

Article

Model-Free Prediction of Multivariate Time Series

Hanieh Saeidi ¹ and Adel Mohammadpour ^{1,2,*} ¹ Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran 15914, Iran² Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

* Correspondence: adel@aut.ac.ir or adel.mohammadpour@ucalgary.ca

Abstract

This paper extends a model-free prediction framework from univariate to multivariate time series. We show that, under a mild uniformly bounded first-moment condition, a multivariate time series admits a VARMA-type representation and an associated infinite-past linear solution. These results are obtained without assuming that the observed data are generated by a parametric VAR, VARMA, or other specifically postulated stochastic model. Motivated by this representation, we develop a practical forecasting method based on truncating the infinite-past solution and estimating the resulting coefficient matrices by multivariate least squares. The proposed procedure is simple to implement and is applicable to a broad class of stationary or nonstationary, linear or nonlinear multivariate time series when the required moment, identification, and empirical design conditions are plausible. We also describe recursive multi-step forecasting and give multivariate analogues of the main theoretical accuracy statements. The representation results require only a uniformly bounded first moment, whereas the forecast-accuracy results are stated under an explicit selected-solution identification assumption and additional empirical design conditions needed for multivariate least-squares estimation. Monte Carlo results, a real-data application, and additional sensitivity checks show that, in the lag-based implementation studied here, the method is empirically comparable to standard linear and regularized forecasting approaches, while not uniformly dominating them, and retains the conceptual advantage of a model-free formulation.

Keywords: prediction; nonparametric methods; multivariate time series**MSC:** 62M10; 62M20

Academic Editor: Xuejun Wang

Received: 1 June 2026

Revised: 9 July 2026

Accepted: 10 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Motivation and Background

This paper builds on the model-free prediction framework of Mohammadi and Li [1] for scalar time series. The starting point of that framework is the familiar difficulty in time series forecasting: parametric methods can be efficient when the assumed model is correct, but they may be sensitive to misspecification, distributional assumptions, stationarity assumptions, and the selected model order. Standard nonparametric methods are more flexible, yet many of them still assume a regression-type data-generating structure, in which the future is represented as an unknown function of the past plus an error term. Thus, such methods are nonparametric in the unknown function, but they are not fully model-free.

The scalar model-free approach replaces these structural assumptions by a mild uniformly bounded first-moment condition. Under this condition, Mohammadi and Li [1] establish an

ARMA-type representation for a broad class of scalar time series. This representation resembles a classical ARMA equation, but it has a different interpretation: the observed process need not be weakly stationary, and the constructed auxiliary sequence need not be white noise. Therefore, the representation should be viewed as a model-free algebraic device for prediction, not as evidence that the data were generated by a classical ARMA model.

Several interpretive remarks are central to that framework. The constructed sequence is not necessarily an innovation sequence; the ARMA-type representation is not unique; the centering constant is not necessarily the mean of the process; and, after deterministic centering, the same idea can accommodate series with a deterministic trend or time-varying location. The representation is also distinct from the Wold decomposition, because it does not rely on weak stationarity or innovation-type components.

A second main contribution of Mohammadi and Li [1] is an infinite-past linear representation. Although this representation resembles an infinite-order autoregression, it is not obtained by assuming an autoregressive data-generating model. It is derived from the model-free ARMA-type construction. Truncating the infinite-past representation gives a practical forecasting algorithm: finite-lag coefficients are estimated from the observed sample and then used recursively for multi-step prediction. The corresponding large-sample result shows that, in the no-intercept case, the one-step prediction error converges to zero in probability as the truncation level increases.

The scalar method has several advantages: weak assumptions, simple implementation, and conceptual robustness to model formulation, in the sense that a parametric data-generating model need not be specified. This should not be read as a claim of uniform empirical dominance in finite samples. It can be applied to stationary or nonstationary, linear or nonlinear, and possibly heavy-tailed series, provided the uniformly bounded first-moment condition is plausible. At the same time, the framework has important limitations. The representations are not unique, the coefficients should not be interpreted as structural parameters, the constructed auxiliary sequence is not classical white noise, and finite-sample performance depends on the truncation level. Large truncation levels may lead to ill-conditioned regressions, so regularization can be useful in practice.

This paper extends the univariate model-free prediction framework of [1] to multivariate time series. In particular, the VARMA-type representation, the infinite-past vector representation, the truncation-based forecasting method, and the predictor accuracy results developed below are multivariate analogues of the scalar ARMA-type representation, infinite-past representation, prediction algorithm, and asymptotic accuracy result in [1]. The main new feature is that past observations and forecasting coefficients are now vector- and matrix-valued, so the method must account for cross-series dependence and higher-dimensional estimation.

The proposed framework is related to several strands of the time-series literature. Classical references on time-series prediction, forecasting, and multivariate time-series analysis include [2–9]. Nonparametric, model-free, robust, and distribution-free approaches to time-series prediction are discussed in [10–19]. For nonstationary and locally stationary processes, related prediction and model-selection issues are studied in [20–25]. The empirical comparisons in this paper are also connected to regularized and high-dimensional multivariate forecasting methods, including ridge, lasso, elastic-net, sparse VAR, reduced-rank, and factor-based forecasting Bai and Ng [26], Basu and Michailidis [27], Nicholson et al. [28], Stock and Watson [29], Tibshirani [30], Yuan and Lin [31], Zou and Hastie [32], Hoerl and Kennard [33].

The remainder of the paper is organized as follows. The rest of this section introduces the notation and basic assumptions. Section 2 establishes the multivariate VARMA-type model-free representation. Section 3 derives the associated infinite-past linear solution and presents the finite-lag forecasting algorithm. Section 4 states the accuracy results under

the additional selected-solution identification and empirical design assumptions. Section 5 reports the Monte Carlo experiments, the real-data application, and the sensitivity checks. Section 6 summarizes the main findings, limitations, and possible extensions.

Preliminaries and Notation

Consider the m -variate time series

$$\{\mathbf{X}_t, t \in \mathbb{Z}\}, \quad \mathbf{X}_t = (X_{1t}, \dots, X_{mt})',$$

where $m \geq 1$ is fixed. The goal is to predict $\mathbf{X}_{n+h}$, for an integer forecast horizon $h \geq 1$, from the observed vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$. In a fully specified parametric model, prediction is usually based on the conditional mean. When no model is specified but weak stationarity is assumed, a common benchmark is linear prediction using matrix-weighted combinations of past vectors.

Our aim is different: we want a forecasting construction that does not require a parametric VAR, VARMA, nonlinear autoregression, or a specified nonparametric regression model. The basic assumption used throughout the paper is

$$\sup_{t \in \mathbb{Z}} \mathbb{E} \|\mathbf{X}_t\| < \infty, \quad (1)$$

where $\|\cdot\|$ denotes a fixed vector norm on $\mathbb{R}^m$. For a matrix $A \in \mathbb{R}^{m \times m}$, we use the same notation for the matrix norm induced by this vector norm, defined by

$$\|A\| = \sup_{\substack{\mathbf{x} \in \mathbb{R}^m \\ \mathbf{x} \neq \mathbf{0}}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|},$$

where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^m$. Equivalently,

$$\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

This induced norm satisfies

$$\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|, \quad \|AB\| \leq \|A\| \|B\|,$$

whenever the products are defined.

The uniformly bounded first-moment condition is weak compared with many second-order stationarity assumptions, but it is not vacuous. It excludes processes whose marginal first moments grow without bound after the chosen centering, such as explosive processes or series with unremoved deterministic growth in scale. It also does not by itself guarantee stable least-squares estimation or forecast accuracy. The representation theorem below uses only this moment condition, whereas the prediction-accuracy results require the additional selected-solution identification and empirical design conditions stated later in the paper.

This assumption allows many stationary and nonstationary processes, provided their first moments remain uniformly bounded after any necessary deterministic centering.

For comparison, a classical VARMA(p, q) model assumes weak stationarity and a white noise sequence $\{\mathbf{Z}_t\}$. In Section 2, we show that condition (1) is enough to obtain an analogous VARMA-type representation, but without requiring stationarity or white noise innovations.

2. Linear Representation

In the multivariate literature, a weakly stationary m -variate process $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ is called a VARMA(p, q) process if

$$\Phi(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t, \quad (2)$$

where B is the backshift operator, $B^j\mathbf{X}_t = \mathbf{X}_{t-j}$,

$$\Phi(z) = I_m - \Phi_1 z - \cdots - \Phi_p z^p, \quad \Theta(z) = I_m - \Theta_1 z - \cdots - \Theta_q z^q,$$

$\Phi_1, \dots, \Phi_p$ and $\Theta_1, \dots, \Theta_q$ are $m \times m$ matrices, and $\{\mathbf{Z}_t\}$ is an m -dimensional white noise sequence. The goal of this section is to obtain a representation with the same algebraic form as (2), but without assuming weak stationarity or white noise innovations.

In the following theorem, we show that there exists a sequence $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$ such that $\mathbf{X}_t$ can be written as a VARMA-type representation based on $\{\mathbf{Z}_t\}$.

The next theorem is the multivariate analogue of the ARMA-type representation theorem of [1].

Theorem 1. Consider the m -variate time series $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ satisfying (1). Fix $p, q \in \mathbb{N}$. Let $\Phi_1, \dots, \Phi_p$ and $\Theta_1, \dots, \Theta_q$ be $m \times m$ matrices, and define

$$\Phi(z) = I_m - \sum_{i=1}^p \Phi_i z^i, \quad \Theta(z) = I_m - \sum_{i=1}^q \Theta_i z^i.$$

Let $\boldsymbol{\mu} \in \mathbb{R}^m$. Suppose that

$$\det \Theta(z) \neq 0, \quad |z| \leq 1.$$

Then there exists an $\mathbb{R}^m$ -valued sequence $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$ satisfying

$$\Phi(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \Theta(B)\mathbf{Z}_t, \quad (3)$$

where

$$\Phi(B)(\mathbf{X}_t - \boldsymbol{\mu}) = (\mathbf{X}_t - \boldsymbol{\mu}) - \Phi_1(\mathbf{X}_{t-1} - \boldsymbol{\mu}) - \cdots - \Phi_p(\mathbf{X}_{t-p} - \boldsymbol{\mu}),$$

and

$$\Theta(B)\mathbf{Z}_t = \mathbf{Z}_t - \Theta_1\mathbf{Z}_{t-1} - \cdots - \Theta_q\mathbf{Z}_{t-q}.$$

Proof. Let

$$\mathbf{Y}_t = \mathbf{X}_t - \boldsymbol{\mu}.$$

Since $\det \Theta(z) \neq 0$ for $|z| \leq 1$, the matrix-valued function

$$C(z) = \Theta(z)^{-1}$$

is analytic for $|z| < Q$ for some $Q > 1$. Hence it admits the power series representation

$$C(z) = \sum_{j=0}^{\infty} C_j z^j, \quad |z| < Q,$$

where each C_j is an $m \times m$ matrix and

$$\sum_{j=0}^{\infty} \|C_j\| < \infty.$$

Moreover,

$$\Theta(z)C(z) = I_m, \quad |z| \leq 1. \quad (4)$$

Now define

$$\Psi(z) = C(z)\Phi(z).$$

Then Ψ is also analytic for $|z| < Q$, so it can be written as

$$\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j, \quad |z| < Q,$$

with $\Psi_j \in \mathbb{R}^{m \times m}$ satisfying

$$\sum_{j=0}^{\infty} \|\Psi_j\| < \infty.$$

For each $t \in \mathbb{Z}$, define

$$\mathbf{Z}_t := \Psi(B)\mathbf{Y}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{Y}_{t-j}. \quad (5)$$

Since

$$\sup_t \mathbb{E}\|\mathbf{Y}_t\| \leq \sup_t \mathbb{E}\|\mathbf{X}_t\| + \|\boldsymbol{\mu}\| < \infty,$$

Tonelli's theorem gives

$$\mathbb{E} \sum_{j=0}^{\infty} \|\Psi_j \mathbf{Y}_{t-j}\| \leq \sum_{j=0}^{\infty} \|\Psi_j\| \mathbb{E}\|\mathbf{Y}_{t-j}\| \leq \left(\sup_t \mathbb{E}\|\mathbf{Y}_t\| \right) \sum_{j=0}^{\infty} \|\Psi_j\| < \infty.$$

Thus the series in (5) converges absolutely almost surely and in L_1 , and $\mathbf{Z}_t$ is well defined for every fixed t . In particular, $\mathbb{E}\|\mathbf{Z}_t\| < \infty$.

Finally, by (5),

$$\Theta(B)\mathbf{Z}_t = \Theta(B)\Psi(B)\mathbf{Y}_t.$$

Since $\Psi(z) = C(z)\Phi(z)$ and $\Theta(z)C(z) = I_m$ from (4), we have

$$\Theta(z)\Psi(z) = \Phi(z), \quad |z| \leq 1.$$

Therefore,

$$\Theta(B)\mathbf{Z}_t = \Phi(B)\mathbf{Y}_t = \Phi(B)(\mathbf{X}_t - \boldsymbol{\mu}),$$

which proves (3). $\square$

The following remarks are motivated by the univariate model-free representation and prediction framework of [1].

Remark 1. The sequence $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$ in the VARMA-type representation (3) is not necessarily a white noise process. In particular, its components need not be uncorrelated or independent.

Remark 2. For any multivariate time series satisfying (1), there exist infinitely many VARMA-type representations of the form (3). Indeed, one may choose arbitrary matrix polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$, together with arbitrary $\boldsymbol{\mu} \in \mathbb{R}^m$, provided that $\det \Theta(z) \neq 0$ for $|z| \leq 1$. Note that $\boldsymbol{\mu}$ in (3) is not necessarily the mean vector of $\mathbf{X}_t$.

Remark 3. If, instead of a constant vector $\boldsymbol{\mu}$, we consider a deterministic sequence $\{\boldsymbol{\mu}_t, t \in \mathbb{Z}\}$ satisfying

$$\sup_{t \in \mathbb{Z}} \mathbb{E}\|\mathbf{X}_t - \boldsymbol{\mu}_t\| < \infty,$$

then the proof of Theorem 1 implies the existence of a sequence $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$ such that

$$\Phi(B)(\mathbf{X}_t - \boldsymbol{\mu}_t) = \Theta(B)\mathbf{Z}_t.$$

Remark 4. For weakly stationary multivariate time series, the classical Wold decomposition expresses the process as the sum of a deterministic component and an infinite moving-average term driven by an innovation sequence. By contrast, the representation (3) is obtained without assuming weak stationarity, and the constructed sequence $\{\mathbf{Z}_t\}$ is not required to be white noise. Thus, (3) is different from the usual Wold representation and should be viewed as a model-free algebraic representation rather than a classical innovation decomposition. For classical multivariate Wold decompositions and related stationary representations, see Brockwell and Davis [3], Wei [9], and Lütkepohl [6].

3. Prediction Method

In this section, we derive an associated linear solution in terms of the infinite past and then use a truncated version of that solution to construct a practical forecasting method. Motivated by the scalar truncation-based predictor of [1], we estimate the truncated vector solution by multivariate least squares.

Throughout this section, an *associated solution* means a process obtained after fixing one VARMA-type representation and one right inverse of the finite block linear system used in the proof below. Such a solution is not asserted to be the original observed process unless an additional identification condition is imposed.

Theorem 2. Let $\{\mathbf{X}_t; t \in \mathbb{Z}\}$ satisfy (1). Then there exist a vector $\mathbf{c} \in \mathbb{R}^m$ and a sequence of matrices $\{\Gamma_j; j \geq 1\}$, $\Gamma_j \in \mathbb{R}^{m \times m}$, such that

$$\sum_{j=1}^{\infty} \|\Gamma_j\| < \infty,$$

and the process $\{\mathbf{X}_t^*; t \in \mathbb{Z}\}$ defined by

$$\mathbf{X}_t^* = \mathbf{c} + \sum_{j=1}^{\infty} \Gamma_j \mathbf{X}_{t-j} \quad (6)$$

is an associated infinite-past linear solution for the VARMA-type representation (3).

Proof. By Theorem 1, for arbitrary matrix polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$ such that $\det \Theta(z) \neq 0$ for $|z| \leq 1$, and an arbitrary vector $\boldsymbol{\mu} \in \mathbb{R}^m$, the process $\{\mathbf{X}_t; t \in \mathbb{Z}\}$ satisfies

$$\mathbf{Y}_t - \Phi_1 \mathbf{Y}_{t-1} - \cdots - \Phi_p \mathbf{Y}_{t-p} = \mathbf{Z}_t - \Theta_1 \mathbf{Z}_{t-1} - \cdots - \Theta_q \mathbf{Z}_{t-q}, \quad (7)$$

where $\mathbf{Y}_t = \mathbf{X}_t - \boldsymbol{\mu}$.

Moreover, from the proof of Theorem 1, there exists a sequence of matrices $\{\Psi_j; j \geq 0\}$ such that

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{Y}_{t-j}, \quad t \in \mathbb{Z}, \quad (8)$$

and

$$\sum_{j=0}^{\infty} \|\Psi_j\| < \infty.$$

Now define the block vectors

$$\mathcal{U}_t = \begin{pmatrix} \mathbf{0} \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p} \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-q} \end{pmatrix} \in \mathbb{R}^{(p+q+1)m}, \quad \mathcal{V}_t = \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p} \\ \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-q} \end{pmatrix} \in \mathbb{R}^{(p+q+2)m}.$$

Then (7) can be written as

$$\mathcal{A}\mathcal{V}_t = \mathcal{U}_t, \quad (9)$$

where the block matrix $\mathcal{A} \in \mathbb{R}^{(p+q+1)m \times (p+q+2)m}$ is given by

$$\mathcal{A} = \begin{pmatrix} I_m & -\Phi_1 & -\Phi_2 & \cdots & -\Phi_p & -I_m & \Theta_1 & \Theta_2 & \cdots & \Theta_q \\ 0 & I_m & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & I_m & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_m & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & I_m & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & I_m \end{pmatrix}.$$

The matrix $\mathcal{A}$ has full row rank $(p+q+1)m$. To verify this, note that the first block row contains the block $-I_m$ in the column corresponding to $\mathbf{Z}_t$, whereas all remaining block rows contain 0 in that column. Hence the first block row is linearly independent of the others. The remaining block rows are shift identities, each containing a single I_m in a distinct block column and zeros elsewhere. Therefore these block rows are linearly independent, and

$$\text{rank}(\mathcal{A}) = (p+q+1)m.$$

Consequently, $\mathcal{A}$ admits a right inverse $\mathcal{A}^R \in \mathbb{R}^{(p+q+2)m \times (p+q+1)m}$ satisfying

$$\mathcal{A}\mathcal{A}^R = I_{(p+q+1)m}.$$

For definiteness, one may take the Moore–Penrose right inverse

$$\mathcal{A}^R = \mathcal{A}'(\mathcal{A}\mathcal{A}')^{-1},$$

which is well defined because $\mathcal{A}$ has full row rank. The particular choice of right inverse selects one associated solution in the existence proof; it is not a tuning parameter in the empirical forecasting algorithm. Hence

$$\mathcal{V}_t^* := \mathcal{A}^R \mathcal{U}_t$$

is a solution of (9). Let

$$E_1 = \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{m \times (p+q+2)m},$$

so that E_1 extracts the first block of $\mathcal{V}_t^*$. Define

$$\mathbf{Y}_t^* = E_1 \mathcal{V}_t^* = E_1 \mathcal{A}^R \mathcal{U}_t.$$

Since $\mathcal{U}_t$ depends linearly on $\mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-p}, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_{t-q}$, there exist matrices $A_1, \dots, A_p$ and $B_1, \dots, B_q$ in $\mathbb{R}^{m \times m}$ such that

$$\mathbf{Y}_t^* = \sum_{i=1}^p A_i \mathbf{Y}_{t-i} + \sum_{i=1}^q B_i \mathbf{Z}_{t-i}.$$

Substituting (8) yields

$$\mathbf{Y}_t^* = \sum_{i=1}^p A_i \mathbf{Y}_{t-i} + \sum_{i=1}^q B_i \sum_{j=0}^{\infty} \Psi_j \mathbf{Y}_{t-i-j}.$$

Therefore there exists a sequence of matrices $\{\tilde{\Gamma}_j; j \geq 1\}$ such that

$$\mathbf{Y}_t^* = \sum_{j=1}^{\infty} \tilde{\Gamma}_j \mathbf{Y}_{t-j}.$$

Because $q < \infty$ and $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$, we have

$$\sum_{j=1}^{\infty} \|\tilde{\Gamma}_j\| < \infty.$$

Finally, since $\mathbf{Y}_t = \mathbf{X}_t - \boldsymbol{\mu}$,

$$\mathbf{X}_t^* - \boldsymbol{\mu} = \sum_{j=1}^{\infty} \tilde{\Gamma}_j (\mathbf{X}_{t-j} - \boldsymbol{\mu}),$$

which is equivalent to

$$\mathbf{X}_t^* = \left(I_m - \sum_{j=1}^{\infty} \tilde{\Gamma}_j \right) \boldsymbol{\mu} + \sum_{j=1}^{\infty} \tilde{\Gamma}_j \mathbf{X}_{t-j}.$$

Thus (6) holds with

$$\Gamma_j = \tilde{\Gamma}_j, \quad \mathbf{c} = \left(I_m - \sum_{j=1}^{\infty} \Gamma_j \right) \boldsymbol{\mu}.$$

This completes the proof. $\square$

Remark 5. The vector $\mathcal{V}_t^* = \mathcal{A}^R \mathcal{U}_t$ is only one solution of (9). In general,

$$\mathcal{A}(\mathcal{V}_t - \mathcal{A}^R \mathcal{U}_t) = \mathbf{0},$$

so the solution is not unique. Consequently, Theorem 2 constructs an associated solution $\mathbf{X}_t^*$; it does not, by itself, prove that $\mathbf{X}_t^* = \mathbf{X}_t$ for the observed process.

Remark 6. The coefficient matrices Γ_j and the vector $\mathbf{c}$ in (6) are not necessarily unique. Their values depend on the choice of $\Phi(\cdot)$, $\Theta(\cdot)$, and $\boldsymbol{\mu}$ in the VARMA-type representation. However, the matrices Γ_j do not depend on t .

Proposition 1 (Truncation bias). *Let*

$$\mathbf{X}_t^{*(k)} = \mathbf{c} + \sum_{j=1}^{k+1} \Gamma_j \mathbf{X}_{t-j}$$

be the truncation of the infinite-past representation (6), and define

$$\tau_k = \sum_{j=k+2}^{\infty} \|\Gamma_j\|.$$

If (1) holds, then

$$\sup_{t \in \mathbb{Z}} \mathbb{E} \|\mathbf{X}_t^* - \mathbf{X}_t^{*(k)}\| \leq \left(\sup_{t \in \mathbb{Z}} \mathbb{E} \|\mathbf{X}_t\| \right) \tau_k \longrightarrow 0, \quad k \rightarrow \infty.$$

Proof. By (6),

$$\mathbf{X}_t^* - \mathbf{X}_t^{*(k)} = \sum_{j=k+2}^{\infty} \Gamma_j \mathbf{X}_{t-j}.$$

Using the triangle inequality and submultiplicativity of the induced matrix norm,

$$\mathbb{E} \|\mathbf{X}_t^* - \mathbf{X}_t^{*(k)}\| \leq \sum_{j=k+2}^{\infty} \|\Gamma_j\| \mathbb{E} \|\mathbf{X}_{t-j}\| \leq \left(\sup_{s \in \mathbb{Z}} \mathbb{E} \|\mathbf{X}_s\| \right) \sum_{j=k+2}^{\infty} \|\Gamma_j\|.$$

The last term tends to zero because $\sum_{j \geq 1} \|\Gamma_j\| < \infty$. $\square$

Theorem 2 constructs an associated infinite-past linear solution under the uniformly bounded first-moment condition. We therefore propose to approximate (6) by truncating the infinite sum after the $(k+1)$ st lag and estimating the resulting finite-dimensional coefficient matrices.

For the selected associated solution, consider the finite-lag approximation

$$\mathbf{X}_t^* \approx \mathbf{c} + \sum_{i=1}^{k+1} \Gamma_i \mathbf{X}_{t-i}, \quad t = k+2, \dots, n.$$

In practice, we fit the corresponding lag regression using the observed sample $\mathbf{X}_1, \dots, \mathbf{X}_n$. A direct identification of this fitted regression with the observed process requires Assumption 1 below. This leads to the one-step-ahead predictor

$$\hat{\mathbf{X}}_{n+1}^{(k)} = \hat{\mathbf{c}} + \sum_{i=1}^{k+1} \hat{\Gamma}_i \mathbf{X}_{n+1-i}, \quad (10)$$

where $\hat{\Gamma}_1, \dots, \hat{\Gamma}_{k+1} \in \mathbb{R}^{m \times m}$ and $\hat{\mathbf{c}} \in \mathbb{R}^m$ are estimated from the observed data. Here Γ_j denotes the coefficient matrices in the theoretical infinite-past representation, whereas $\hat{\Gamma}_1, \dots, \hat{\Gamma}_{k+1}$ denote their finite-lag least-squares estimates used for forecasting. This finite-lag regression should be understood as an empirical forecasting approximation to the associated solution, rather than as a claim that the coefficients are unique structural parameters of the observed process.

Define

$$\hat{B}_k = \begin{pmatrix} \hat{\Gamma}_1 & \cdots & \hat{\Gamma}_{k+1} & \hat{\mathbf{c}} \end{pmatrix} \in \mathbb{R}^{m \times ((k+1)m+1)},$$

and let

$$\mathcal{Y}(n, k) = \begin{pmatrix} \mathbf{X}'_n \\ \mathbf{X}'_{n-1} \\ \vdots \\ \mathbf{X}'_{k+2} \end{pmatrix} \in \mathbb{R}^{(n-k-1) \times m},$$

$$\mathcal{X}(n, k) = \begin{pmatrix} \mathbf{X}'_{n-1} & \mathbf{X}'_{n-2} & \cdots & \mathbf{X}'_{n-k-1} & 1 \\ \mathbf{X}'_{n-2} & \mathbf{X}'_{n-3} & \cdots & \mathbf{X}'_{n-k-2} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{X}'_{k+1} & \mathbf{X}'_k & \cdots & \mathbf{X}'_1 & 1 \end{pmatrix} \in \mathbb{R}^{(n-k-1) \times ((k+1)m+1)}.$$

We estimate $\hat{B}_k$ by multivariate least squares:

$$\hat{B}'_k = \arg \min_{M \in \mathbb{R}^{((k+1)m+1) \times m}} \|\mathcal{Y}(n, k) - \mathcal{X}(n, k)M\|_F^2, \quad (11)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

A least-squares solution is given by

$$\hat{B}'_k = \mathcal{X}(n, k)^+ \mathcal{Y}(n, k), \quad (12)$$

where $\mathcal{X}(n, k)^+$ is the Moore–Penrose pseudoinverse of $\mathcal{X}(n, k)$. Equivalently,

$$\hat{B}_k = \mathcal{Y}(n, k)' (\mathcal{X}(n, k)^+)'.$$

If $\mathcal{X}(n, k)$ has full column rank, then

$$\hat{B}'_k = (\mathcal{X}(n, k)' \mathcal{X}(n, k))^{-1} \mathcal{X}(n, k)' \mathcal{Y}(n, k).$$

If $\mathcal{X}(n, k)$ has full row rank, then

$$\hat{B}'_k = \mathcal{X}(n, k)' (\mathcal{X}(n, k) \mathcal{X}(n, k)')^{-1} \mathcal{Y}(n, k).$$

Remark 7. The Moore–Penrose pseudoinverse provides a unified expression for the estimator without restricting attention to a particular rank regime. In particular, (12) is well defined regardless of whether $\mathcal{X}(n, k)$ has full column rank or full row rank, although the explicit formulas above simplify in those special cases.

Using (10), the h -step-ahead forecast is obtained recursively as

$$\hat{\mathbf{X}}_{n+h}^{(k)} = \hat{\mathbf{c}} + \sum_{i=1}^{k+1} \hat{\Gamma}_i \tilde{\mathbf{X}}_{n+h-i}, \quad h \geq 1, \quad (13)$$

where

$$\tilde{\mathbf{X}}_{n+h-i} = \begin{cases} \hat{\mathbf{X}}_{n+h-i}^{(k)} & \text{if } h-i > 0, \\ \mathbf{X}_{n+h-i} & \text{if } h-i \leq 0. \end{cases}$$

The preceding construction can be summarized as the following implementation-ready procedure (Algorithm 1).

Algorithm 1 Model-free multivariate forecasting.

Given observations $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^m$, a truncation parameter $0 \leq k \leq n-2$, and a forecast horizon $h \geq 1$, compute forecasts $\hat{\mathbf{X}}_{n+1}^{(k)}, \dots, \hat{\mathbf{X}}_{n+h}^{(k)}$ as follows.

Step 1: Construct the response matrix $\mathcal{Y}(n, k)$ and the design matrix $\mathcal{X}(n, k)$ from the observed series according to the definitions above.

Step 2: Compute the least-squares estimator

$$\hat{B}'_k = \mathcal{X}(n, k)^\dagger \mathcal{Y}(n, k),$$

and extract from $\hat{B}'_k$ the coefficient matrices $\hat{\Gamma}_1, \dots, \hat{\Gamma}_{k+1}$ and the intercept vector $\hat{\mathbf{c}}$.

Step 3: Form the one-step-ahead forecast

$$\hat{\mathbf{X}}_{n+1}^{(k)} = \hat{\mathbf{c}} + \sum_{i=1}^{k+1} \hat{\Gamma}_i \mathbf{X}_{n+1-i}.$$

Step 4: For each $s = 2, \dots, h$, compute the recursive forecast

$$\hat{\mathbf{X}}_{n+s}^{(k)} = \hat{\mathbf{c}} + \sum_{i=1}^{k+1} \hat{\Gamma}_i \tilde{\mathbf{X}}_{n+s-i},$$

where $\tilde{\mathbf{X}}_{n+s-i} = \hat{\mathbf{X}}_{n+s-i}^{(k)}$ if $s-i > 0$, and $\tilde{\mathbf{X}}_{n+s-i} = \mathbf{X}_{n+s-i}$ otherwise.

Step 5: Return $\hat{\mathbf{X}}_{n+1}^{(k)}, \dots, \hat{\mathbf{X}}_{n+h}^{(k)}$.

Implementation choices and non-uniqueness. The theoretical representation in Theorems 1 and 2 is not unique. Different choices of the VARMA-type polynomials, the centering vector, and the right inverse in the associated linear system may lead to different associated infinite-past solutions. This non-uniqueness is a feature of the existence theory and should not be interpreted as introducing additional tuning parameters in the forecasting algorithm.

In the empirical implementation, the practitioner does not choose among all possible theoretical VARMA-type representations. Instead, the forecasting procedure uses the observed lagged vectors to build a finite-lag design matrix and estimates the truncated coefficient matrix directly by least squares, or by a regularized version such as ridge regression. Thus, the practical implementation choices are the candidate lag grid, the validation rule used to select the truncation parameter k , the intercept convention, and the regularization rule when a regularized estimator is used.

In the numerical studies below, k is selected by minimizing the validation-set forecasting error over a fixed candidate lag grid. The resulting fitted coefficient matrices are therefore forecasting coefficients for the selected finite-lag approximation. They should not be interpreted as unique structural parameters or as coefficients of a uniquely identified VARMA data-generating model.

Remark 8. If we impose $\boldsymbol{\mu} = \mathbf{0}$ in the VARMA-type representation, then the predictor becomes

$$\hat{\mathbf{X}}_{n+1}^{(k)} = \sum_{i=1}^{k+1} \hat{\Gamma}_i \mathbf{X}_{n+1-i}.$$

Since

$$\mathbf{c} = \left(I_m - \sum_{j=1}^{\infty} \Gamma_j \right) \boldsymbol{\mu},$$

it follows that $\mu = \mathbf{0}$ implies $\mathbf{c} = \mathbf{0}$ in the theoretical infinite-past representation, which motivates dropping the intercept in the finite-lag predictor. In this case,

$$\widehat{\mathbf{G}}_k = \begin{pmatrix} \widehat{\Gamma}_1 & \cdots & \widehat{\Gamma}_{k+1} \end{pmatrix} \in \mathbb{R}^{m \times (k+1)m}$$

may be estimated by solving

$$\widehat{\mathbf{G}}'_k = \arg \min_{M \in \mathbb{R}^{(k+1)m \times m}} \|\mathcal{Y}(n, k) - \mathcal{X}_0(n, k)M\|_F^2,$$

where

$$\mathcal{X}_0(n, k) = \begin{pmatrix} \mathbf{X}'_{n-1} & \mathbf{X}'_{n-2} & \cdots & \mathbf{X}'_{n-k-1} \\ \mathbf{X}'_{n-2} & \mathbf{X}'_{n-3} & \cdots & \mathbf{X}'_{n-k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}'_{k+1} & \mathbf{X}'_k & \cdots & \mathbf{X}'_1 \end{pmatrix} \in \mathbb{R}^{(n-k-1) \times (k+1)m}.$$

A least-squares solution is

$$\widehat{\mathbf{G}}'_k = \mathcal{X}_0(n, k)^\dagger \mathcal{Y}(n, k).$$

Remark 9. Other multivariate regression methods may also be used in place of ordinary least squares. For example, ridge regression Hoerl and Kennard [33], reduced-rank regression Izenman [34], lasso-type penalties Tibshirani [30], elastic-net penalties Zou and Hastie [32], and group penalties Yuan and Lin [31] can be useful when the columns of $\mathcal{X}(n, k)$ are highly correlated or when the dimension m is moderately large. Related high-dimensional time-series and structured VAR methods include Basu and Michailidis [27], Nicholson et al. [28]. A ridge-regularized model-free estimator can be defined by

$$\widehat{\mathbf{B}}'_{k, \lambda} = \arg \min_M \left\{ \|\mathcal{Y}(n, k) - \mathcal{X}(n, k)M\|_F^2 + \lambda \|M_{-0}\|_F^2 \right\},$$

where M_{-0} denotes the rows of M corresponding to lagged variables, so the intercept row is not penalized. Similarly, lasso and elastic-net versions may be obtained by replacing the ridge penalty with

$$\lambda \left[\alpha \|M_{-0}\|_1 + (1 - \alpha) \|M_{-0}\|_F^2 \right].$$

The tuning parameters λ and α may be selected on a validation sample. These regularized variants are especially relevant when m or k is large relative to the training length.

Remark 10. Although the focus here is forecasting, the same truncated representation may also be used to reconstruct the observed series. Specifically, for $t = k + 2, \dots, n$, one may define

$$\widehat{\mathbf{X}}_t^{(k)} = \widehat{\mathbf{c}} + \sum_{i=1}^{k+1} \widehat{\Gamma}_i \mathbf{X}_{t-i},$$

or, in the zero-intercept case,

$$\widehat{\mathbf{X}}_t^{(k)} = \sum_{i=1}^{k+1} \widehat{\Gamma}_i \mathbf{X}_{t-i}.$$

4. Accuracy of the Multivariate Predictor

The representation results above require only the uniformly bounded first-moment condition (1). Forecast accuracy of the estimated least-squares predictor, however, also depends on the numerical and probabilistic behavior of the lag-design matrix. This section states multivariate analogues of the univariate accuracy result, separating these two levels of assumptions.

In this section, $\|\cdot\|$ on vectors in the stacked lag space $\mathbb{R}^{d_k}$ denotes the Euclidean norm, $\|\cdot\|_{\text{op}}$ denotes the associated spectral norm on matrices, and $\|\cdot\|_F$ denotes the Frobenius norm. The generic compatible norms used in Sections 2 and 3 remain unchanged.

Assumption 1 (Selected-solution identification). *For the representation choice fixed in Theorem 2, the selected associated solution coincides with the observed process; that is, $\mathbf{X}_t^* = \mathbf{X}_t$ for all $t \in \mathbb{Z}$.*

Assumption 1 is an additional condition for the accuracy results and is not asserted to hold automatically for every data-generating process considered in the empirical study.

Throughout this section, we work under Assumption 1 and study prediction for the observed process $\mathbf{X}_t$. This assumption is an identifying convention for the selected associated solution. It is not implied by Theorem 2 and should not be interpreted as uniqueness of the representation; rather, it is the extra condition needed to read the accuracy results as statements about the observed future vectors instead of only about the selected associated solution.

For a truncation level k , define the stacked lag vector

$$\mathbf{W}_t^{(k)} = \begin{pmatrix} \mathbf{X}'_{t-1} & \mathbf{X}'_{t-2} & \cdots & \mathbf{X}'_{t-k-1} & 1 \end{pmatrix}' \in \mathbb{R}^{d_k}, \quad d_k = (k+1)m+1,$$

and write

$$B_k = \begin{pmatrix} \Gamma_1 & \cdots & \Gamma_{k+1} & \mathbf{c} \end{pmatrix} \in \mathbb{R}^{m \times d_k}.$$

Under Assumption 1, the finite-lag approximation can be written as

$$\mathbf{X}_t = B_k \mathbf{W}_t^{(k)} + \mathbf{R}_t^{(k)}, \quad \mathbf{R}_t^{(k)} = \sum_{j=k+2}^{\infty} \Gamma_j \mathbf{X}_{t-j},$$

where $\mathbf{R}_t^{(k)}$ is the truncation remainder. Proposition 1 implies $\mathbf{R}_t^{(k)} \rightarrow 0$ in L_1 , uniformly in t , as $k \rightarrow \infty$.

Assumption 2 (Prediction-design conditions). *Let $k = k_n \rightarrow \infty$, $d_n = (k_n + 1)m + 1$, and $N_n = n - k_n - 1$. Suppose that:*

(A1) $d_n/N_n \rightarrow 0$.

(A2) *With probability tending to one, the empirical Gram matrix*

$$\widehat{\Sigma}_n = \frac{1}{N_n} \sum_{t=k_n+2}^n \mathbf{W}_t^{(k_n)} \mathbf{W}_t^{(k_n)'}.$$

is nonsingular and satisfies $\lambda_{\min}(\widehat{\Sigma}_n) \geq c_0$ for some constant $c_0 > 0$.

(A3) *The regressor–remainder term is small at the prediction scale:*

$$\left\| \frac{1}{N_n} \sum_{t=k_n+2}^n \mathbf{W}_t^{(k_n)} \mathbf{R}_t^{(k_n)'} \right\|_F = o_P(d_n^{-1/2}).$$

(A4) *The new lag vector satisfies $\|\mathbf{W}_{n+1}^{(k_n)}\| = O_P(d_n^{1/2})$.*

Remark 11. *Assumption 2 is a high-level form of the conditions needed for stability of the multivariate least-squares predictor. It can be verified, for example, under standard weak dependence and finite $2 + \delta$ moment conditions, together with a nondegenerate lag-design matrix and a sufficiently*

small truncation remainder. These conditions are not needed for the infinite-past representation; they are imposed only to justify the estimated predictor.

The following result is a multivariate analogue of the scalar one-step accuracy theorem of [1].

Theorem 3 (One-step consistency). Assume (1), Assumption 1, $\sum_{j \geq 1} \|\Gamma_j\| < \infty$, and Assumption 2. Let $\widehat{B}_{k_n}$ be the least-squares estimator in (11); under Assumption 2 this is equivalently given by the usual full-rank formula. Define

$$\widehat{\mathbf{X}}_{n+1}^{(k_n)} = \widehat{B}_{k_n} \mathbf{W}_{n+1}^{(k_n)}.$$

Then

$$\|\widehat{\mathbf{X}}_{n+1}^{(k_n)} - \mathbf{X}_{n+1}\| \xrightarrow{P} 0.$$

Proof. Write $Y_n = \mathcal{Y}(n, k_n)$, $D_n = \mathcal{X}(n, k_n)$, and let R_n be the $N_n \times m$ matrix with rows $\mathbf{R}_t^{(k_n)'}$, $t = k_n + 2, \dots, n$. Then D_n has rows $\mathbf{W}_t^{(k_n)'}$ and the regression identity is

$$Y_n = D_n B'_{k_n} + R_n.$$

Moreover,

$$D'_n D_n = N_n \widehat{\Sigma}_n,$$

so Assumption 2 (A2) yields

$$\left\| \left(\frac{1}{N_n} D'_n D_n \right)^{-1} \right\|_{\text{op}} = \|\widehat{\Sigma}_n^{-1}\|_{\text{op}} \leq c_0^{-1}$$

with probability tending to one. When $D'_n D_n$ is nonsingular,

$$\widehat{B}'_{k_n} - B'_{k_n} = (D'_n D_n)^{-1} D'_n R_n.$$

Hence, by Assumption 2,

$$\|\widehat{B}_{k_n} - B_{k_n}\|_F \leq c_0^{-1} \left\| \frac{1}{N_n} D'_n R_n \right\|_F = o_P(d_n^{-1/2}).$$

Therefore,

$$\|(\widehat{B}_{k_n} - B_{k_n}) \mathbf{W}_{n+1}^{(k_n)}\| \leq \|\widehat{B}_{k_n} - B_{k_n}\|_F \|\mathbf{W}_{n+1}^{(k_n)}\| = o_P(1).$$

Finally,

$$\widehat{\mathbf{X}}_{n+1}^{(k_n)} - \mathbf{X}_{n+1} = (\widehat{B}_{k_n} - B_{k_n}) \mathbf{W}_{n+1}^{(k_n)} - \mathbf{R}_{n+1}^{(k_n)}.$$

The first term is $o_P(1)$, and the second term converges to zero in probability by Proposition 1. This proves the result. $\square$

Theorem 4 (Fixed-horizon recursive consistency). Let $h \geq 1$ be fixed. Suppose that the conditions of Theorem 3 hold for the fitted predictor based on $\mathbf{X}_1, \dots, \mathbf{X}_n$, and that the lag vectors and truncation remainders appearing in the recursive forecasts satisfy the same boundedness and small-remainder conditions uniformly for $r = 1, \dots, h$. Then the recursive predictor (13) satisfies

$$\max_{1 \leq r \leq h} \|\widehat{\mathbf{X}}_{n+r}^{(k_n)} - \mathbf{X}_{n+r}\| \xrightarrow{P} 0.$$

Proof. The case $r = 1$ follows from Theorem 3. For $r > 1$, subtract the infinite-past approximation from the recursive forecast. The error is a sum of three terms: the coefficient-estimation error, the truncation remainder, and a finite linear combination of the previous forecast errors $\hat{\mathbf{X}}_{n+s}^{(k_n)} - \mathbf{X}_{n+s}$, $s < r$, premultiplied by $\hat{\Gamma}_i$. The coefficient-estimation term and the truncation remainder are $o_P(1)$ by the same argument used for the one-step forecast. Moreover, $\|\hat{B}_{k_n} - B_{k_n}\|_F = o_P(d_n^{-1/2})$ and $\sum_{j \geq 1} \|\Gamma_j\| < \infty$ imply that any fixed finite collection of estimated matrices $\hat{\Gamma}_1, \dots, \hat{\Gamma}_{h-1}$ is $O_P(1)$. Hence multiplication by these matrices preserves $o_P(1)$ errors. Since h is fixed, the finite number of propagated earlier forecast errors is therefore $o_P(1)$ by induction. Thus the result follows for every fixed $r \leq h$. $\square$

Remark 12 (Long horizons and high dimension). *The fixed-horizon result does not automatically imply uniform accuracy when $h = h_n \rightarrow \infty$. For growing horizons, a stability condition on the companion matrix associated with $\Gamma_1, \dots, \Gamma_{k_n+1}$ is needed to prevent recursive forecast errors from accumulating. Similarly, when m or k_n is large relative to n , unrestricted least squares can become unstable because the number of regression coefficients is of order $m((k_n + 1)m + 1)$. In such cases, ridge regression [33], reduced-rank regression [34], or sparsity/group penalties [30,31] are natural regularized versions of the proposed model-free predictor.*

5. Simulation and Empirical Study

We compare the proposed model-free predictor with ridge regression and a channel-independent linear benchmark across four simulation settings: a stable VAR(2) model, a two-regime threshold VAR (TVAR), a VAR(2) model with Student- t innovations, and a VAR(2) model with one structural break. For each setting, we consider dimensions $m = 3, 5, 10$ and conduct 100 Monte Carlo replications. In each replication, the generated series is split chronologically into training, validation, and test sets. The lag parameter is selected on the validation set, and forecasting performance is evaluated on the test set using rolling multi-step prediction. Tables 1–4 report the confidence intervals of MAE and MSE across replications.

The channel-independent linear benchmark fits separate univariate linear regressions for the individual component series, using the same candidate lag set and validation rule as the multivariate methods. The main-text tables therefore compare the proposed model-free predictor with two distinct linear benchmarks: ridge regression and channel-independent linear forecasting.

For full reproducibility, we record the exact simulation designs used in the code.

In the stable VAR(2) setting, the process is generated as

$$\mathbf{X}_t = \mathbf{c} + A_1 \mathbf{X}_{t-1} + A_2 \mathbf{X}_{t-2} + \boldsymbol{\varepsilon}_t, \quad t \geq 2,$$

where $\mathbf{c} \in \mathbb{R}^m$ has entries

$$c_i = 0.25 - \frac{0.30(i-1)}{m-1}, \quad i = 1, \dots, m.$$

The diagonal entries of A_1 and A_2 are

$$(A_1)_{ii} = 0.52 - \frac{0.20(i-1)}{m-1}, \quad (A_2)_{ii} = -0.20 + \frac{0.10(i-1)}{m-1}.$$

For $i = 1, \dots, m-1$, the nearest-neighbor coefficients are

$$(A_1)_{i,i+1} = 0.08, \quad (A_1)_{i+1,i} = 0.05, \quad (A_2)_{i,i+1} = 0.03, \quad (A_2)_{i+1,i} = 0.02,$$

and for $i = 1, \dots, m - 2$ the second-neighbor coefficients are

$$(A_1)_{i,i+2} = -0.03, \quad (A_2)_{i+2,i} = 0.02,$$

with all remaining entries equal to zero. The Gaussian innovations satisfy

$$\varepsilon_t \sim N_m(0, \Sigma),$$

where

$$\Sigma_{ij} = 0.12 s_i s_j 0.35^{|i-j|}, \quad s_i = 1 - \frac{0.15(i-1)}{m-1}, \quad i, j = 1, \dots, m.$$

In the TVAR setting, the process is generated as

$$\mathbf{X}_t = \begin{cases} \mathbf{c}^{(1)} + A_1^{(1)} \mathbf{X}_{t-1} + A_2^{(1)} \mathbf{X}_{t-2} + \varepsilon_t, & X_{t-1,1} \leq 0, \\ \mathbf{c}^{(2)} + A_1^{(2)} \mathbf{X}_{t-1} + A_2^{(2)} \mathbf{X}_{t-2} + \varepsilon_t, & X_{t-1,1} > 0, \end{cases} \quad t \geq 2,$$

so the threshold variable is the first component, the delay is 1, and the threshold value is 0. Here $(\mathbf{c}^{(1)}, A_1^{(1)}, A_2^{(1)})$ and $(\mathbf{c}^{(2)}, A_1^{(2)}, A_2^{(2)})$ are constructed from the stable VAR(2) matrices as follows. First, set

$$\mathbf{c}^{(1)} = \mathbf{c}, \quad \mathbf{c}^{(2)} = \mathbf{c} + (0.03, \dots, -0.03)'$$

with linear interpolation across coordinates. For regime 1, increase each diagonal entry of the base matrix A_1 by 0.05, cap it at 0.65, and then multiply both A_1 and A_2 by 0.92. For regime 2, decrease each diagonal entry of the base matrix A_1 by 0.10, floor it at 0.20, and multiply the nearest-neighbor coefficients of A_1 by 1.8 and those of A_2 by 1.5, leaving the remaining coefficients unchanged. The innovation covariance matrix is again Σ .

In the Student- t VAR(2) setting, the coefficient matrices and intercept are the same as in the stable VAR(2) design, but the innovations are heavy-tailed. Specifically,

$$\varepsilon_t = Z_t \sqrt{\nu / U_t},$$

where $Z_t \sim N_m(0, \Sigma)$, $U_t \sim \chi_\nu^2$, the two are independent, and $\nu = 5$. Thus the innovations have a multivariate Student- t distribution with 5 degrees of freedom and scale matrix Σ . Equivalently, their covariance matrix is

$$\text{Cov}(\varepsilon_t) = \frac{\nu}{\nu - 2} \Sigma = \frac{5}{3} \Sigma.$$

In the structural-break VAR(2) setting, the process is generated as

$$\mathbf{X}_t = \begin{cases} \mathbf{c}^{(1)} + A_1^{(1)} \mathbf{X}_{t-1} + A_2^{(1)} \mathbf{X}_{t-2} + \varepsilon_t, & t \leq t_0, \\ \mathbf{c}^{(2)} + A_1^{(2)} \mathbf{X}_{t-1} + A_2^{(2)} \mathbf{X}_{t-2} + \varepsilon_t, & t > t_0, \end{cases} \quad t \geq 2,$$

with break point

$$t_0 = 200 + \lfloor 0.5 \times 1200 \rfloor = 800,$$

that is, after 600 post-burn observations. The pre-break parameters $(\mathbf{c}^{(1)}, A_1^{(1)}, A_2^{(1)})$ are the stable VAR(2) values. The post-break intercept is

$$\mathbf{c}^{(2)} = \mathbf{c}^{(1)} + (0.04, \dots, -0.04)'$$

with linear interpolation across coordinates. The post-break diagonal entries of $A_1^{(2)}$ are obtained from those of $A_1^{(1)}$ by subtracting 0.10 and then adding an alternating adjustment of ± 0.04 , starting with -0.04 in the first component, with a lower bound of 0.18. The post-break diagonal entries of $A_2^{(2)}$ are obtained by adding 0.05 to those of $A_2^{(1)}$ and capping them above by -0.05 . For nearest-neighbor terms, the upper-diagonal coefficients of $A_1^{(2)}$ and $A_2^{(2)}$ are obtained by multiplying those of the stable VAR(2) design by 1.6 and 1.2, respectively, while the lower-diagonal coefficients are obtained by multiplying by 0.4 and 0.6, respectively. The innovation covariance matrix remains Σ .

In every replication, we generate a post-burn sample of length $n = 1200$ after a burn-in of 200 observations and split the series chronologically into 720 training observations, 240 validation observations, and 240 test observations. The forecast horizon is $h = 12$, the rolling evaluation step size is 12, and the candidate lag set for validation is

$$\{0, 1, 2, 3, 5, 7\}.$$

For lag-based regression methods, the fitted empirical lag order corresponding to a truncation level k is

$$p_{\text{lag}} = k + 1,$$

so the implied candidate lag-order set is $\{1, 2, 3, 4, 6, 8\}$. Ridge regression is implemented with a fixed penalty parameter $\alpha = 1$; no ridge-penalty grid is used. For each method, the lag parameter is selected by minimizing the validation-set mean absolute error.

For each rolling test window w , the window-level MAE and MSE are computed as

$$\text{MAE}_w = \frac{1}{hm} \sum_{r=1}^h \sum_{j=1}^m |X_{t_w+r,j} - \hat{X}_{t_w+r,j}|,$$

and

$$\text{MSE}_w = \frac{1}{hm} \sum_{r=1}^h \sum_{j=1}^m (X_{t_w+r,j} - \hat{X}_{t_w+r,j})^2.$$

The reported test MAE and test MSE for a given replication are the averages of these window-level quantities across all rolling windows in the test sample. Tables 1–4 report the Monte Carlo means of MAE and MSE together with 95% confidence intervals for those means.

Several conclusions emerge from these results. First, ridge regression often attains slightly smaller errors than the proposed model-free predictor, especially in lower-dimensional settings, which suggests that mild regularization can improve finite-sample stability. Second, the channel-independent linear benchmark performs best in several settings with $m = 5$ and $m = 10$, particularly under heavy-tailed innovations and structural breaks. This suggests that, in moderate or high dimension, reduced model complexity may offset the loss from ignoring cross-series dependence. Finally, in the TVAR setting the differences among the reported methods are small, indicating that the nonlinear regime-switching mechanism considered here is still reasonably well approximated by linear predictors at the chosen forecast horizon.

Overall, the Monte Carlo study shows that the proposed lag-based implementation is competitive with standard linear forecasting approaches across a variety of linear, nonlinear, heavy-tailed, and nonstationary settings. However, it does not uniformly dominate the competing methods: ridge regression is often slightly better in low dimension, whereas the channel-independent benchmark becomes more attractive as the dimension increases.

Table 1. Monte Carlo results for the stable VAR(2) setting. Reported values are means with 95% confidence intervals.

<i>m</i>	Method	MAE [95% CI]	MSE [95% CI]
3	Ridge	0.284112 [0.282201, 0.286023]	0.127959 [0.126252, 0.129666]
	Model-free	0.284124 [0.282213, 0.286035]	0.127969 [0.126262, 0.129676]
	Channel-linear	0.284213 [0.282280, 0.286146]	0.127994 [0.126268, 0.129720]
5	Channel-linear	0.282634 [0.280977, 0.284291]	0.126457 [0.125054, 0.127860]
	Ridge	0.282660 [0.281017, 0.284303]	0.126476 [0.125099, 0.127853]
	Model-free	0.282672 [0.281028, 0.284316]	0.126484 [0.125106, 0.127862]
10	Channel-linear	0.283889 [0.282631, 0.285147]	0.127030 [0.125980, 0.128080]
	Ridge	0.284102 [0.282816, 0.285388]	0.127227 [0.126153, 0.128301]
	Model-free	0.284122 [0.282835, 0.285409]	0.127244 [0.126170, 0.128318]

Table 2. Monte Carlo results for the two-regime TVAR setting. Reported values are means with 95% confidence intervals.

<i>m</i>	Method	MAE [95% CI]	MSE [95% CI]
3	Ridge	0.280358 [0.278470, 0.282246]	0.124623 [0.122952, 0.126294]
	Model-free	0.280366 [0.278479, 0.282253]	0.124631 [0.122960, 0.126302]
	Channel-linear	0.280587 [0.278682, 0.282492]	0.124790 [0.123098, 0.126482]
5	Ridge	0.279581 [0.277951, 0.281211]	0.123742 [0.122382, 0.125102]
	Model-free	0.279585 [0.277952, 0.281218]	0.123737 [0.122373, 0.125101]
	Channel-linear	0.279711 [0.278063, 0.281359]	0.123875 [0.122482, 0.125268]
10	Ridge	0.281404 [0.280127, 0.282681]	0.124794 [0.123733, 0.125855]
	Model-free	0.281417 [0.280138, 0.282696]	0.124807 [0.123744, 0.125870]
	Channel-linear	0.281526 [0.280266, 0.282786]	0.124897 [0.123849, 0.125945]

Table 3. Monte Carlo results for the VAR(2) setting with Student-*t* innovations. Reported values are means with 95% confidence intervals.

<i>m</i>	Method	MAE [95% CI]	MSE [95% CI]
3	Ridge	0.345114 [0.341751, 0.348477]	0.212498 [0.207095, 0.217901]
	Channel-linear	0.345135 [0.341749, 0.348521]	0.212517 [0.207081, 0.217953]
	Model-free	0.345150 [0.341785, 0.348515]	0.212544 [0.207138, 0.217950]
5	Channel-linear	0.342997 [0.340036, 0.345958]	0.210308 [0.205302, 0.215314]
	Ridge	0.343067 [0.340102, 0.346032]	0.210262 [0.205315, 0.215209]
	Model-free	0.343096 [0.340130, 0.346062]	0.210296 [0.205347, 0.215245]
10	Channel-linear	0.343194 [0.340639, 0.345749]	0.207701 [0.204000, 0.211402]
	Model-free	0.343433 [0.340832, 0.346034]	0.207957 [0.204210, 0.211704]
	Ridge	0.343455 [0.340859, 0.346051]	0.207969 [0.204227, 0.211711]

Table 4. Monte Carlo results for the VAR(2) setting with one structural break. Reported values are means with 95% confidence intervals.

<i>m</i>	Method	MAE [95% CI]	MSE [95% CI]
3	Ridge	0.275404 [0.273617, 0.277191]	0.120113 [0.118588, 0.121638]
	Model-free	0.275411 [0.273624, 0.277198]	0.120119 [0.118594, 0.121644]
	Channel-linear	0.275529 [0.273730, 0.277328]	0.120189 [0.118647, 0.121731]
5	Ridge	0.274430 [0.272882, 0.275978]	0.119278 [0.118021, 0.120535]
	Model-free	0.274444 [0.272895, 0.275993]	0.119293 [0.118037, 0.120549]
	Channel-linear	0.274457 [0.272898, 0.276016]	0.119265 [0.117994, 0.120536]
10	Channel-linear	0.276482 [0.275343, 0.277621]	0.120389 [0.119444, 0.121334]
	Ridge	0.276484 [0.275325, 0.277643]	0.120415 [0.119456, 0.121374]
	Model-free	0.276498 [0.275339, 0.277657]	0.120427 [0.119468, 0.121386]

The real-data macroeconomic application is motivated by the empirical forecasting literature using macroeconomic variables, factor methods, and high-dimensional prediction tools Bai and Ng [26], Stock and Watson [29], Bernanke et al. [35], Banbura et al. [36], Mc-

Cracken and Ng [37]. The purpose of this empirical example is not to claim dominance over all benchmarks, but to show that the proposed model-free finite-lag predictor can be implemented directly on real multivariate time-series data and can achieve competitive forecast accuracy with a simple fitting procedure. For the real-data application, the truncation parameter k is selected from $\{0, 1, 2, 3, 5, 7\}$ by validation MAE. For ridge regression, the penalty parameter is selected from the grid $\{0.01, 0.1, 1, 10, 100\}$ on the validation sample. Because small differences in forecasting errors may not be meaningful, formal forecast-comparison tools such as the Diebold–Mariano test and the Model Confidence Set provide useful complementary evidence Diebold and Mariano [38], Hansen et al. [39].

The selected tuning parameters and forecast-accuracy results for the real-data application are reported in Table 5.

Table 5. Real-data macroeconomic forecasting application. The model-free rows with fixed k illustrate the effect of the truncation parameter. Here $p_{\text{lag}} = k + 1$. Forecasts are non-overlapping four-quarter recursive forecasts, and accuracy is reported in standardized units.

Method	p_{lag}	k	α	MAE	MSE
Model-free, fixed $k = 0$	1	0	–	0.5964	0.7454
Model-free, fixed $k = 1$	2	1	–	0.6520	0.8315
Model-free, fixed $k = 2$	3	2	–	0.6705	0.8626
Model-free, selected	1	0	–	0.5964	0.7454
Ridge, selected	3	2	10.00	0.6559	0.8358
Channel-linear, selected	3	2	–	0.5969	0.7385

In this real-data application, the validation rule selects $k = 0$, equivalently $p_{\text{lag}} = 1$, for the model-free predictor. The fixed- k rows show that increasing the truncation level worsens forecast accuracy in this dataset, suggesting that the additional lag coefficients increase estimation variance more than they reduce truncation bias. The model-free predictor attains the smallest MAE among the reported methods, whereas the channel-independent linear benchmark attains the smallest MSE. Thus, the real-data results support competitive performance rather than uniform dominance. Paired tests over the 11 test windows gave a model-free minus ridge MAE difference of -0.0595 with 95% confidence interval $[-0.1466, 0.0275]$, so the difference was not statistically significant at the 5% level.

To assess the robustness of the empirical conclusions under more strongly nonlinear, heavy-tailed, and contaminated settings, Table 6 summarizes the average forecast-error ranks for the expanded set of competing methods.

Table 6. Average ranks in additional screening simulations with bilinear, smooth nonlinear, heavier-tailed Student- t ($\nu = 3$), and contaminated-shock designs. Lower ranks indicate smaller forecast errors. These results summarize the expanded benchmark set used to assess robustness claims.

Method	Average MAE Rank	Average MSE Rank
Channel-linear	1.00	1.00
LASSO	2.50	2.67
Factor-ridge	3.17	2.92
Elastic net	3.67	3.75
Ridge	4.83	4.83
Model-free	5.83	5.83

The additional screening simulations involved genuinely nonlinear and more strongly heavy-tailed or contaminated designs, together with LASSO, elastic-net, factor-ridge, and channel-independent benchmarks. They do not support a claim of uniform empirical superiority. Instead, they support the more cautious interpretation that the model-free

implementation is a flexible forecasting framework whose unregularized least-squares form can be matched or outperformed by simpler or regularized procedures in finite samples.

Additional sensitivity analyses concerning regularization, the sample-to-parameter ratio, the truncation level, and the series dimension are reported in Table 7.

Table 7. Additional sensitivity checks for regularization, sample/parameter ratio, truncation, and dimension. Here p_{lag} denotes the empirical fitted lag order, so $k = p_{\text{lag}} - 1$. The regularized/benchmark column reports model-free ridge for the first five rows and channel-linear for the last two rows.

Check	Setting	Model-Free	Reg./Bench.	Main Implication
Reg. model-free	Nonlinear tanh, $m = 10$, $p_{\text{lag}} = 4$	0.9133	0.8838	Ridge regularization improves the model-free estimator.
Sample/param.	Stable VAR(2), $m = 10$, $p_{\text{lag}} = 8$, training = 90	5.7644	1.1358	Unrestricted least squares is unstable when the sample/parameter ratio is near 1.
Sample/param.	Stable VAR(2), $m = 10$, $p_{\text{lag}} = 8$, training = 220	0.9295	0.8970	Larger samples reduce instability, but regularization remains helpful.
Truncation	Nonlinear tanh, $m = 10$, $p_{\text{lag}} = 1$	0.8185	0.8147	Small truncation gives low variance but may be biased.
Truncation	Nonlinear tanh, $m = 10$, $p_{\text{lag}} = 4$	0.8939	0.8575	Moderate truncation increases flexibility, but ridge regularization remains helpful.
Truncation	Nonlinear tanh, $m = 10$, $p_{\text{lag}} = 12$	3.9825	1.1986	Large truncation can greatly increase estimation variance.
High dimension	Stable VAR(2), $m = 20$, $p_{\text{lag}} = 4$	1.2223	0.7830	Channel-wise forecasting was best in this screening setting.
High dimension	Stable VAR(2), $m = 50$, $p_{\text{lag}} = 4$	1.3889	0.7757	Dimension reduction or regularization is needed at larger m .

These additional results clarify the finite-sample bias–variance tradeoff. A small truncation level may omit useful lag information, while a large truncation level increases the number of coefficients and can make unrestricted least squares unstable. The number of scalar coefficients in the unrestricted estimator with intercept is $(k + 1)m^2 + m$, or equivalently $(k + 1)m + 1$ regressors per equation. This explains why regularization or dimension reduction is practically important when m or k is large relative to the training sample.

6. Discussion and Conclusions

We extended the scalar model-free prediction framework of Mohammadi and Li [1] to multivariate time series. Under a mild uniformly bounded first-moment condition, we established a VARMA-type representation and constructed an associated infinite-past vector linear solution. This led to a practical forecasting method based on truncation of the infinite-past expansion and estimation of the resulting coefficient matrices by multivariate least squares.

The theoretical development separates two levels of assumptions. The representation results are model-free and require only a uniformly bounded first moment. The stated accuracy results for the estimated predictor require the selected-solution identification assumption and additional prediction-design conditions, reflecting the fact that multivariate least squares must be stable in finite samples. Under Assumption 1 and these design conditions, the one-step predictor is consistent, and the recursive predictor is consistent for every fixed forecast horizon.

The Monte Carlo study showed that the proposed lag-based implementation is competitive with standard linear forecasting approaches across a range of linear, nonlinear,

heavy-tailed, and nonstationary settings. In particular, its forecasting performance was close to that of the ridge and channel-independent linear benchmarks, although these competing methods were sometimes preferable in finite samples depending on the dimension and the form of misspecification.

The real-data application and additional sensitivity checks support the same cautious interpretation. The model-free predictor can be practically competitive, but its unregularized least-squares implementation does not uniformly dominate ridge, lasso, elastic-net, factor-based, or channel-independent alternatives. When the number of lagged regressors is large relative to the training length, regularized model-free variants or dimension-reduction methods are preferable.

Finally, the proposed method is intended for prediction rather than structural analysis. The estimated coefficient matrices are forecasting coefficients arising from a truncated model-free linear representation. They should not be interpreted as structural parameters, causal effects, impulse-response coefficients, or policy multipliers. Structural analysis, counterfactual policy evaluation, and causal identification require additional assumptions not imposed in this paper.

Further work may study sharper high-dimensional theory, asymptotic distributions for the estimated coefficient matrices, regularized and reduced-rank versions of the method, and applications to real multivariate datasets.

Author Contributions: Conceptualization, A.M.; Methodology, A.M.; Software, A.M.; Validation, H.S.; Formal analysis, H.S.; Resources, A.M.; Data curation, A.M.; Writing—original draft, H.S. and A.M.; Review and editing, H.S.; Supervision, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors sincerely thank the Associate Editor and the three anonymous reviewers for their careful reading of the manuscript and for their constructive comments and suggestions, which helped improve the clarity, presentation, and empirical evaluation of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mohammadi, M.; Li, M. Model-free prediction of time series: A nonparametric approach. *J. Nonparametric Stat.* **2024**, *36*, 804–824.
2. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.
3. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*; Springer: Berlin/Heidelberg, Germany, 1991.
4. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.
5. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021.
6. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2005.
7. Pourahmadi, M. *Foundations of Time Series Analysis and Prediction Theory*; Wiley: Hoboken, NJ, USA, 2001.
8. Tsay, R.S. *Multivariate Time Series Analysis: With R and Financial Applications*; Wiley: Hoboken, NJ, USA, 2014.
9. Wei, W.W.S. *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed.; Pearson: London, UK, 2006.
10. Biau, G.; Bleakley, K.; Györfi, L.; Ottucsák, G. Nonparametric sequential prediction of time series. *J. Nonparametric Stat.* **2010**, *22*, 297–317. [[CrossRef](#)]
11. Chernozhukov, V.; Wüthrich, K.; Zhu, Y. Distributional conformal prediction. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2107794118. [[CrossRef](#)] [[PubMed](#)]
12. Härdle, W.; Lütkepohl, H.; Chen, R. A review of nonparametric time series analysis. *Int. Stat. Rev.* **1997**, *65*, 49–72. [[CrossRef](#)]
13. Politis, D.N. Model-free model-fitting and predictive distributions. *TEST* **2013**, *22*, 183–221. [[CrossRef](#)]
14. Politis, D.N. *Model-Free Prediction and Regression: A Transformation-Based Approach to Inference*; Springer: Berlin/Heidelberg, Germany, 2015.

15. Tschernig, R.; Yang, L. Nonparametric lag selection for time series. *J. Time Ser. Anal.* **2000**, *21*, 457–487. [[CrossRef](#)]
16. Vilar-Fernández, J.M.; Cao, R. Nonparametric forecasting in time series: A comparative study. *Commun. Stat. Simul. Comput.* **2007**, *36*, 311–334. [[CrossRef](#)]
17. Wang, Y.; Politis, D.N. Model-free bootstrap for a general class of stationary time series. *Bernoulli* **2022**, *28*, 744–770. [[CrossRef](#)]
18. Zeng, Z.; Li, M. Bayesian median autoregression for robust time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1000–1010. [[CrossRef](#)]
19. Zhang, G.P.; Qi, M. Neural network forecasting for seasonal and trend time series. *Eur. J. Oper. Res.* **2005**, *160*, 501–514. [[CrossRef](#)]
20. Cherstvy, A.G.; Vinod, D.; Aghion, E.; Chechkin, A.V.; Metzler, R. Time averaging, ageing and delay analysis of financial time series. *New J. Phys.* **2017**, *19*, 063045. [[CrossRef](#)]
21. Dahlhaus, R. Fitting time series models to nonstationary processes. *Ann. Stat.* **1997**, *25*, 1–37. [[CrossRef](#)]
22. Dette, H.; Wu, W.B. Prediction in locally stationary time series. *J. Bus. Econ. Stat.* **2022**, *40*, 370–381.
23. Fryzlewicz, P.; Van Belleghem, S.; von Sachs, R. Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Stat. Math.* **2003**, *55*, 737–764. [[CrossRef](#)]
24. Kley, T.; Preuß, P.; Fryzlewicz, P. Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **2019**, *13*, 3710–3774. [[CrossRef](#)]
25. Nason, G.P.; von Sachs, R.; Kroisandt, G. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B* **2000**, *62*, 271–292. [[CrossRef](#)]
26. Bai, J.; Ng, S. Determining the number of factors in approximate factor models. *Econometrica* **2002**, *70*, 191–221. [[CrossRef](#)]
27. Basu, S.; Michailidis, G. Regularized estimation in sparse high-dimensional time series models. *Ann. Stat.* **2015**, *43*, 1535–1567. [[CrossRef](#)]
28. Nicholson, W.B.; Matteson, D.S.; Bien, J. VARX-L: Structured regularization for large vector autoregressions with exogenous variables. *Int. J. Forecast.* **2017**, *33*, 627–651. [[CrossRef](#)]
29. Stock, J.H.; Watson, M.W. Forecasting using principal components from a large number of predictors. *J. Am. Stat. Assoc.* **2002**, *97*, 1167–1179. [[CrossRef](#)]
30. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
31. Yuan, M.; Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* **2006**, *68*, 49–67.
32. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [[CrossRef](#)]
33. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
34. Izenman, A.J. Reduced-rank regression for the multivariate linear model. *J. Multivar. Anal.* **1975**, *5*, 248–264. [[CrossRef](#)]
35. Bernanke, B.S.; Boivin, J.; Elias, P. Measuring the effects of monetary policy: A factor-augmented vector autoregressive approach. *Q. J. Econ.* **2005**, *120*, 387–422. [[CrossRef](#)]
36. Banbura, M.; Giannone, D.; Reichlin, L. Large Bayesian vector autoregressions. *J. Appl. Econom.* **2010**, *25*, 71–92.
37. McCracken, M.W.; Ng, S. FRED-MD: A monthly database for macroeconomic research. *J. Bus. Econ. Stat.* **2016**, *34*, 574–589. [[CrossRef](#)]
38. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
39. Hansen, P.R.; Lunde, A.; Nason, J.M. The model confidence set. *Econometrica* **2011**, *79*, 453–497. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.