (2014) is increasing in H and decreasing in the sample size T. Both properties are clearly desirable.

- For the cases when the underlying model is misspecified, the (modified) method of Jordà and Marcellino (2010) suffers from severe undercoverage already for the short forecast horizon and completely fails for $H \in \{12, 24\}$, independently of the type of data generating process (VMA or TVAR). The NP Heuristic method of Staszewska-Bystrova (2011) performs reliably for the VMA processes and the TVAR process, when $H = 6$ and $T = 100$. For larger forecast horizons, there is mild to massive undercoverage. The performance is considerably better for $T = 400$, but there is still mild undercoverage for $H = 24$. The WW-JPR of Wolf and Wunderli (2014) performs very good for both types of processes when $T = 400$. However, there is mild undercoverage for $H = 24$ when $T = 100$. All in all, the method of Wolf and Wunderli (2014) is by far the most reliable when the data generating process is misspecified which happens often in applied work.

- The WW and the NP Heuristic joint predictions based on a VAR($\hat{p}$) model outperform their univariate counterparts. However, the method of Wolf and Wunderli (2014) produces reliable joint prediction regions when based on the univariate methodology and $H = 6$. Curiously, the (modified) Scheffé joint prediction region exhibit a sligthly better performance when based on the univariate model.

Nominal Coverage $1-\alpha=90\%$

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	85.3	84.3	83.5	85.8	83.2	81.1	87.3	85.1	83.9
NP Heuristic	88.2	87.4	86.3	87.4	84.9	81.9	88.9	87.0	85.1
WW	88.7	87.9	86.4	89.1	87.3	85.4	88.8	87.6	86.9
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	86.0	85.2	84.1	85.9	84.3	82.0	86.5	85.3	84.1
NP Heuristic	87.8	87.2	86.8	86.8	85.5	82.9	87.5	86.5	85.9
WW	89.6	89.2	88.7	89.1	88.2	86.1	89.4	88.7	88.1
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	67.5	48.1	31.6	70.3	52.2	32.2	71.6	53.3	33.4
NP Heuristic	87.4	84.2	79.5	85.5	80.9	71.1	87.7	85.5	80.7
WW	90.1	89.1	88.0	89.0	87.2	82.1	89.1	87.6	84.0
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.3	45.6	24.9	70.6	50.8	28.4	69.6	48.9	26.9
NP Heuristic	87.0	84.1	79.8	85.5	81.0	71.8	87.5	85.1	79.7
WW	89.9	89.6	88.8	89.0	87.2	82.4	89.0	87.6	83.8
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.0	47.2	29.6	69.9	51.6	31.0	72.0	53.3	33.4
NP Heuristic	87.1	84.0	78.5	85.4	85.4	70.5	87.9	85.3	81.0
WW	90.0	89.0	87.6	89.0	89.2	82.0	89.2	87.6	84.4
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.6	87.1	87.7	87.2	84.5	83.5	87.3	85.5	84.7
NP Heuristic	88.2	88.2	86.8	86.8	83.7	78.4	88.8	88.2	85.6
WW	90.0	90.4	90.2	88.3	86.7	83.6	89.7	89.4	88.0

Table 1: Known Lag, $T = 100$: Empirical Coverages.

Nominal Coverage $1-\alpha=90\%$

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.4	87.0	87.2	88.1	86.9	85.7	89.1	88.4	87.4
NP Heuristic	88.6	87.5	85.4	88.2	86.9	84.7	89.1	88.0	86.1
WW	89.7	89.5	88.9	90.1	89.8	89.3	90.0	89.7	89.1
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.3	87.3	87.3	88.0	86.8	85.1	88.4	88.0	87.1
NP Heuristic	88.6	87.6	86.3	88.2	87.0	85.1	88.7	87.8	86.3
WW	89.8	89.8	89.8	90.0	89.8	89.3	90.0	90.0	89.6
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	68.2	45.8	19.6	73.0	53.4	27.7	73.5	54.0	27.6
NP Heuristic	88.1	86.5	83.6	87.8	85.8	81.1	88.9	87.7	84.6
WW	89.7	89.6	89.3	89.8	89.6	88.3	89.9	89.6	88.4
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	68.4	44.9	17.0	73.2	52.6	26.6	71.1	49.7	22.6
NP Heuristic	88.4	86.6	83.2	88.3	85.9	81.7	88.4	87.2	84.1
WW	89.9	89.6	89.3	90.3	89.6	88.9	89.7	89.5	88.4
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	68.0	43.5	17.0	72.6	52.0	26.1	73.0	53.1	27.0
NP Heuristic	88.6	86.7	83.2	87.8	85.5	81.3	88.8	87.8	85.2
WW	90.1	89.7	89.3	89.8	89.3	88.6	89.7	89.7	88.9
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.0	86.6	86.2	86.7	84.1	82.0	87.2	85.7	84.1
NP Heuristic	88.6	87.5	85.6	88.4	86.4	83.8	89.3	88.2	86.3
WW	90.1	90.0	89.8	90.0	89.4	89.2	90.0	89.7	89.2

Table 2: Known Lag, $T = 400$: Empirical Coverages.

Nominal Coverage $1-\alpha=90\%$

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	85.2	84.0	83.3	85.6	83.2	81.1	87.2	85.5	83.7
NP Heuristic	88.0	87.0	85.8	85.6	82.0	75.0	88.6	87.0	85.1
WW	88.6	87.4	86.3	88.6	86.7	83.6	89.2	88.1	86.9
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	86.0	85.5	84.0	85.8	84.0	82.0	86.6	85.6	84.1
NP Heuristic	87.9	87.5	86.7	86.9	85.3	83.5	87.7	86.6	85.8
WW	89.8	89.5	88.5	88.9	87.7	85.6	89.7	88.9	88.2
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.8	47.7	32.1	70.1	51.6	32.7	72.3	53.2	34.4
NP Heuristic	87.3	84.1	79.8	85.2	79.8	70.9	88.1	85.4	80.9
WW	90.0	89.1	88.0	88.8	86.5	82.1	89.3	87.6	84.0
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.8	45.4	24.5	69.4	50.2	28.3	70.3	49.1	27.0
NP Heuristic	87.3	84.2	79.4	85.1	80.6	71.1	87.9	85.1	79.9
WW	90.4	89.7	88.6	88.6	86.9	81.7	89.4	87.6	84.0
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.3	47.3	29.7	70.3	50.8	32.5	71.8	53.3	32.9
NP Heuristic	87.2	84.6	78.1	85.3	80.1	70.3	88.0	85.4	80.7
WW	90.1	89.5	87.3	88.8	86.7	82.3	89.5	87.7	84.2
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	82.7	78.5	75.7	83.6	78.8	71.4	84.5	78.9	71.9
NP Heuristic	87.4	85.4	82.5	86.5	83.2	75.8	87.7	84.8	80.4
WW	88.8	87.9	86.8	88.6	86.9	82.7	89.3	88.1	85.9

Table 3: BIC order selection, $T = 100$: Empirical Coverages.

Nominal Coverage $1-\alpha=90\%$

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.3	87.0	87.2	87.7	87.0	85.7	88.9	88.2	87.6
NP Heuristic	88.6	87.4	86.3	87.7	86.2	82.8	88.8	87.9	85.9
WW	89.8	89.5	89.5	89.6	89.9	89.1	90.0	89.6	89.1
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	87.3	87.3	87.3	88.1	86.7	85.4	88.3	87.3	87.3
NP Heuristic	88.5	87.9	86.3	88.2	87.0	84.5	88.6	87.5	85.7
WW	89.8	90.0	89.7	90.2	89.8	89.0	89.9	89.6	89.2
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	68.4	45.8	19.3	73.3	53.2	28.0	73.4	53.5	27.4
NP Heuristic	88.5	86.7	83.7	87.9	85.8	81.5	88.7	87.3	84.5
WW	89.9	89.7	89.5	90.0	89.6	88.6	89.7	89.3	88.3
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	69.1	44.3	17.6	73.0	52.1	26.2	71.2	49.8	22.4
NP Heuristic	88.8	86.6	83.7	87.9	85.8	81.7	88.6	87.0	83.8
WW	90.2	89.9	89.6	89.9	89.5	88.9	89.8	89.3	88.2
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	68.4	44.4	17.2	73.0	51.9	27.0	72.9	52.6	26.8
NP Heuristic	88.4	86.6	83.1	88.1	85.8	81.4	88.6	87.4	84.7
WW	90.0	89.7	89.3	90.1	89.6	88.7	89.6	89.4	88.2
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	86.8	86.6	86.3	86.7	84.1	81.5	87.0	85.6	84.0
NP Heuristic	88.6	87.6	85.5	88.3	86.5	83.1	88.9	88.1	86.0
WW	90.0	90.0	90.0	90.0	89.5	88.5	89.8	89.7	89.2

Table 4: BIC order selection, $T = 400$: Empirical Coverages.

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	7.0	9.6	12.7	6.9	9.4	12.4	7.1	9.6	12.8
NP Heuristic	6.6	8.5	10.6	6.4	8.3	9.7	6.5	8.3	10.7
WW	6.5	8.2	10.0	6.9	9.2	11.4	6.9	8.9	11.1
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	7.2	9.9	12.9	6.9	9.5	12.6	7.1	9.9	13.0
NP Heuristic	6.6	8.2	10.0	6.6	8.8	11.3	6.7	8.6	10.5
WW	6.6	8.3	10.0	6.9	9.3	12.0	7.1	9.1	11.2
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	4.0	4.2	3.8	3.9	4.0	4.0	4.0	4.2
NP Heuristic	4.9	5.4	5.8	5.3	6.5	7.4	4.8	5.4	6.0
WW	5.1	5.7	6.2	6.1	8.0	9.3	6.1	7.3	8.3
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	3.9	3.9	3.8	3.8	3.8	3.9	3.9	3.9
NP Heuristic	4.9	5.4	5.8	5.3	6.6	7.6	5.0	5.7	6.3
WW	5.1	5.7	6.2	6.0	7.9	9.6	6.2	7.4	8.3
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	4.0	4.1	3.8	3.9	4.1	3.9	3.9	4.1
NP Heuristic	4.9	5.4	5.7	5.4	6.5	7.7	4.8	5.4	5.9
WW	5.1	5.7	6.1	6.1	7.9	10.1	6.1	7.3	8.3
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.0	3.6	4.3	3.0	3.7	4.4	3.1	3.6	4.2
NP Heuristic	3.0	3.5	4.0	3.0	3.8	4.4	2.9	3.5	4.0
WW	3.0	3.5	4.1	3.3	4.2	5.1	3.3	4.0	4.7

Table 5: BIC order selection, $T = 100$: Empirical Geometric-Average Widths.

	Normal distribution			t-distribution			χ^2-distribution		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	7.2	9.9	13.2	7.0	9.8	12.8	7.2	9.9	13.0
NP Heuristic	6.4	7.9	9.5	6.3	8.2	10.1	6.0	7.8	9.4
WW	6.4	8.0	9.6	6.5	8.8	11.3	6.6	8.6	10.6
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	7.2	9.9	13.2	7.0	9.8	12.8	7.1	9.9	13.0
NP Heuristic	6.4	8.0	9.4	6.3	8.3	10.4	6.5	8.3	10.0
WW	6.4	8.0	9.6	6.6	8.8	11.3	7.2	9.9	13.0
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	3.9	3.8	3.9	3.8	3.8	3.9	3.9	3.8
NP Heuristic	4.9	5.4	5.9	5.4	6.7	8.0	4.6	5.2	5.8
WW	5.0	5.2	6.0	5.8	7.5	9.7	5.9	7.2	8.4
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	3.9	3.8	3.9	3.8	3.7	3.9	3.9	3.7
NP Heuristic	4.9	5.4	5.9	5.4	6.7	8.0	4.8	5.6	6.2
WW	5.0	5.5	6.1	5.8	7.4	9.7	6.0	7.2	8.4
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.9	3.9	3.7	3.9	3.8	3.7	3.9	3.8	3.7
NP Heuristic	4.9	5.4	5.8	5.4	6.7	8.0	4.5	5.2	5.8
WW	5.0	5.5	6.0	5.8	7.4	9.7	5.9	7.2	8.4
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.1	3.9	4.9	3.0	3.9	4.9	3.0	3.9	4.9
NP Heuristic	2.6	3.4	3.8	2.9	3.7	4.5	2.7	3.3	3.9
WW	2.9	3.4	3.9	3.1	4.0	5.1	3.1	3.9	4.7

Table 6: BIC order selection, $T = 400$: Empirical Geometric-Average Widths.

Nominal Coverage $1-\alpha = 90\%$

	$T=100$			$T=400$		
DGP 7	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	63.9	42.9	26.7	63.0	37.8	12.9
NP Heuristic	86.6	83.8	77.1	88.3	86.5	83.3
WW	88.8	88.3	85.6	89.6	89.4	88.8
DGP 8	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	57.7	33.4	14.8	55.7	26.9	6.6
NP Heuristic	86.7	83.5	77.2	88.2	86.4	83.0
WW	89.0	88.6	85.9	89.7	89.6	89.1
DGP 9	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	55.9	30.4	11.9	55.4	26.4	6.1
NP Heuristic	86.8	83.1	76.4	88.2	86.3	83.3
WW	89.0	88.4	85.7	89.6	89.2	89.2
DGP 10	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	79.5	70.8	54.8	79.4	67.9	46.8
NP Heuristic	86.7	86.0	82.6	88.1	87.1	84.9
WW	88.0	88.7	87.9	89.3	89.4	89.4
DGP 11	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	80.2	73.4	66.2	80.1	69.5	53.4
NP Heuristic	86.6	85.7	82.3	87.5	86.9	84.6
WW	87.3	87.7	86.8	88.5	89.1	88.9

Table 7: VMA-model, BIC order selection: Empirical Coverages.

Nominal Coverage $1-\alpha = 90\%$

	$T=100$			$T=400$		
DGP 12	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	72.8	57.8	42.4	74.8	57.0	31.6
NP Heuristic	88.0	85.8	80.6	89.3	88.0	84.6
WW	89.6	88.9	86.8	90.5	90.5	89.7

Table 8: TVAR-model, BIC order selection: Empirical Coverages.

Nominal Coverage $1-\alpha = 90\%$

	T=100			T=400		
DGP 7	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.8	3.8	4.0	3.7	3.6	3.5
NP Heuristic	4.8	5.3	5.6	4.8	5.3	5.7
WW	4.9	5.5	5.9	4.9	5.4	5.9
DGP 8	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.5	3.5	3.5	3.4	3.3	3.2
NP Heuristic	4.8	5.2	5.5	4.8	5.3	5.6
WW	4.9	5.4	5.8	4.8	5.4	5.8
DGP 9	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	3.5	3.4	3.4	3.4	3.3	3.2
NP Heuristic	4.8	5.2	5.5	4.8	5.2	5.7
WW	4.9	5.4	5.8	4.8	5.3	5.8
DGP 10	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	5.3	5.6	5.7	5.1	5.3	5.3
NP Heuristic	5.7	6.4	6.9	5.7	6.4	6.9
WW	5.8	6.5	7.2	5.7	6.4	7.1
DGP 11	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	6.9	7.5	8.1	6.6	6.9	7.1
NP Heuristic	7.3	8.3	9.0	7.2	8.1	8.9
WW	7.3	8.3	9.1	7.2	8.2	9.0

Table 9: VMA-model, BIC order selection: Empirical Geometric-Average Widths.

Nominal Coverage $1-\alpha = 90\%$

	$T=100$			$T=400$		
DGP 12	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	4.6	4.7	4.8	4.5	4.6	4.5
NP Heuristic	5.4	6.0	6.4	5.5	6.0	6.5
WW	5.5	6.2	6.7	5.5	6.1	6.7

Table 10: TVAR-model, BIC order selection: Empirical Geometric-Average Widths.

Nominal Coverage $1-\alpha = 90\%$

	JPR's based on VAR($\hat{p}$)			JPR's based on AR($\hat{p}$)		
DGP 1	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	85.6	84.3	83.0	86.6	85.4	83.9
NP Heuristic	88.6	87.4	85.4	88.0	86.2	82.3
WW	88.9	87.7	85.4	88.5	87.1	83.8
DGP 2	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	85.7	85.4	83.8	86.9	86.6	84.3
NP Heuristic	87.5	87.6	86.6	88.4	87.3	85.0
WW	89.3	89.6	88.4	89.4	89.0	87.6
DGP 3	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.6	48.9	31.1	67.9	50.2	32.3
NP Heuristic	87.4	84.8	79.3	86.9	83.6	77.6
WW	89.9	89.6	87.8	88.8	87.9	85.9
DGP 4	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.6	45.5	24.9	67.4	46.3	26.1
NP Heuristic	87.0	84.0	79.6	86.6	82.9	78.0
WW	90.0	89.5	88.9	88.9	88.0	86.7
DGP 5	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	66.0	47.2	29.4	67.1	48.6	31.5
NP Heuristic	87.0	84.2	78.1	86.6	83.6	77.0
WW	90.0	89.3	87.9	88.8	87.9	85.8
DGP 6	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Scheffé	83.7	80.5	75.8	84.5	80.9	75.9
NP Heuristic	88.1	86.4	82.9	87.7	85.6	81.2
WW	89.3	88.5	86.9	88.7	87.4	84.8

Table 11: BIC order selection, $T = 100$, multivariate vs. univariate: Empirical Coverages.

4 Conclusion

Path forecasts, defined as sequences of individual forecasts, generated by vector autoregressions are widely used in applied work. It has been recognized that a profound econometric analysis requires, besides the path forecast, a joint prediction region that contains the whole future path with a prespecified coverage probability. This paper investigates the finite-sample performance of the methods of constructing joint prediction regions of Jordà and Marcellino (2010), Staszewska-Bystrova (2011) and Wolf and Wunderli (2014) in various scenarios through a Monte Carlo simulation.

The finite-sample performance of the asymptotic method of Jordà and Marcellino (2010) is clearly inferior to the bootstrap based methods of Staszewska-Bystrova (2011) and Wolf and Wunderli (2014). Furthermore, the asymptotic method is not robust with respect to the characteristics of the underlying model; the undercoverage ranges from mild to severe. Misspecification of the underlying model leads to a failure of the asymptotic method in the sense that the joint prediction region is much too narrow resulting in a massive undercoverage.

The method Staszewska-Bystrova (2011) exhibts good and robust coverage properties for the short forecast horizon. For larger forecast horizons, there is systematic undercoverage where the extent depends on the model characteristics as well as the distribution of the errors. The same properties apply for the case of model misspecification.

The best overall performance exhibits the method of Wolf and Wunderli (2014) in the sense that the empirical coverage is closest to the nominal coverage for all considered methods and scenarios. The method is generally robust with respect to model characteristics and produces reliable joint prediction regions already for small sample sizes. Moreover, the method evem works reliably when the model is misspecified.

References

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(04):813–841.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.

Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2012). Correcting estimation bias in dynamic term structure models. *Journal of Business and Economic Statistics*, 30(3):454–467.

Bowden, D. C. (1970). Simultaneous confidence bands for linear regression models. *Journal of the American Statistical Association*, 65(329):413–421.

Engsted, T. and Pedersen, T. Q. (2014). Bias-correction in vector autoregressive models: A simulation study. *Econometrics*, 2(1):45–71.

Galbraith, J., Ullah, A., and Zinde-Walsh, V. (2002). Estimation of the vector moving average model by vector autoregression. *Econometric Reviews*, 21(2):205–219.

Hurvich, C. M. and Tsai, C.-L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, 14(3):271–279.

Jordà, O. and Marcellino, M. (2010). Path forecast evaluation. *Journal of Applied Econometrics*, 25(4):635–662.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80(2):218–230.

Kilian, L. (2001). Impulse response analysis in vector autoregressions with unknown lag order. *Journal of Forecasting*, 20(3):161–179.

Kim, J. H. (2001). Bootstrap-after-bootstrap prediction intervals for autoregressive models. *Journal of Business Economic Statistics*, 19(1):117–128.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Meyer, B. and Zaman, S. (2013). Its not just for inflation: The usefulness of the median CPI in BVAR forecasting. Working Paper 1303, Federal Reserve Bank of Cleveland.

Nicholls, D. F. and Pope, A. L. (1988). Bias in the estimation of multivariate autoregressions. *Australian Journal of Statistics*, 30A(12):296–309.

Pascual, L., Ruiz, E., and Fresoli, D. (2011). Bootstrap forecast of multivariate VAR models without using the backward representation. Working Paper 11–34, Universidad Carlos III de Madrid.

Pope, A. L. (1990). Biases of estimators in multivariate non-gaussian autoregressions. *Journal of Time Series Analysis*, 11(3):249–258.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40(1-2):87–110.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons, New York.

Schorfheide, F. and Song, D. (2012). Real-time forecasting with a mixed-frequency VAR. Working Papers 701, Federal Reserve Bank of Minneapolis.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Staszewska-Bystrova, A. (2011). Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting*, 30(8):721–735.

Staszewska-Bystrova, A. (2013). Modified scheffs prediction bands. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik), Justus-Liebig University Giessen, Department of Statistics and Economics*, 233(5-6):680–690.

Staszewska-Bystrova, A. and Winker, P. (2013). Constructing narrowest pathwise bootstrap prediction bands using threshold accepting. *International Journal of Forecasting*, 29(2):221–233.

Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association*, 82(400):1072–1078.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

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

Wolf, M. and Wunderli, D. (2014). Bootstrap joint prediction regions. *Journal of Time Series Analysis*. Forthcoming.

Yamamoto, T. and Kunitomo, N. (1981). Asymptotic bias of ordinary least squares estimator for multivariate autoregressive models. *Annals of the Institute of Statistical Mathematics*, 36:419–430.